# ADVANCED SCIENTIFIC LIBRARY
# ASL
## User's Guide
## <Basic Functions Vol.6>

# PREFACE

This manual describes general concepts, functions, and specifications for use of the Advanced Scientific Library (ASL).

The manuals corresponding to this product consist of seven volumes, which are divided into the chapters shown below. This manual describes the basic functions, volume 6.

Basic Functions Volume 1

| Chapter | Title | Contents |
|---|---|---|
| 1 | Introduction | Explanation of the organization of this manual, how to view each item, and usage limitations. |
| 2 | Storage Mode Conversion | Explanation of algorithms, method of using, and usage example of subroutine related to storage mode conversion of array data. |
| 3 | Basic Matrix Algebra | Explanation of algorithms, method of using, and usage example of subroutine related to basic calculations involving matrices. |
| 4 | Eigenvalues and Eigenvectors | Explanation of algorithms, method of using, and usage example of subroutine related to **the standard eigenvalue problem** for real matrices, complex matrices, real symmetric matrices, Hermitian matrices, real symmetric band matrices, real symmetric tridiagonal matrices, real symmetric random sparse matrices, Hermitian random sparse matrices and **the generalized eigenvalue problem** for real matrices, real symmetric matrices, Hermitian matrices, real symmetric band matrices. |

Basic Functions Volume 2

| Chapter | Title | Contents |
|---|---|---|
| 1 | Introduction | Explanation of the organization of this manual, how to view each item, and usage limitations. |
| 2 | Simultaneous Linear Equations (Direct Method) | Explanation of algorithms, method of using, and usage example of subroutine related to simultaneous linear equations corresponding to real matrices, complex matrices, positive symmetric matrices, real symmetric matrices, Hermitian matrices, real band matrices, positive symmetric band matrices, real tridiagonal matrices, real upper triangular matrices, and real lower triangular matrices. |

Basic Functions Volume 3

| Chapter | Title | Contents |
|---|---|---|
| 1 | Introduction | Explanation of the organization of this manual, how to view each item, and usage limitations. |
| 2 | Fourier Transforms and their applications | Explanation of algorithms, method of using, and usage example of subroutine related to one-, two- and three-dimensional complex Fourier transforms and real Fourier transforms, one-, two- and three-dimensional convolutions, correlations, and power spectrum analysis, wavelet transforms, and inverse Laplace transforms. |

Basic Functions Volume 4

| Chapter | Title | Contents |
|---|---|---|
| 1 | Introduction | Explanation of the organization of this manual, how to view each item, and usage limitations. |
| 2 | Differential Equations and Their Applications | Explanation of algorithms, method of using, and usage example of subroutine related to **ordinary differential equations initial value problems** for high-order simultaneous ordinary differential equations, implicit simultaneous ordinary differential equations, matrix type ordinary differential equations, stiff problem high-order simultaneous ordinary differential equations, simultaneous ordinary differential equations, first-order simultaneous ordinary differential equations, and high-order ordinary differential equations, and **ordinary differential equations boundary value problems** for high-order simultaneous ordinary differential equations, first-order simultaneous ordinary differential equations, high-order ordinary differential equations, high-order linear ordinary differential equations, and second-order linear ordinary differential equations, and **integral equations** for Fredholm's integral equations of second kind and Volterra's integral equations of first kind, and **partial differential equations** for two- and three-dimensional inhomogeneous Helmholtz equation. |
| 3 | Numerical Differentials | Explanation of algorithms, method of using, and usage example of subroutine related to numerical differentials of one-variable functions and multi-variable functions. |
| 4 | Numerical Integration | Explanation of algorithms, method of using, and usage example of subroutine related to numerical integration over a finite interval, semi-infinite interval, fully infinite interval, two-dimensional finite interval, and multi-dimensional finite interval. |
| 5 | Interpolations and Approximations | Explanation of algorithms, method of using, and usage example of subroutine related to interpolations, surface interpolations, least squares approximations, least squares surface approximations, and Chebyshev's approximations. |
| 6 | Spline Functions | Explanation of algorithms, method of using, and usage example of subroutine related to interpolation, smoothing, numerical derivatives, and numerical integrals using cubic splines, bicubic splines and B–splines. |

Basic Functions Volume 5

| Chapter | Title | Contents |
|---|---|---|
| 1 | Introduction | Explanation of the organization of this manual, how to view each item, and usage limitations. |
| 2 | Special Functions | Explanation of algorithms, method of using, and usage example of subroutine related to Bessel functions, modified Bessel functions, spherical Bessel functions, functions related to Bessel functions, Gamma functions, functions related to Gamma functions, elliptic functions, indefinite integrals of elementary functions, associated Legendre functions, orthogonal polynomials, and other special functions. |
| 3 | Sorting and Ranking | Explanation and usage examples of subroutine related to sorting and ranking. |
| 4 | Roots of Equations | Explanation of algorithms, method of using, and usage example of subroutine related to roots of algebraic equations, nonlinear equations, and simultaneous nonlinear equations. |
| 5 | Extremal Problems and Optimization | Explanation of algorithms, method of using, and usage example of subroutine related to minimization of functions with no constraints, minimization of the sum of the squares of functions with no constraints, minimization of one-variable functions with constraints, minimization of multi-variable functions with constraints, and shortest path problem. |

Basic Functions Volume 6

| Chapter | Title | Contents |
|---|---|---|
| 1 | Introduction | Explanation of the organization of this manual, how to view each item, and usage limitations. |
| 2 | Random Number Tests | Explanation and usage examples of subroutine related to uniform random number tests, and distribution random number tests. |
| 3 | Probability Distributions | Explanation and usage examples of subroutine related to continuous distributions and discrete distributions. |
| 4 | Basic Statistics | Explanation and usage examples of subroutine related to basic statistics, variance-covariance and correlation. |
| 5 | Tests and Estimates | Explanation and usage examples of subroutine related to interval estimates and tests. |
| 6 | Analysis of Variance and Design of Experiments | Explanation and usage examples of subroutine related to one-way layout, two-way layout, multiple-way layout, randomized block design, Greco-Latin square method, cumulative Method. |
| 7 | Nonparametric Tests | Explanation and usage examples of subroutine related to tests using $\chi^2$ distribution and tests using other distributions. |
| 8 | Multivariate Analysis | Explanation and usage examples of subroutine related to principal component analysis, factor analysis, canonical correlation analysis, discriminant analysis, cluster analysis. |
| 9 | Time Series Analysis | Explanation and usage examples of subroutine related to autocorrelation, cross correlation, autocovariance, cross covariance, smoothing and demand forecasting. |
| 10 | Regression analysis | Explanation and usage examples of subroutine related to linear Regression and nonlinear Regression. |

Shared Memory Parallel Functions

| Chapter | Title | Contents |
|---|---|---|
| 1 | Introduction | Explanation of the organization of this manual, how to view each item, and usage limitations. |
| 2 | Basic Matrix Algebra | Explanation of algorithms, method of using, and usage example of subroutine related to obtain the product of real matrices and complex matrices. |
| 3 | Simultaneous Linear Equations (Direct Method) | Explanation of algorithms, method of using, and usage example of subroutine related to simultaneous linear equations corresponding to real matrices, complex matrices, real symmetric matrices, and Hermitian matrices. |
| 4 | Simultaneous Linear Equations (Iteration Method) | Explanation of algorithms, method of using, and usage example of subroutine related to simultaneous linear equations corresponding to real positive definite symmetric sparse matrices, real symmetric sparse matrices and real asymmetric sparse matrices. |
| 5 | Eigenvalues and Eigenvectors | Explanation of algorithms, method of using, and usage example of subroutine related to the eigenvalue problem for real symmetric matrices and Hermitian matrices. |
| 6 | Fourier Transforms and their applications | Explanation of algorithms, method of using, and usage example of subroutine related to one-, two- and three-dimensional complex Fourier transforms and real Fourier transforms, two- and three-dimensional convolutions, correlations, and power spectrum analysis. |
| 7 | Sorting | Explanation and usage examples of subroutine related to sorting and ranking. |

Document Version 3.0.0-230301 for ASL, March 2023

**Remarks**

(1) This manual corresponds to ASL 1.1. All functions described in this manual are program products.

(2) Proper nouns such as product names are registered trademarks or trademarks of individual manufacturers.

(3) This library was developed by incorporating the latest numerical computational techniques. Therefore, to keep up with the latest techniques, if a newly added or improved function includes the function of an existing function may be removed.

# Contents

**1 INTRODUCTION** **1**
1.1 OVERVIEW . . . 1
1.1.1 Introduction to The Advanced Scientific Library ASL . . . 1
1.1.2 Distinctive Characteristics of ASL . . . 1
1.2 KINDS OF LIBRARIES . . . 2
1.3 ORGANIZATION . . . 3
1.3.1 Introduction . . . 3
1.3.2 Organization of Subroutine Description . . . 3
1.3.3 Contents of Each Item . . . 3
1.4 SUBROUTINE NAMES . . . 7
1.5 NOTES . . . 9

**2 RANDOM NUMBER TESTS** **10**
2.1 INTRODUCTION . . . 10
2.1.1 Explanation . . . 11
2.1.2 Reference Bibliography . . . 18
2.2 UNIFORM RANDOM NUMBER TESTS . . . 19
2.2.1 DJTEUN, RJTEUN
Uniform Random Number Tests . . . 19
2.3 CONTINUOUS DISTRIBUTION RANDOM NUMBER TESTS . . . 24
2.3.1 DJTENO, RJTENO
Normal Distribution Random Number Test . . . 24
2.3.2 DJTEEX, RJTEEX
Exponential Distribution Random Number Test . . . 28
2.3.3 DJTECC, RJTECC
Cauchy Distribution Random Number Test . . . 31
2.3.4 DJTEGU, RJTEGU
Gumbel Distribution Random Number Test . . . 34
2.3.5 DJTEWE, RJTEWE
Weibull Distribution Random Number Tests . . . 38
2.3.6 DJTEGM, RJTEGM
Gamma Distribution Random Number Test . . . 42
2.3.7 DJTELG, RJTELG
Logistic Distribution Random Number Test . . . 45
2.4 DISCRETE DISTRIBUTION RANDOM NUMBER TESTS . . . 49
2.4.1 RJTEBI
Binomial Distribution Random Number Test . . . 49
2.4.2 RJTENG
Geometric Distribution Random Number Test . . . 52
2.4.3 RJTEPO
Poisson Distribution Random Number Test . . . 55

**3 PROBABILITY DISTRIBUTIONS** **58**
3.1 INTRODUCTION . . . 58
3.1.1 Explanation . . . 60
3.1.2 Reference Bibliography . . . 65
3.2 CONTINUOUS DISTRIBUTIONS . . . 66
3.2.1 D1CDNO, R1CDNO
Normal Distribution . . . 66
3.2.2 D1CDIN, R1CDIN
Inverse of Normal Distribution . . . 69
3.2.3 D1CDBN, R1CDBN
Bivariate Normal Distribution . . . 72
3.2.4 D1CDCH, R1CDCH
$\chi^2$ Distribution . . . 75
3.2.5 D1CDIC, R1CDIC
Inverse of $\chi^2$ Distribution . . . 78
3.2.6 D1CDNC, R1CDNC
Noncentral $\chi^2$ Distribution . . . 81
3.2.7 D1CDIX, R1CDIX
Inverse Noncentral $\chi^2$ Distribution . . . 85
3.2.8 D1CDTB, R1CDTB
$t$ Distribution . . . 88
3.2.9 D1CDIT, R1CDIT
Inverse of $t$ Distribution . . . 91
3.2.10 D1CDNT, R1CDNT
Noncentral $t$ Distribution . . . 94
3.2.11 D1CDIS, R1CDIS
Inverse Noncentral $t$ Distribution . . . 97
3.2.12 D1CDFB, R1CDFB
$F$ Distribution . . . 100
3.2.13 D1CDIF, R1CDIF
Inverse of $F$ Distribution . . . 104
3.2.14 D1CDGM, R1CDGM
Gamma Distribution . . . 107
3.2.15 D1CDIG, R1CDIG
Inverse Gamma Distribution . . . 111
3.2.16 D1CDBT, R1CDBT
Beta Distribution . . . 114
3.2.17 D1CDIB, R1CDIB
Inverse Beta Distribution . . . 117
3.2.18 D1CDUF, R1CDUF
Uniform Distribution . . . 120
3.2.19 D1CDTR, R1CDTR
Triangular Distribution . . . 123
3.2.20 D1CDPA, R1CDPA
Pareto Distribution . . . 126
3.2.21 D1CDWE, R1CDWE
Weibull Distribution . . . 129
3.2.22 D1CDEX, R1CDEX
Exponential Distribution . . . 132
3.2.23 D1CDGU, R1CDGU
Gumbel Distribution . . . 135
3.2.24 D1CDLD, R1CDLD
Logarithmic Distribution . . . 138
3.2.25 D1CDLN, R1CDLN
Log-Normal Distribution . . . 141

3.2.26 D1CDLG, R1CDLG
Logistic Distribution . . . 144
3.2.27 D1CDCC, R1CDCC
Cauchy Distribution . . . 147
3.3 DISCRETE DISTRIBUTIONS . . . 150
3.3.1 D1DDBP, R1DDBP
Binomial Distribution and Negative Binomial Distribution . . . 150
3.3.2 D1DDGO, R1DDGO
Geometric Distribution . . . 154
3.3.3 D1DDPO, R1DDPO
Poisson Distribution . . . 157
3.3.4 D1DDHG, R1DDHG
Hypergeometric Distribution . . . 159
3.3.5 D1DDHN, R1DDHN
Negative Hypergeometric Distribution . . . 162

**4 SAMPLE STATISTICS 166**
4.1 INTRODUCTION . . . 166
4.1.1 Explanation . . . 167
4.1.2 Reference Bibliography . . . 172
4.2 BASIC STATISTICS . . . 173
4.2.1 D2BA1T, R2BA1T
One Sample Basic Statistics . . . 173
4.2.2 D2BA2S, R2BA2S
Two Samples Basic Statistics . . . 179
4.2.3 D2BAMS, R2BAMS
$m$ Samples Basic Statistics . . . 186
4.2.4 D2BAGM, R2BAGM
Geometric Mean . . . 191
4.2.5 D2BAMO, R2BAMO
Moment . . . 195
4.2.6 D2BAHM, R2BAHM
Harmonic Mean . . . 199
4.2.7 D2BASM, R2BASM
Root Mean Square . . . 203
4.3 VARIANCE-COVARIANCE . . . 207
4.3.1 D2VCMT, R2VCMT
Variance-Covariance Matrices . . . 207
4.3.2 D2VCGR, R2VCGR
Variance-Covariance Matrices (Grouped Data) . . . 212
4.4 CORRELATION COEFFICIENTS . . . 220
4.4.1 D2CCMT, R2CCMT
Correlation Matrices . . . 220
4.4.2 D2CCMA, R2CCMA
Multiple Correlation Coefficients . . . 225
4.4.3 D2CCPR, R2CCPR
Partial Correlation Coefficients . . . 231

**5 TIME SERIES ANALYSIS 237**
5.1 INTRODUCTION . . . 237
5.1.1 Explanation . . . 238
5.1.2 Reference Bibliography . . . 242
5.2 AUTOCOVARIANCE AND CROSS COVARIANCE . . . 243
5.2.1 DFCVSC, RFCVSC
Autocovariances . . . 243

5.2.2 DFCVCS, RFCVCS
Cross Covariances . . . 246
5.3 AUTOCORRELATION AND CROSS CORRELATION . . . 250
5.3.1 DFCRSC, RFCRSC
Autocorrelation Coefficients . . . 250
5.3.2 DFCRCZ, RFCRCZ
Cross Correlation Coefficients (Zero Means) . . . 252
5.3.3 DFCRCS, RFCRCS
Cross Correlation Coefficients . . . 254
5.4 SMOOTHING AND DEMAND FORECASTING . . . 256
5.4.1 DFASMA, RFASMA
Moving Averages . . . 256
5.4.2 DFDPES, RFDPES
Single Exponential Smoothing . . . 260
5.4.3 DFDPED, RFDPED
Double Exponential Smoothing . . . 262
5.4.4 DFDPET, RFDPET
Triple Exponential Smoothing . . . 265

**6 TESTS AND ESTIMATES 268**
6.1 INTRODUCTION . . . 268
6.1.1 Explanation . . . 269
6.1.2 Reference Bibliography . . . 289
6.2 INTERVAL ESTIMATES . . . 290
6.2.1 D3IERA, R3IERA
Interval Estimation of the Population Ratio According to One Set of Samples . . . 290
6.2.2 D3IEME, R3IEME
Interval Estimation of the Population Mean According to One Set of Samples . . . 293
6.2.3 D3IESU, R3IESU
Interval Estimation of the Difference of the Population Means According to Two Sets of Independent Samples . . . 297
6.2.4 D3IEVA, R3IEVA
Interval Estimation of the Population Variance Due to One Set of Samples . . . 301
6.2.5 D3IETC, R3IETC
Interval Estimation of the Population Correlation Coefficient According to One Set of Samples 304
6.2.6 D3IECD, R3IECD
Interval Estimation of the Difference of the Population Correlation Coefficient According to Two Sets of Independent Samples . . . 307
6.2.7 D3IESR, R3IESR
Interval Estimation in the Simple Linear Regression . . . 311
6.3 TESTS . . . 317
6.3.1 D3TSRA, R3TSRA
Test of the Population Ratio According to One Set of Samples . . . 317
6.3.2 D3TSRD, R3TSRD
Test of the Difference of the Population Ratios According to Two Sets of Independent Samples 321
6.3.3 D3TSME, R3TSME
Test of the Population Mean According to One Set of Samples . . . 326
6.3.4 D3TSSU, R3TSSU
Test of the Difference of the Population Means According to Two Sets of Independent Samples 331
6.3.5 D3TSVA, R3TSVA
Test of the Population Variance Due to One Set of Samples . . . 338
6.3.6 D3TSTC, R3TSTC
Test of the Population Correlation Coefficient According to One Set of Samples . . . 342

6.3.7 D3TSCD, R3TSCD
Test of the Difference of the Population Correlation Coefficients According to Two Sets of Independent Samples . . . 347
6.3.8 D3TSSR, R3TSSR
Test in the Simple Linear Regression . . . 350

**7 ANALYSIS OF VARIANCE AND DESIGN OF EXPERIMENTS 359**
7.1 INTRODUCTION . . . 359
7.1.1 Explanation . . . 360
7.1.2 Reference Bibliography . . . 362
7.2 ONE-WAY LAYOUT . . . 363
7.2.1 D41WR1, R41WR1
One-Way Layout Analysis of Variance . . . 363
7.3 TWO-WAY LAYOUT . . . 369
7.3.1 D42WRN, R42WRN
Two-Way Layout Analysis of Variance . . . 369
7.3.2 D42WRM, R42WRM
Two-Way Layout Analysis of Variance (With Missing Values) . . . 375
7.3.3 D42WR1, R42WR1
Two-Way Layout Analysis of Variance (Repetition Data) . . . 383
7.4 MULTIPLE-WAY LAYOUT . . . 391
7.4.1 D4MWRF, R4MWRF
Multiple-Way Layout Analysis of Variance . . . 391
7.4.2 D4MWRM, R4MWRM
Multiple-Way Layout Analysis of Variance (With Missing Values) . . . 402
7.5 CUMULATIVE METHOD . . . 415
7.5.1 D4MU01, R4MU01
Analysis of Variance Using the Cumulative Method . . . 415
7.6 RANDOMIZED BLOCK DESIGN . . . 430
7.6.1 D4RB01, R4RB01
Analysis of Variance Using Randomized Block Design . . . 430
7.7 GRECO-LATIN SQUARE METHOD . . . 434
7.7.1 D4GL01, R4GL01
Analysis of Variance Using the Greco-Latin Square Method . . . 434
7.8 BALANCED INCOMPLETE BLOCK DESIGN . . . 438
7.8.1 D4BI01, R4BI01
Analysis of Variance Using Balanced Incomplete Block Design . . . 438

**8 NONPARAMETRIC TESTS 443**
8.1 INTRODUCTION . . . 443
8.1.1 Explanation . . . 444
8.1.2 Reference Bibliography . . . 446
8.2 TESTS USING $\chi^2$ DISTRIBUTION . . . 447
8.2.1 D5CHEF, R5CHEF
Test of Goodness of Fit . . . 447
8.2.2 D5CHTT, R5CHTT
$\chi^2$ Test ($2 \times 2$ Contingency Table) . . . 450
8.2.3 D5CHMN, R5CHMN
$\chi^2$ Test ($m \times n$ Contingency Table) . . . 453
8.2.4 D5CHMD, R5CHMD
Median Test . . . 456
8.3 TESTS USING OTHER DISTRIBUTION . . . 459
8.3.1 D5TESG, R5TESG
Sign Test . . . 459
8.3.2 D5TEWL, R5TEWL
Wilcoxon Test . . . 462

8.3.3 D5TEMH, R5TEMH
Mann-Whitney's U Test . . . 466
8.3.4 D5TESP, R5TESP
Spearman's Rank Correlation Test . . . 470

**9 MULTIVARIATE ANALYSIS** **474**
9.1 INTRODUCTION . . . 474
9.1.1 Explanation . . . 475
9.1.2 Reference Bibliography . . . 481
9.2 PRINCIPAL COMPONENT ANALYSIS . . . 482
9.2.1 D6CPCC, R6CPCC
Principal Component Cumulative Contribution Ratio . . . 482
9.2.2 D6CPSC, R6CPSC
Principal Component Scores . . . 484
9.3 FACTOR ANALYSIS . . . 489
9.3.1 D6FALD, R6FALD
Factor Loading Matrix . . . 489
9.3.2 D6FAVR, R6FAVR
Rotation According to the Varimax Criterion . . . 491
9.4 CANONICAL CORRELATION ANALYSIS . . . 495
9.4.1 D6CVAN, R6CVAN
Canonical Correlation Analysis . . . 495
9.4.2 D6CVSC, R6CVSC
Canonical Variate Scores . . . 498
9.5 DISCRIMINANT ANALYSIS . . . 503
9.5.1 D6DAFN, R6DAFN
Discriminant Functions . . . 503
9.5.2 D6DASC, R6DASC
Discriminant Function Scores . . . 507
9.6 CLUSTER ANALYSIS . . . 513
9.6.1 D6CLDS, R6CLDS
Dissimilarity Measures . . . 513
9.6.2 D6CLAN, R6CLAN
Cluster Analysis (Input Dissimilarity Measure or Similarity Measure) . . . 518
9.6.3 D6CLDA, R6CLDA
Cluster Analysis (Input Observation Data, Use Dissimilarity Measure) . . . 523

**10 REGRESSION ANALYSIS** **529**
10.1 INTRODUCTION . . . 529
10.1.1 Explanation . . . 530
10.1.2 Reference Bibliography . . . 536
10.2 LINEAR REGRESSION ANALYSIS . . . 537
10.2.1 DNLNRG, RNLNRG
Linear Regression Analysis . . . 537
10.2.2 DNLNRR, RNLNRR
Linear Regression Analysis (Repetitive Data) . . . 543
10.2.3 DNLNMA, RNLNMA
Multiple Regression Analysis . . . 550
10.3 NONLINEAR REGRESSION ANALYSIS . . . 555
10.3.1 DNNLPO
Polynomial Regression Analysis . . . 555
10.3.2 DNNLGF, RNNLGF
Regression According to an Arbitrary Function . . . 560

**A METHODS OF HANDLING ARRAY DATA** **567**
A.1 Methods of handling array data corresponding to matrix . . . . 567
A.2 Data storage modes . . . . 568
A.2.1 Real matrix (two-dimensional array type) . . . . 568
A.2.2 Real symmetric matrix and positive symmetric matrix . . . . 569

**B MACHINE CONSTANTS USED IN ASL** **570**
B.1 Units for Determining Error . . . . 570
B.2 Maximum and Minimum Values of Floating Point Data . . . . 570

# Chapter 1

# INTRODUCTION

## 1.1 OVERVIEW

### 1.1.1 Introduction to The Advanced Scientific Library ASL

Table 1−1 shows correspondences among product categories, functions of ASL and supported hardware platforms. In the same version of ASL, interfaces of subroutines of the same name are common among hardware platforms.

Table 1−1 Classification of functions included in ASL

| Classification of Functions | Volume |
|---|---|
| Basic functions | Vol. 1-6 |
| Shared memory parallel functions | Vol. 7 |

### 1.1.2 Distinctive Characteristics of ASL

ASL has the following distinctive characteristics.

(1) Subroutines are optimized using compiler optimization to take advantage of corresponding system hardware features.

(2) Special-purpose subroutines for handling matrices are provided so that the optimum processing can be performed according to the type of matrix (symmetric matrix, Hermitian matrix, or the like). Generally, processing performance can be increased and the amount of required memory can be conserved by using the special-purpose subroutines.

(3) Subroutines are modularized according to processing procedures to improve reliability of each component subroutine as well as the reliability and efficiency of the entire system.

(4) Error information is easy to access after a subroutine has been used since error indicator numbers have been systematically determined.

## 1.2 KINDS OF LIBRARIES

Table 1−2 Kinds of libraries providing ASL

| Size of variable(byte) | | Declaration of arguments | Kind | Kind of library |
|---|---|---|---|---|
| integer | real | | | |
| 4 | 8 | INTEGER(4)<br>REAL(8) | 32bit integer Double-precision subroutine | 32bit integer library<br>(link option: -lasl_sequential) |
| 4 | 4 | INTEGER(4)<br>REAL(4) | 32bit integer Single-precision subroutine | |
| 8 | 8 | INTEGER(8)<br>REAL(8) | 64bit integer Double-precision subroutine | 64bit integer library<br>(link option: -lasl_sequential_i64) |
| 8 | 4 | INTEGER(8)<br>REAL(4) | 64bit integer Single-precision subroutine | |

**(∗1)** Functions that appear in this documentation do not always support all of the four kinds of subroutines listed above. For those functions that do not support some of those subroutine kinds, relevant notes will appear in the corresponding subsections.

**(∗2)** The string "(4)" that specifies 32bit (4 byte) can be omitted.

## 1.3 ORGANIZATION

This section describes the organization of Chapters 2 and later.

### 1.3.1 Introduction

The first section of each chapter is a general introduction describing functions corresponding to each subroutines.

### 1.3.2 Organization of Subroutine Description

The second section of each chapter sequentially describes the following topics for each subroutine.

(1) Function

(2) Usage

(3) Arguments

(4) Restrictions

(5) Error indicator

(6) Notes

(7) Example

Each item is described according to the following principles.

### 1.3.3 Contents of Each Item

(1) **Function**

Function briefly describes the purpose of the ASL subroutine.

(2) **Usage**

Usage describes the subroutine name and the order of its arguments. In general, arguments are arranged as follows.

CALL subroutine-name (input-arguments, input/output-arguments, output-arguments, ISW, work, IERR)

ISW is an input argument for specifying the processing procedure. IERR is an error indicator. In some cases, input/output arguments precede input arguments. The following general principles also apply.

- Array are placed as far to the left as possible according to their importance.
- The dimension of an array immediately follows the array name. If multiple arrays have the same dimension, the dimension is assigned as an argument of only the first array name. It is not assigned as an argument of subsequent array names.

(3) **Arguments**

Arguments are explained in the order described above in paragraph (2). The explanation format is as follows.

| Arguments | Type | Size | Input/Output | Contents |
|---|---|---|---|---|
| (a) | (b) | (c) | (d) | (e) |

(a) Arguments

Arguments are explained in the order they are designated in the Usage paragraph.

(b) Type

Type indicates the data type of the argument. Any of the following codes may appear as the type.

**I** : Integer type

**D** : Double precision real

**R** : Real

**Z** : Double precision complex

**C** : Complex

There are 64-bit integer and 32-bit integer for integer type arguments. In a 32-bit (64-bit) integer type subroutine, all the integer type arguments are 32-bit (64-bit) integer. In other words, kinds of libraries determine the sizes of integer type arguments (Refer to 1.4). In the user program, a 32-bit/64-bit integer type argument must be declared by `INTEGER/ INTEGER(8)`, respectively.

(c) Size

Size indicates the required size of the specified argument. If the size is greater than 1, the required area must be reserved in the program calling this subroutine.

**1** : Indicates that argument is a variable.

N : Indicates that the argument is a vector (one-dimensional array) having N elements. The argument N indicating the size of this vector is defined immediately after the specified vector. However, if the size of a vector or array defined earlier, it is omitted following subsequently defined vectors or arrays. The size may be specified by only a numeric value or in the form of a product or sum such as $3 \times \mathrm{N}$ or $\mathrm{N} + \mathrm{M}$.

M**,** N : Indicates that the argument is a two-dimensional array having M rows and N columns. If M and N indicating the size of this array have not been defined before this array is specified, they are defined as arguments immediately following this array.

(d) Input/Output

Input/Output indicates whether the explanation of argument contents applies to input time or output time.

i. When only "Input" appears

When the control returns to the program using this subroutine, information when the argument is input is preserved. The user must assign input-time information unless specifically instructed otherwise.

ii. When only "Output" appears

Results calculated within the subroutine are output to the argument. No data is entered at input time.

iii. When both "Input" and "Output" appear

Argument contents change between the time control passes to the subroutine and the time control returns from the subroutine. The user must assign input-time information unless specifically instructed otherwise.

iv. When "Work" appears

Work indicates that the argument is an area used when performing calculations within the subroutine. A work area having the specified size must be reserved in the program calling this subroutine. The contents of the work area may have to be maintained so they can be passed along to the next calculation.

(e) Contents

Contents describes information held by the argument at input time or output time.

- A sample Argument description follows.

**Example**

The statement of the subroutine (DBGMLC, RBGMLC) that obtains the LU decomposition and the condition number of a real matrix is as follows.

Double precision:
  CALL DBGMLC (A, LNA, N, IPVT, COND, W1, IERR)
Single precision:
  CALL RBGMLC (A, LNA, N, IPVT, COND, W1, IERR)

The explanation of the arguments is as follows.

Table 1−3 Sample Arguments

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}D\\R\end{array}\right\}^{\text{Note}}$ | LNA, N | Input | Real matrix $A$(two-dimensional array) |
| | | | | Output | The matrix $A$ decomposed into the matrix $LU$ where $U$ is a unit upper triangular matrix and $L$ is a lower triangular matrix. |
| 2 | LNA | I | 1 | Input | Adjustable dimension size of array A |
| 3 | N | I | 1 | Input | Order $n$ of matrix $A$ |
| 4 | IPVT | I | N | Output | Pivoting information<br>IPVT(i): Number of the row exchanged with row $i$ in the $i$-th step. |
| 5 | COND | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Output | Reciprocal of the condition number |
| 6 | W1 | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | N | Work | Work area |
| 7 | IERR | I | 1 | Output | Error indicator |

To use this subroutine, arrays A, IPVT and W1 must first be allocated in the calling program so they can be used as arguments. A is a $\left\{\begin{array}{c}\text{double-precision}\\ \text{single-precision}\end{array}\right\}^{\text{Note}}$ real array of size (LNA , N) , IPVT is an integer array of size N and W1 is a $\left\{\begin{array}{c}\text{double-precision}\\ \text{single-precision}\end{array}\right\}$ real array of size N.
When the 64-bit integer version is used, all integer-type arguments (LNA, N, IPVT and IERR) must be declared by using `INTEGER(8)`, not `INTEGER`.

**Note** The entries enclosed in brace { } mean that the array should be declared double precision type (code D) when using subroutine DBGMLC and real type (code R) when using subroutine RBGMLC. Braces are used in this manner throughout the remainder of the text unless specifically stated otherwise.

Data must be stored in A, LNA and N before this subroutine is called. The LU decomposition and condition number of the assigned matrix are calculated with in the subroutine, and the results are stored in array A and variable COND. In addition, pivoting information is stored in IPVT for use by subsequent subroutines.

IERR is an argument used to notify the user of invalid input data or an error that may occur during processing. If processing terminates normally, IERR is set to zero.

Since W1 is a work area used only within the subroutine, its contents at input and output time have no special meaning.

(4) **Restrictions**

Restrictions indicate limiting ranges for subroutine arguments.

(5) **Error indicator**

Each subroutine has been given an error indicator as an output argument. This error indicator, which has uniformly been given the variable name IERR, is placed at the end of the arguments. If an error is detected within the subroutine, a corresponding value is output to IERR. Error indicator values are divided into five levels.

Table 1−4 Classification of Error Indicator Output Values

| Level | IERR value | Meaning | Processing result |
|---|---|---|---|
| Normal | 0 | Processing is terminated normally. | Results are guaranteed. |
| Warning | 1000∼2999 | Processing is terminated under certain conditions. | Results are conditionally guaranteed. |
| Fatal | 3000∼3499 | Processing is aborted since an argument violated its restrictions. | Results are not guaranteed. |
| | 3500∼3999 | Obtained results did not satisfy a certain condition. | Obtained results are returned (the results are not guaranteed). |
| | 4000 or more | A fatal error was detected during processing. Usually, processing is aborted. | Results are not guaranteed. |

(6) **Notes**

Notes describes ambiguous items and points requiring special attention when using the subroutine.

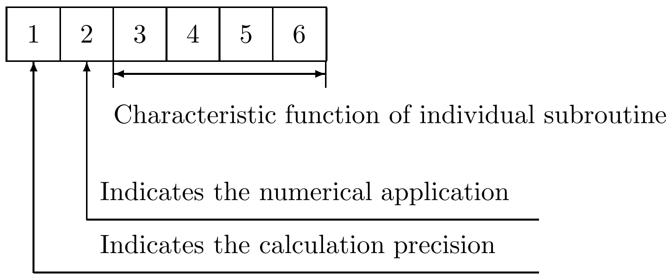
(7) **Example**

Here gives an example of how to use the subroutine. Note that in some cases, multiple subroutines are combined in a single example. The output results are given in the 32-bit integer version, and may differ within the range of rounding error if the compiler or intrinsic functions are different.
The source codes of examples in this document are included in User's Guide. Input data, if required, is also included in it. To build up an executable files by compiling these example source codes, they should be linked with this product library.

## 1.4 SUBROUTINE NAMES

The subroutines name of ASL basic functions consists of ⟨six alphanumeric characters⟩.

Figure 1−1 Subroutine Name Components




**"1" in Figure** 1−1 **:** The following eight letters are used to indicate the calculation precision.

| | |
|---|---|
| D, W | Double precision real-type calculation |
| R, V | Single precision real-type calculation |
| Z, J | Double precision complex-type calculation |
| C, I | Single precision complex-type calculation |

However, the complex type calculations listed above do not necessarily require complex arguments.

**"2" in Figure** 1−1 **:** Currently, the following letters lettererererere are used to indicate the application field in the ASL related products.

| Letter | Application Field | Volume |
|---|---|---|
| A | Storage mode conversion | 1 |
| | Basic matrix algebra | 1, 7 |
| B | Simultaneous linear equations (direct method) | 2, 7 |
| C | Eigenvalues and eigenvectors | 1, 7 |
| F | Fourier transforms and their applications | 3, 7 |
| | Time series analysis | 6 |
| G | Spline function | 4 |
| H | Numeric integration | 4 |
| I | Special function | 5 |
| J | Random number tests | 6 |
| K | Ordinary differential equation (initial value problems) | 4 |
| L | Roots of equations | 5 |
| M | Extremum problems and optimization | 5 |
| N | Approximation and regression analysis | 4, 6 |
| O | Ordinary differential equations (boundary value problems), integral equations and partial differential equations | 4 |
| P | Interpolation | 4 |
| Q | Numerical differentials | 4 |
| S | Sorting and ranking | 5, 7 |

| Letter | Application Field | Volume |
|---|---|---|
| X | Basic matrix algebra | 1 |
| | Simultaneous linear equations (iterative method) | 7 |
| 1 | Probability distributions | 6 |
| 2 | Basic statics | 6 |
| 3 | Tests and estimates | 6 |
| 4 | Analysis of variance and design of experiments | 6 |
| 5 | Nonparametric tests | 6 |
| 6 | Multivariate analysis | 6 |

**"3–6" in Figure** 1−1 **:** These characters indicate the characteristic function of the individual subroutine.

## 1.5 NOTES

(1) Use the subroutines of double precision version whenever possible. They not only provide higher precision solutions but also are more stable than single precision versions, in particular, for eigenvalue and eigenvector problems.

(2) To suppress compiler operation exceptions, ASL subroutines are set to so that they conform to the compiler parameter indications of a user's main program. Therefore, the main program must suppress any operation exceptions.

(3) The numerical calculation programs generally deal with operations on finite numbers of digits, so the precision of the results cannot exceed the number of operation digits being handled. For example, since the number of operation digits (in the mantissa part) for double-precision operations is on the order of 15 decimal digits, when using these floating point modes to calculate a value that mathematically becomes 1, an error on the order of $10^{-15}$ may be introduced at any time. Of course, if multiple length arithmetic is emulated such as when performing operations on an arbitrary number of digits, this kind of error can be controlled. However, in this case, when constants such as $\pi$ or function approximation constants, which are fixed in double-precision operations, for example, are also to be subject to calculations that depend on the length of the multiple length arithmetic operations, the calculation efficiency will be worse than for normal operations.

(4) A solution cannot be obtained for a problem for which no solution exists mathematically. For example, a solution of simultaneous linear equations having a singular (or nearly singular) matrix for its coefficient matrix theoretically cannot be obtained with good precision mathematically. Numerical calculations cannot strictly distinguish between mathematically singular and nearly singular matrices. Of course, it is always possible to consider a matrix to be singular if the calculation value for the condition number is greater than or equal to an established criterion value.

(5) Generally, if data is assigned that causes a floating point exception during calculations (such as a floating point overflow), a normal calculation result cannot be expected. However, a floating point underflow that occurs when adding residuals in an iterative calculation is an exception to this.

(6) For problems that are handled using numerical calculations (specifically, problems that use iterative techniques as the calculation method), there are cases in which a solution cannot be obtained with good precision and cases in which no solution can be obtained at all, by a special-purpose subroutine.

(7) Depending on the problem being dealt with, there may be cases when there are multiple solutions, and the execution result differs in appearance according to the compiler used or the computer or OS under which the program is executed. For example, when an eigenvalue problem is solved, the eigenvectors that are obtained may differ in appearance in this way.

(8) The mark "DEPRECATED" denotes that the subroutine will be removed in the future. Use **ASL Unified Interface**, the higher performance alternative practice instead.

## Chapter 2

# RANDOM NUMBER TESTS

## 2.1 INTRODUCTION

This chapter describes the subroutines for testing random numbers.

The subroutine for testing uniform random numbers can perform the following kinds of tests.

(1) Frequency one–dimensional test

(2) Frequency two–dimensional test

(3) Frequency three–dimensional test

(4) Run (ascending/descending) test

(5) Run (above/below) test

(6) Combination test

(7) Gap test

In addition, there are subroutines that perform frequency one–dimensional tests on each type of distribution random numbers.

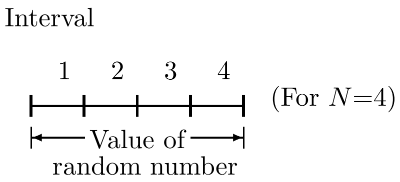
### 2.1.1 Explanation

(1) **Frequency one–dimensional test of uniform random numbers**

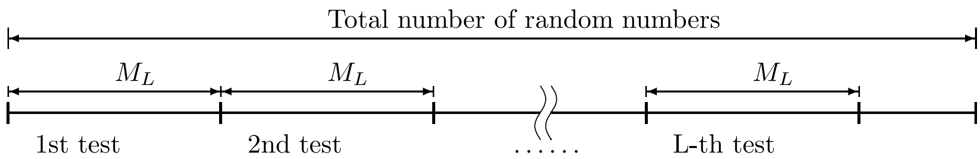
This is a frequency one–dimensional test of uniform random numbers in the range 0.0 through 1.0 that obtains the $\chi^2$ value $X2$ shown below for each test, where $N$ is the number of subdivisions and $M_L$ is the number of random numbers used in a single test.

$$X2 = \sum_{i=1}^{N} \frac{(f_{P_i} - f_{T_i})^2}{f_{P_i}}$$

$f_{T_i}$ : Total number of random numbers in interval $i$ when the interval from 0.0 through 1.0 is subdivided into $N$ equal parts.




The following figure shows which random numbers are used for each test.




$f_{P_i}$ : $f_{P_i} = \dfrac{M_L}{N}$

The test is considered be passed if the following condition holds:
$X2 < \chi^2$ distribution percentile for the entered significance level

(2) **Frequency two–dimensional test of uniform random numbers**

This is a frequency two–dimensional test of uniform random numbers in the range 0.0 through 1.0 that obtains the $\chi^2$ value $X2$ shown below for each test, where $N$ is the number of subdivisions and $M_L$ is the number of random numbers used in a single test.

$$X2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{(f_{P_{ij}} - f_{T_{ij}})^2}{f_{P_{ij}}}$$

$f_{T_{ij}}$ : If the random numbers $U_t, U_{t+1}$ are distributed as two–dimensional points in the X and Y coordinate system within a lattice created by subdividing the intervals from 0.0 through 1.0 on the X and Y axis into $N$ equal parts, then $f_{T_{ij}}$ is the number of points within the corresponding lattice sector (See Fig. 2−1,2−2).

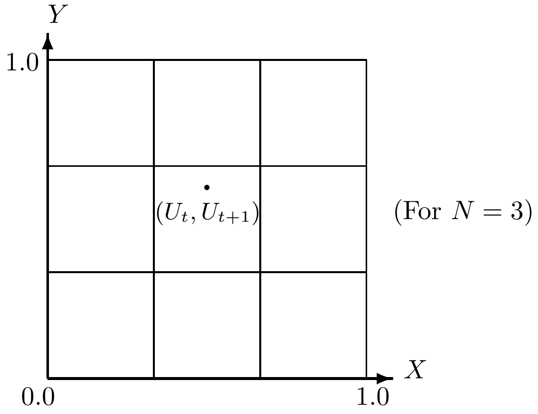



Figure 2−1

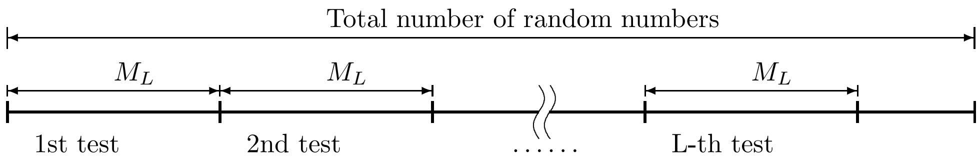



Figure 2−2

$$f_{P_{ij}} \quad : \quad f_{P_{ij}} = \frac{M_L}{N^2}$$

The test is considered to be passed if the following condition holds:
$X2 < \chi^2$ distribution percentile for the entered significance level

(3) **Frequency three–dimensional test of uniform random numbers**

This is a frequency three–dimensional test of uniform random numbers in the range 0.0 through 1.0 that obtains the $\chi^2$ value $X2$ shown below for each test, where $N$ is the number of subdivisions and $M_L$ is the number of random numbers used in a single test.

$$X2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} \frac{(f_{P_{ijk}} - f_{T_{ijk}})^2}{f_{P_{ijk}}}$$

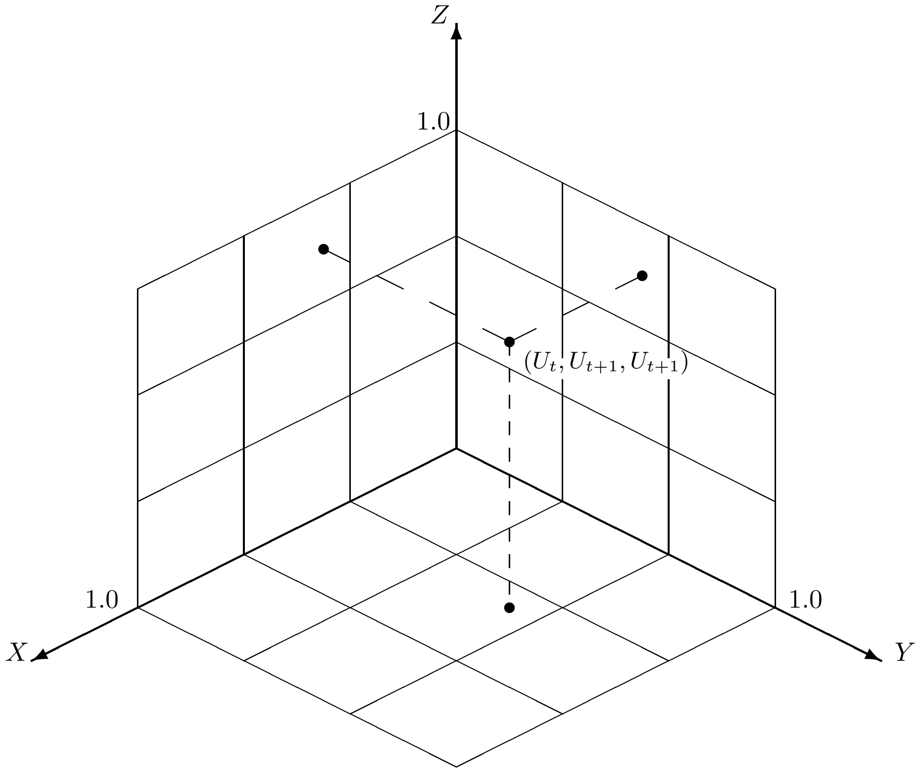
$f_{T_{ijk}}$ : If the random numbers $U_t, U_{t+1}, U_{t+2}$ are distributed as three–dimensional points in the X, Y and Z coordinate system within a lattice created by subdividing the intervals from 0.0 through 1.0 on the X, Y and Z axis into $N$ equal parts, then $f_{T_{ijk}}$is the number of points within the corresponding lattice sector.(See Fig. 2−3 and 2−4)

$$f_{P_{ijk}} \quad : \quad f_{P_{ijk}} = \frac{M_L}{N^3}$$

The test is considered to be passed if the following condition holds:
$X2 < \chi^2$ distribution percentile for the entered significance level

(4) **Run (ascending/descending) test of uniform random numbers**

This is a run (ascending/descending) test of uniform random numbers in the range 0.0 through 1.0 that




(For $N = 3$)

Figure 2−3

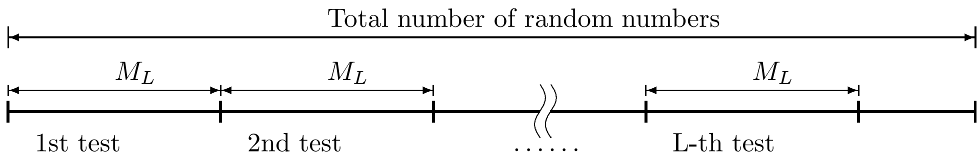



Figure 2−4

obtains the $\chi^2$ value $X2$ shown below for each test, where $N$ is the maximum run length and $M_L$ is the number of random numbers used in a single test.

$$X2 = \sum_{i=1}^{N} \frac{(f_{P_i} - f_{T_i})^2}{f_{P_i}}$$

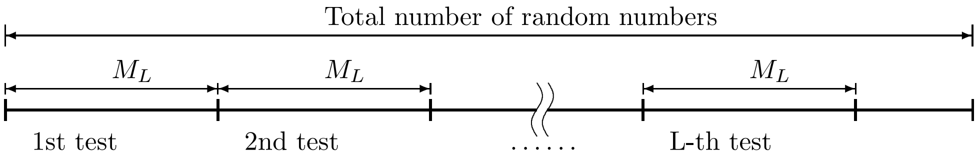
$f_{T_i}$ : Number of ascending or descending runs of length $i$ within the random number sequence.
(However, for $i = N, f_{T_i}$ is the number of ascending runs of length at least $N$.)

Example of an ascending run of length $i$

$$U_{t-1} > U_t < U_{t+1} < \cdots < U_{t+i} > U_{t+i+1}$$

The following figure shows which random numbers are used for each test.

$f_{P_i}$ : $f_{P_i} = f'_{P_i}$ for $1 \leq i \leq N-1$




$$f_{P_N} = \sum_{k=N}^{M_L-1} f'_{P_k}$$

$$f'_{P_i} = 2 \times M_L \times \frac{i^2 + 3 \times i + 1}{(i+3)!} - 2 \times \frac{i^3 + 3 \times i^2 - i - 4}{(i+3)!}$$

To improve test precision, this subroutine uses a modified value for $f_{P_i}$ defined by the following expression.

$$f^*_{P_i} \quad = \quad f_{P_i} \times \frac{\sum_{i=1}^{N} f_{T_i}}{\sum_{i=1}^{N} f_{P_i}}$$

The test is considered to be passed if the following condition holds:
$X2 < \chi^2$ distribution percentile for the entered significance level

(5) **Run (above/below) test of uniform random numbers**
This is a run (above/below) test of uniform random numbers in the range 0.0 through 1.0 that obtains the $\chi^2$ value $X2$ shown below for each test, where $N$ is the maximum run length and $M_L$ is the number of random numbers used in a single test.

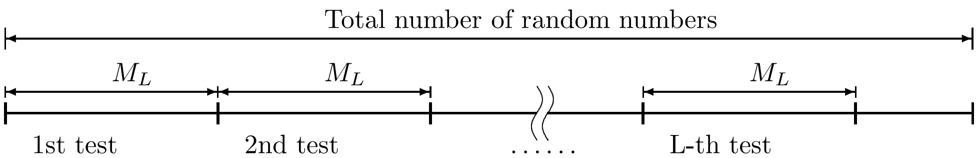
$$X2 \quad = \quad \sum_{i=1}^{N} \frac{(f_{P_i} - f_{T_i})^2}{f_{P_i}}$$

$f_{T_i}$ : Number of runs above 0.5 of length $i$ within the random number sequence.
(However, for $i = N, f_{T_i}$ is the number of runs above 0.5 of length at least $N$.)

Example of a run above 0.5 of length 4

0.3, 0.6, 0.9, 0.7, 0.8, 0.1

The following figure shows which random numbers are used for each test.




$f_{P_i}$ : $f_{P_i} = f'_{P_i}$ for $1 \leq i \leq N-1$

$$f_{P_N} = \sum_{k=N}^{M_L} f'_{P_k}$$

$$f'_{P_i} = \frac{M_L - i + 3}{2^{i+1}}$$

To improve test precision, this subroutine uses a modified value for $f_{P_i}$ defined by the following expression.

$$f^*_{P_i} = f_{P_i} \times \frac{\sum_{i=1}^{N} f_{T_i}}{\sum_{i=1}^{N} f_{P_i}}$$

The test is considered to be passed if the following condition holds:
$X2 < \chi^2$ distribution percentile for the entered significance level

(6) **Combination test of uniform random numbers**

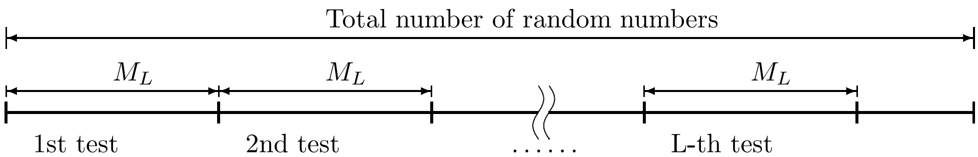
This is a combination test of uniform random numbers in the range 0.0 through 1.0. Conventionally, a combination test checks the number of "0" and "1" bits within the bit pattern of a single random number. However, in order to handle real random numbers, this subroutine collects the register–length number of random numbers and tests how many of them are at least 0.5. The subroutine obtains the $\chi^2$ value $X2$ shown below for each test.

$$X2 = \sum_{i=0}^{N_B} \frac{(f_{P_i} - f_{T_i})^2}{f_{P_i}}$$

where $N_B$, which is the register length, is 32.

$f_{T_i}$ : Number of groups of random numbers taken $N_B$ at a time for which $i$ numbers are at least 0.5.

The following figure shows which random numbers are used for each test.




$$f_{P_i} \ : \ f_{P_i} = \binom{N_B}{i} \times \left(\frac{1}{2}\right)^{N_B} \times \sum_{i=1}^{N_B} f_{T_i} \times \left[\frac{M_L}{N_B}\right]$$

The actual subdivision is as follows.

$$i = 0 \sim 8, 9, 10, \cdots, 22, 23, 24 \sim 32$$

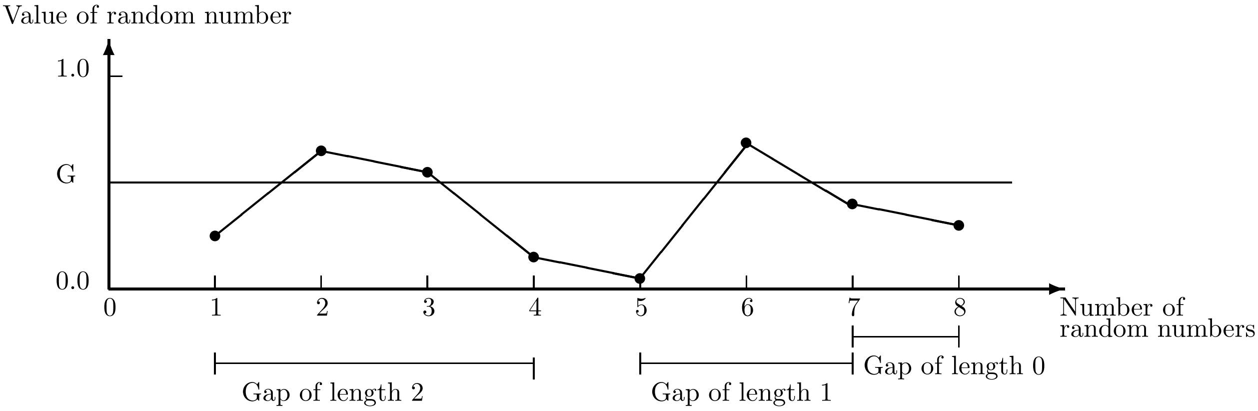
The test is considered to be passed if the following condition holds:
$X2 < \chi^2$ distribution percentile for the entered significance level

(7) **Gap test**

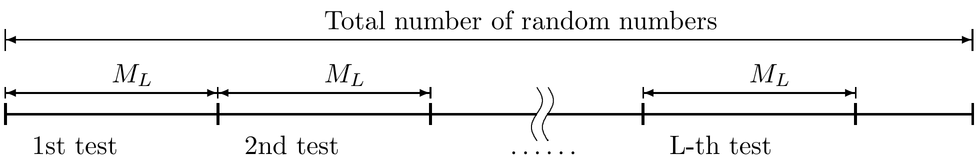
This is a gap test of uniform random numbers in the range 0.0 through 1.0 that obtains the $\chi^2$ value $X2$ shown below for each test, where $N$ is the maximum gap length, $G$ is the gap value, and $M_L$ is the number of random numbers used in a single test.

$$X2 = \sum_{i=0}^{N} \frac{(f_{P_i} - f_{T_i})^2}{f_{P_i}}$$

$f_{T_i}$ : Number of gaps of length $i$ in which sequences of random number that fall within the range 0.0 through $G$ occur.




The following figure shows which random numbers are used for each test.




$$f_{P_i} \quad : \quad f_{P_i} = f'_{P_i} \quad \text{for} \quad 0 \le i \le N-1$$

$$f_{P_N} = \sum_{k=N}^{M_L-1} f'_{P_k}$$

$$f'_{P_i} = G \times (1.0 - G)^i \times M_L$$

To improve test precision, this subroutine uses a modified value for $f_{P_i}$ defined by the following expression.

$$f^*_{P_i} = f_{P_i} \times \frac{\sum_{i=0}^{N} f_{T_i}}{\sum_{i=0}^{N} f_{P_i}}$$

The test is considered to be passed if the following condition holds:
$X2 < \chi^2$ distribution percentile for the entered significance level

(8) **Tests of distribution random numbers**
A frequency one–dimensional test of distribution random numbers is performed.

(a) For a Continuous distribution
The test obtains the $\chi^2$ value $X2$ shown below for each test, where $U_L$ and $U_P$ are the lower and upper limits of the test interval, $N$ is the number of subdivisions, and $M_L$ is the number of random numbers used in a single test.

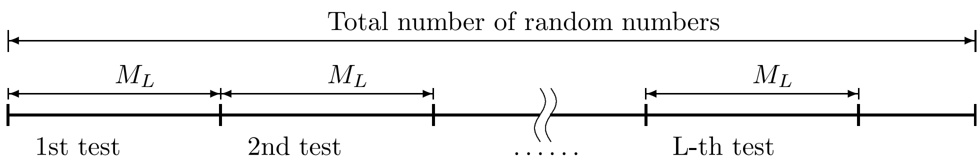
$$X2 = \sum_{i=1}^{N} \frac{(f_{P_i} - f_{T_i})^2}{f_{P_i}}$$

$U_L$ and $U_P$ are determined as follows depending on the type of distribution random numbers.

| | Normal distribution<br>Cauchy distribution<br>Gumbel distribution<br>Gamma distribution | Exponential distribution<br>Weibull distribution |
|---|---|---|
| $U_P$ | Entered test interval upper limit | Entered test interval upper limit |
| $U_L$ | Entered test interval lower limit | 0.0 |

$f_{T_i}$ : Number of random numbers in interval $i$ when the interval from $U_L$ through $U_P$ is subdivided into $N$ equal parts. (Random numbers having values less than or equal to $U_L$ are counted in interval 1 and those having values greater than or equal to $U_P$ are counted in interval $N$.)

The following figure shows which random numbers are used in each test.




$f_{P_i}$ : Expected frequency of random numbers in interval $i$.

If there is an expected frequency that a random number value will be less than or equal to $U_L$ or greater than or equal to $U_P$, that expected frequency is added to the expected frequency of interval 1 or $N$, respectively.

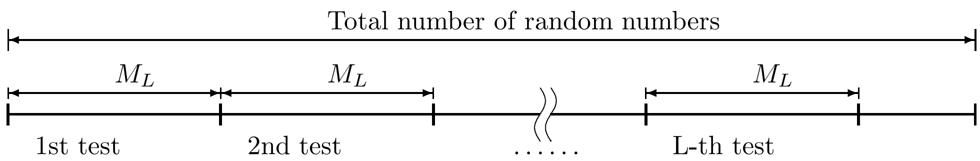
(b) For a Discrete distribution

The test obtains the $\chi^2$ value $X2$ shown below for each test, where the integer $U_P$ is the upper limit of the test interval and $M_L$ is the number of random numbers used in a single test.

$$X2 = \sum_{i=0}^{U_P} \frac{(f_{P_i} - f_{T_i})^2}{f_{P_i}}$$

$f_{T_i}$ : Number of random numbers having value $i$. (Random numbers having values greater than $U_P$ are counted in numbercorresponding to $U_P$.)

The following figure shows which random numbers are used in each test.




$f_{P_i}$ : Expected frequency of random numbers having value $i$. (If there is an expected frequency that a random number value will be greater than $U_P$, then that expected frequency is added to the expected frequency of $U_P$.)

### 2.1.2 Reference Bibliography

(1) Knuth, D. E. , "The Art of Computer Programming Volume II (2nd ed.) Seminumerical Algorithms/Random Numbers", Addison-Wesley (1981).

(2) Heringa, J. R. , Blote, H. W. J. , Compagner, A. , Int. J. Mod. Phys. C 3, 561 (1992)

(3) Kirpatrick, S. , Stoll, E. P. , "A Very Fast Shift–Register Sequence Random Number Generator", Journal of Computational Physics, Vol.40, pp.517–526, (1981).

## 2.2 UNIFORM RANDOM NUMBER TESTS

### 2.2.1 DJTEUN, RJTEUN
### Uniform Random Number Tests

(1) **Function**

Tests the given uniform random numbers in the range 0.0 through 1.0.

(2) **Usage**

Double precision:

CALL DJTEUN (U, M, LT, N, G, ALF, K, X2, CX, ISW, WK, IERR)

Single precision:

CALL RJTEUN (U, M, LT, N, G, ALF, K, X2, CX, ISW, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: {INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer}

R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | U | {D / R} | M | Input | Random numbers |
| 2 | M | I | 1 | Input | Total number of random numbers |
| 3 | LT | I | 1 | Input | Test iteration count (See Note (a)) |
| 4 | N | I | 1 | Input | ISW=1, 2 or 3: Number of subdivisions<br>ISW=4 or 5: Maximum run length<br>ISW=6: Not used<br>ISW=7: Maximum gap length<br>(See Note (b)) |
| 5 | G | {D / R} | 1 | Input | Gap value (See Note (c)) |
| 6 | ALF | {D / R} | 1 | Input | Significance (%) |
| 7 | K | I | 1 | Output | Number of passed tests |
| 8 | X2 | {D / R} | LT | Output | Test result $\chi^2$ value |
| 9 | CX | {D / R} | 1 | Output | $\chi^2$ value for significance level |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 10 | ISW | I | 1 | Input | Processing switch<br>ISW=1: Frequency one-dimensional test<br>ISW=2: Frequency two-dimensional test<br>ISW=3: Frequency three-dimensional test<br>ISW=4: Run (ascending/descending) test<br>ISW=5: Run (above/below) test<br>ISW=6: Combination test<br>ISW=7: Gap test |
| 11 | WK | $\begin{Bmatrix} \mathrm{D} \\ \mathrm{R} \end{Bmatrix}$ | See Contents | Work | Work area<br>**Size**:<br>ISW=1: N<br>ISW=2: $\mathrm{N}^2$<br>ISW=3: $\mathrm{N}^3$<br>ISW=4 or 5: $2 \times \mathrm{N}$<br>ISW=6: 1<br>ISW=7: $2 \times (\mathrm{N}+1)$ |
| 12 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $1 \leq \mathrm{ISW} \leq 7$

(b) $\mathrm{LT} \geq 1$

(c) $0.0 < \mathrm{ALF} < 100.0$

(d) $\mathrm{M} \geq \mathrm{LT}$

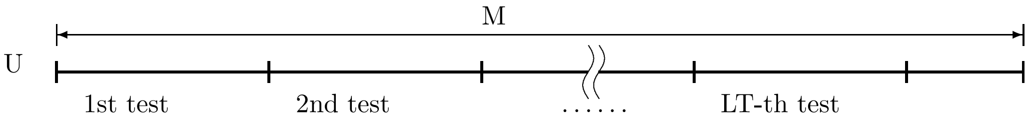
(e) For $\mathrm{ISW} = 1 \sim 5$: $\mathrm{N} \geq 2$
For $\mathrm{ISW} = 6$: $\mathrm{M} \geq 32 \times \mathrm{LT}$
For $\mathrm{ISW} = 7$: $\mathrm{N} \geq 2$ and $0.0 < \mathrm{G} < 1.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | When $\mathrm{ISW} \neq 6$, N was too large or M was too small. (See Note (b)) | Processing continues. (Test precision gets worse.) |
| 3000 | Restriction (a), (b), (c), (d) or (e) was not satisfied. | Processing is aborted. |
| 4000 | $\mathrm{U(i)} < 0.0$ or $\mathrm{U(i)} > 1.0$, $\mathrm{i} = 1, \cdots, \mathrm{M}$. | |

(6) **Notes**

(a) The following figure shows the relationship between the total number of random numbers M, the test iteration count LT, and the random numbers U used for each test.




$\lfloor M/LT \rfloor$ random numbers U are taken at a time for each test, where $\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.

(b) For ISW = 1, 2 or 3:

If the total number of random numbers M is too small or if the number of subdivisions N is too large, the expected frequency of random numbers in each interval decreases and test precision worsens.

Generally, the expected frequency $F_{TN}$ should satisfy the following condition:

$$F_{TN} \geq 5.0$$

If this condition is not satisfied in these subroutines, then IERR=1000.

$F_{TN}$ is calculated from the expressions shown below for each test, where $\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.

Frequency one-dimensional test (ISW=1):

$$F_{TN} = \lfloor M/LT \rfloor \times \frac{1.0}{N}$$

Frequency two-dimensional test (ISW=2):

$$F_{TN} = \lfloor M/LT \rfloor \times \frac{1.0}{2.0 \times N^2}$$

Frequency three-dimensional test (ISW=3):

$$F_{TN} = \lfloor M/LT \rfloor \times \frac{1.0}{3.0 \times N^3}$$

For ISW=4, 5 or 7:

If the total number of random numbers M is too small or if the maximum run length N or maximum gap length N is too large, the expected frequency for which the length is N decreases and test precision worsens.

Generally, the expected frequency $F_{TN}$ should satisfy the following condition:

$$F_{TN} \geq 5.0$$

If this condition is not satisfied in these subroutines, then IERR = 1000. Expressions for calculating N, which is used as the test criterion, and its values are shown below for each test, where $\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.

For a run (ascending/descending) test (ISW=4)

$$N = \lfloor 1 + \text{ LOG10 } \lfloor M/LT \rfloor \rfloor$$

For example:

| $\lfloor M/LT \rfloor$ | 100 | 1, 000 | 10, 000 | 100, 000 | 1, 000, 000 |
|---|---|---|---|---|---|
| N | 3 | 4 | 5 | 6 | 7 |

For a run (above/below) test (ISW = 5)

$$\mathrm{N} = \left\lfloor \frac{\mathrm{LOG10}(\frac{\lfloor \mathrm{M/LT} \rfloor}{10.0})}{\mathrm{LOG10(2)}} \right\rfloor$$

For example:

| ⌊M/LT⌋ | 100 | 1, 000 | 10, 000 | 100, 000 | 1, 000, 000 |
|---|---|---|---|---|---|
| N | 3 | 6 | 9 | 13 | 16 |

For a gap test (ISW = 7)

$$\mathrm{N} = \left\lfloor \frac{\mathrm{LOG10}\ (\frac{5.0}{\mathrm{G} \times \lfloor \mathrm{M/LT} \rfloor})}{\mathrm{LOG10}(1.0 - \mathrm{G})} \right\rfloor$$

For example, when G = 0.1:

| ⌊M/LT⌋ | 100 | 1, 000 | 10, 000 | 100, 000 | 1, 000, 000 |
|---|---|---|---|---|---|
| N | 6 | 28 | 50 | 72 | 93 |

(c) The gap value is used only when performing a gap test.
For explanation for gap value, see Section 2.1.1 Explanation (7).

(d) See the arguments table for the WK size.

### (7) **Example**

(a) Problem

Perform a frequency one-dimensional test on 1000 uniform random numbers.

(b) Main program

```
      PROGRAM BJTEUN
! *** EXAMPLE OF DJTEUN ***
      IMPLICIT REAL(8) (A-H,O-Z)
      PARAMETER ( M=1000, L=10   , ISW=1 , N=10 )
      DIMENSION  U (M)  , X2(L)  , WK(N)
!
      WRITE(6,1000)
      IX = 1
      IY = 1
!
      CALL DJUFSP  ( M,IX,IY,U,IERR )
!
      ALF = 1.0D0
      WRITE(6,1100) M,L,N,ALF,ISW
!
      CALL DJTEUN  ( U,M,L,N,G,ALF,K,X2,CX,ISW,WK,IERR )
!
      WRITE(6,1200) IERR
      WRITE(6,1300) K
      WRITE(6,1400) (X2(I),I=1,L)
      WRITE(6,1500) CX
!
      STOP
 1000 FORMAT(' ',/,/,' *** DJTEUN ***',/)
 1100 FORMAT('    **  INPUT  **',/,10X,'  M = ',I5,5X,'  L = ',   1&
      I5,/,10X,'  N = ',I5,5X,' ALF= ',F5.1,/,10X,'ISW = ',I5,/)
 1200 FORMAT('    **  OUTPUT  **',/,/,10X,' IERR = ',I5,/)
 1300 FORMAT(10X,' NUMBER OF PASSED TEST (K) = ',I3,/)
 1400 FORMAT(10X,'TEST NO.     1       2       3       4       5   ',&
                '    6       7       8       9      10   ',/,/,&
              8X,'CHI-SQUARE',/,8X,'VALUE (X2)',10F8.1,/)
 1500 FORMAT(10X,' CHI-SQUARE VALUE FOR PERCENT POINT (CX) = ',F8.1)
      END
```

(c) Output results

```
*** DJTEUN ***

   **  INPUT  **
          M =  1000       L =    10
          N =    10      ALF=   1.0
        ISW =     1

   **  OUTPUT  **
```

```
   IERR =     0

   NUMBER OF PASSED TEST (K) =  10

  TEST NO.     1       2       3       4       5       6       7       8       9      10

CHI-SQUARE
VALUE (X2)     7.4    14.8     6.4     9.4     7.2     6.0    10.4     1.8    13.2    10.2

   CHI-SQUARE VALUE FOR PERCENT POINT (CX) =     21.7
```

## 2.3 CONTINUOUS DISTRIBUTION RANDOM NUMBER TESTS

### 2.3.1 DJTENO, RJTENO<br>Normal Distribution Random Number Test

(1) **Function**

Performs a frequency one-dimensional test on normal distribution random numbers.

(2) **Usage**

Double precision:

CALL DJTENO (U, M, LT, N, ALF, UL, UP, AM, SG, K, X2, CX, WK, IERR)

Single precision:

CALL RJTENO (U, M, LT, N, ALF, UL, UP, AM, SG, K, X2, CX, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: {INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer}

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | U | {D / R} | M | Input | Random numbers |
| 2 | M | I | 1 | Input | Total number of random numbers |
| 3 | LT | I | 1 | Input | Test iteration count (See Note (a)) |
| 4 | N | I | 1 | Input | Number of subdivisions |
| 5 | ALF | {D / R} | 1 | Input | Significance level (%) |
| 6 | UL | {D / R} | 1 | Input | Test interval lower limit (See Note (b)) |
| 7 | UP | {D / R} | 1 | Input | Test interval upper limit (See Note (b)) |
| 8 | AM | {D / R} | 1 | Input | Mean value |
| 9 | SG | {D / R} | 1 | Input | Standard deviation |
| 10 | K | I | 1 | Output | Number of passed tests |
| 11 | X2 | {D / R} | LT | Output | Test result $\chi^2$ value |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 12 | CX | {D, R} | 1 | Output | $\chi^2$ value for significance level |
| 13 | WK | {D, R} | $2 \times$ N | Work | Work area |
| 14 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) M $\geq$ LT

(b) LT $\geq 1$

(c) N $\geq 2$

(d) UP $>$ UL

(e) $0.0 <$ ALF $< 100.0$

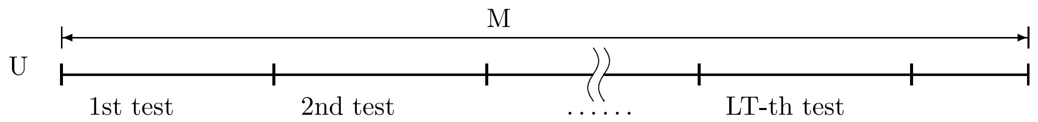
(f) SG $> 0.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | UL was too small, UP was too large, or M was too small. (See Note (b)) | Processing continues. (Test precision gets worse.) |
| 3000 | Restriction (a), (b), (c), (d), (e) or (f) was not satisfied. | Processing is aborted. |

(6) **Notes**

(a) The following figure shows the relationship between the total number of random numbers M, the test iteration count LT, and the random numbers U used for each test.




$\lfloor$M/LT$\rfloor$ random numbers U are taken at a time for each test, where $\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.

(b) If the test range lower limit UL is too small or if the upper limit UP is too large, then an extremely small expected frequency range is tested and test precision worsens.
If $F_{Ti}$ is the expected frequency in each partial interval, then IERR = 1000 if the following condition occurs in the subroutine:

$$F_{Ti} < 5 \qquad (i = 1, \cdots, N)$$

Expressions for calculating UL and UP, which are used as test criteria, are shown below.

$$\begin{aligned} \mathrm{UL} &= \mathrm{AM} - \mathrm{SG} \times \mathrm{D} \\ \mathrm{UP} &= \mathrm{AM} + \mathrm{SG} \times \mathrm{D} \end{aligned}$$

where D is obtained from the following equation.

$$\mathrm{D} \times \mathrm{e}^{-\frac{\mathrm{D}^2}{2}} = \frac{5 \times \mathrm{N}}{\lfloor \mathrm{M/LT} \rfloor} \times \sqrt{\frac{\pi}{2}}$$

$\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.
Sample values of D are shown below.

| $\frac{\mathrm{N}}{\lfloor \mathrm{M/LT} \rfloor}$ | 0.1 | 0.01 | 0.001 | 0.0001 | 0.00001 |
|---|---|---|---|---|---|
| D | 1.5 | 2.5 | 3.5 | 4.0 | 4.5 |

(7) **Example**

(a) Problem
Perform the test 10 times with 10 subdivisions on 10000 normal distribution random numbers having mean 0.0 and standard deviation 1.0.

(b) Main program

```
      PROGRAM BJTENO
! *** EXAMPLE OF DJTENO ***
      IMPLICIT REAL(8) (A-H,O-Z)
      PARAMETER ( M=10000, L=10  , N=10   )
      DIMENSION   U(M) , X2(L) , WK(N,2)
!
      WRITE(6,1000)
      IX = 1
      IY = 1
      AM = 0.0D0
      SG = 1.0D0
!
      CALL DJDBNO ( M,AM,SG,IX,IY,U,IERR )
!
      ALF = 1.0D0
      UL = -2.5D0
      UP =  2.5D0
      WRITE(6,1100)  M,L,N,ALF,UL,UP,AM,SG
!
      CALL DJTENO ( U,M,L,N,ALF,UL,UP,AM,SG,K,X2,CX,WK,IERR )
!
      WRITE(6,1200) IERR
      WRITE(6,1300) K
      WRITE(6,1400) (X2(I),I=1,L)
      WRITE(6,1500) CX
!
      STOP
 1000 FORMAT(' ',/,/,' *** DJTENO ***',/)
 1100 FORMAT('    **  INPUT  **',/,10X,'  M = ',I5,5X,'  L = ',I5,/,&
      10X,'  N = ',I5,5X,' ALF= ',F5.1,/,10X,' UL = ',F5.1,5X,&
      ' UP = ',F5.1,/,10X,' AM = ',F5.1,5X,' SG = ',F5.1,/)
 1200 FORMAT('    **  OUTPUT  **',/,/,10X,' IERR = ',I5,/)
 1300 FORMAT(10X,' NUMBER OF PASSED TEST (K) = ',I3,/)
 1400 FORMAT(10X,'TEST NO.     1       2       3       4       5   ',&
                 '    6       7       8       9      10   ',/,/,&
             8X,'CHI-SQUARE',/,8X,'VALUE (X2)',10F8.1,/)
 1500 FORMAT(10X,' CHI-SQUARE VALUE FOR PERCENT POINT (CX) = ',F8.1)
      END
```

(c) Output results

```
*** DJTENO ***

   **  INPUT  **
           M = 10000        L =    10
           N =    10      ALF=   1.0
          UL =  -2.5      UU =   2.5
          AM =   0.0      SG =   1.0

   **  OUTPUT  **

          IERR =     0

          NUMBER OF PASSED TEST (K) =  10
```

| TEST NO. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| CHI-SQUARE VALUE (X2) | 7.6 | 8.1 | 7.0 | 13.4 | 9.0 | 12.8 | 7.4 | 10.1 | 16.4 | 8.2 |

```
CHI-SQUARE VALUE FOR PERCENT POINT (CX) =     21.7
```

### 2.3.2 DJTEEX, RJTEEX<br>Exponential Distribution Random Number Test

(1) **Function**

Performs a frequency one-dimensional test on exponential distribution random numbers.

(2) **Usage**

Double precision:

CALL DJTEEX (U, M, LT, N, ALF, UP, AM, K, X2, CX, WK, IERR)

Single precision:

CALL RJTEEX (U, M, LT, N, ALF, UP, AM, K, X2, CX, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\\text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | U | $\begin{Bmatrix}D\\R\end{Bmatrix}$ | M | Input | Random numbers |
| 2 | M | I | 1 | Input | Total number of random numbers |
| 3 | LT | I | 1 | Input | Test iteration count (See Note (a).) |
| 4 | N | I | 1 | Input | Number of subdivisions |
| 5 | ALF | $\begin{Bmatrix}D\\R\end{Bmatrix}$ | 1 | Input | Significance level (%) |
| 6 | UP | $\begin{Bmatrix}D\\R\end{Bmatrix}$ | 1 | Input | Test interval upper limit (See Note (b)) |
| 7 | AM | $\begin{Bmatrix}D\\R\end{Bmatrix}$ | 1 | Input | Mean value |
| 8 | K | I | 1 | Output | Number of passed tests |
| 9 | X2 | $\begin{Bmatrix}D\\R\end{Bmatrix}$ | LT | Output | Test result $\chi^2$ value |
| 10 | CX | $\begin{Bmatrix}D\\R\end{Bmatrix}$ | 1 | Output | $\chi^2$ value for significance level |
| 11 | WK | $\begin{Bmatrix}D\\R\end{Bmatrix}$ | $2 \times N$ | Work | Work area |
| 12 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $\mathrm{M} \geq \mathrm{LT}$

(b) $\mathrm{LT} \geq 1$

(c) $\mathrm{N} \geq 2$

(d) $0.0 < \mathrm{ALF} < 100.0$

(e) $\mathrm{AM} > 0.0$

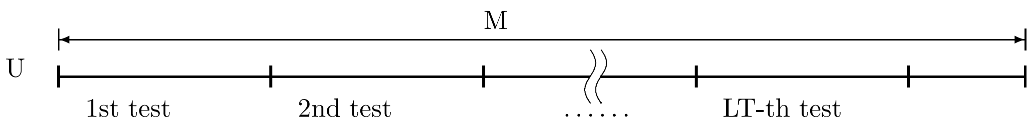
(f) $\mathrm{UP} > 0.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | UP was too large or M was too small. (See Note (b)) | Processing continues. (Test precision gets worse.) |
| 3000 | Restriction (a), (b), (c), (d), (e) or (f) was not satisfied. | Processing is aborted. |
| 4000 | $\mathrm{U(i)} < 0.0\,;\ \mathrm{i} = 1, \cdots, \mathrm{M}$ | |

(6) **Notes**

(a) The following figure shows the relationship between the total number of random numbers M, the test iteration count LT, and the random numbers U used for each test.




$\lfloor \mathrm{M/LT} \rfloor$ random numbers U are taken at a time for each test, where $\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.

(b) If the test range upper limit UP is too large, then an extremely small expected frequency range is tested and test precision worsens.
If $\mathrm{F_{Ti}}$ is the expected frequency in each partial interval, then IERR = 1000 if the following condition occurs in this subroutine:

$$\mathrm{F_{Ti}} < 5 \qquad (\mathrm{i} = 1, \cdots, \mathrm{N})$$

The value of UP, which is used as a test criterion, is determined so that the following condition is satisfied.

$$\mathrm{UP} \times \mathrm{e}^{-\frac{\mathrm{UP}}{\mathrm{AM}}} = \frac{5 \times \mathrm{N} \times \mathrm{AM}}{\lfloor \mathrm{M/LT} \rfloor}$$

$\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.
Sample values of UP are shown below for AM=1.0.

| $\frac{\mathrm{N}}{\lfloor \mathrm{M/LT} \rfloor}$ | 0.01 | 0.001 | 0.0001 | 0.00001 |
|---|---|---|---|---|
| UP | 4 | 7 | 9 | 12 |

(7) **Example**

(a) Problem

Perform the test 10 times with subdivisions on 10000 exponential distribution random numbers having mean value 1.0.

(b) Main program

```
      PROGRAM BJTEEX
! *** EXAMPLE OF DJTEEX ***
      IMPLICIT REAL(8) (A-H,O-Z)
      PARAMETER ( M=10000, L=10   , N=10   )
      DIMENSION   U(M) , X2(L) , WK(N,2)
!
      WRITE(6,1000)
      IX = 1
      IY = 1
      AM = 1.0D0
!
      CALL DJDBEX (M,AM,IX,IY,U,IERR)
!
      ALF = 1.0D0
      UP =  4.0D0
      WRITE(6,1100)  M,L,N,ALF,UP,AM
!
      CALL DJTEEX ( U,M,L,N,ALF,UP,AM,K,X2,CX,WK,IERR )
!
      WRITE(6,1200) IERR
      WRITE(6,1300) K
      WRITE(6,1400) (X2(I),I=1,L)
      WRITE(6,1500) CX
      STOP
 1000 FORMAT(' ',/,/,' *** DJTEEX ***',/)
 1100 FORMAT('    **  INPUT  **',/,10X,'  M = ',I5,5X,'  L = ',I5,/,&
      10X,'  N = ',I5,5X,' ALF= ',F5.1,/,10X,' UP = ',F5.3,5X,&
      ' AM = ',F5.1,/)
 1200 FORMAT('    **  OUTPUT  **',/,/,10X,' IERR = ',I5,/)
 1300 FORMAT(10X,' NUMBER OF PASSED TEST (K) = ',I3,/)
 1400 FORMAT(10X,'TEST NO.     1       2       3       4       5   ',&
                 '    6       7       8       9      10   ',/,/,&
             8X,'CHI-SQUARE',/,8X,'VALUE (X2)',10F8.1,/)
 1500 FORMAT(10X,' CHI-SQUARE VALUE FOR PERCENT POINT (CX) = ',F8.1)
      END
```

(c) Output results

```
*** DJTEEX ***

   **  INPUT  **
          M = 10000       L =    10
          N =    10      ALF=   1.0
         UP = 4.000      AM =   1.0

   **  OUTPUT  **

         IERR =     0

         NUMBER OF PASSED TEST (K) =   9

        TEST NO.     1       2       3       4       5       6       7       8       9      10

      CHI-SQUARE
      VALUE (X2)    16.0     6.3    16.3     6.0     7.4    22.4     8.6     9.5     6.5     7.4

         CHI-SQUARE VALUE FOR PERCENT POINT (CX) =     21.7
```

### 2.3.3 DJTECC, RJTECC Cauchy Distribution Random Number Test

(1) **Function**

Performs a frequency one-dimensional test on Cauchy distribution random numbers.

(2) **Usage**

Double precision:

CALL DJTECC (U, M, LT, N, ALF, UL, UP, A, B, K, X2, CX, WK, IERR)

Single precision:

CALL RJTECC (U, M, LT, N, ALF, UL, UP, A, B, K, X2, CX, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{ \begin{array}{l} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{array} \right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | U | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | M | Input | Random numbers |
| 2 | M | I | 1 | Input | Total number of random numbers |
| 3 | LT | I | 1 | Input | Test iteration count (See Note (a)) |
| 4 | N | I | 1 | Input | Number of subdivisions |
| 5 | ALF | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Significance level (%) |
| 6 | UL | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Test interval lower limit (See Note (b)) |
| 7 | UP | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Test interval upper limit (See Note (b)) |
| 8 | A | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Value of the location parameter $\alpha$. |
| 9 | B | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Value of the scale parameter $\beta$. |
| 10 | K | I | 1 | Output | Number of passed tests |
| 11 | X2 | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | LT | Output | Test result $\chi^2$ value |
| 12 | CX | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Output | $\chi^2$ value for significance level |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 13 | WK | $\left\{\begin{matrix}\mathrm{D}\\\mathrm{R}\end{matrix}\right\}$ | $2 \times \mathrm{N} + 4$ | Work | Work area |
| 14 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $\mathrm{M} \geq \mathrm{LT}$

(b) $\mathrm{LT} \geq 1$

(c) $\mathrm{N} \geq 2$

(d) $\mathrm{UP} > \mathrm{UL}$

(e) $0.0 < \mathrm{ALF} < 100.0$

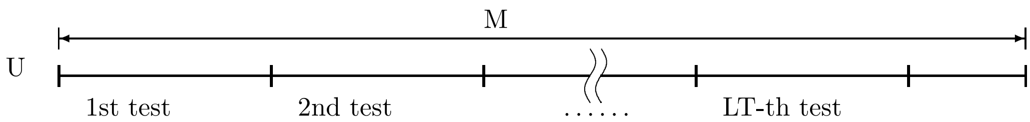
(f) $\mathrm{BX} > 0.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | UL was too small, UP was too large, or M was too small. (See Note (b)) | Processing continues. (Test precision gets worse.) |
| 3000 | Restriction (a), (b), (c), (d), (e) or (f) was not satisfied. | Processing is aborted. |

(6) **Notes**

(a) The following figure shows the relationship between the total number of random numbers M, the test iteration count LT, and the random numbers U used for each test.




$\lfloor \mathrm{M}/\mathrm{LT} \rfloor$ random numbers U are taken at a time for each test, where $\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.

(b) If the test range lower limit UL is too small or if the upper limit UP is too large, then an extremely small expected frequency range is tested and test precision worsens.
If $\mathrm{F}_{\mathrm{Ti}}$ is the expected frequency in each partial interval, then IERR = 1000 if the following condition occurs in the subroutine:

$$\mathrm{F}_{\mathrm{Ti}} < 5 \qquad (\mathrm{i} = 1, \cdots, \mathrm{N})$$

Expressions for calculating UL and UP, which are used as test criteria, are shown below.

$$\mathrm{UL} = \mathrm{AM} - \mathrm{SG} \times \mathrm{D}$$
$$\mathrm{UP} = \mathrm{AM} + \mathrm{SG} \times \mathrm{D}$$

where D is obtained from the following equation.

$$\mathrm{D} \times \mathrm{e}^{-\frac{\mathrm{D}^2}{2}} = \frac{5 \times \mathrm{N}}{\lfloor \mathrm{M/LT} \rfloor} \times \sqrt{\frac{\pi}{2}}$$

$\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.
Sample values of D are shown below.

| $\frac{\mathrm{N}}{\lfloor \mathrm{M/LT} \rfloor}$ | 0.1 | 0.01 | 0.001 | 0.0001 | 0.00001 |
|---|---|---|---|---|---|
| D | 1.5 | 2.5 | 3.5 | 4.0 | 4.5 |

(7) **Example**

(a) Problem

Perform the test 10 times with 10 subdivisions on 10000 normal distribution random numbers with parameters $\alpha = 0.0$ and $\beta = 1.0$.

(b) Main program

```
      PROGRAM BJTECC
! *** EXAMPLE OF DJTECC ***
      IMPLICIT REAL(8) (A-H,O-Z)
      PARAMETER ( M=10000, L=10   , N=10   )
      DIMENSION   U(M) , X2(L) , WK(0:N+1,2)
!
      WRITE(6,1000)
      IX = 1
      IY = 1
      A = 0.0D0
      B = 1.0D0
!
      CALL DJDBCC ( M,A,B,IX,IY,U,IERR )
!
      ALF = 1.0D0
      UL = -2.5D0
      UP =  2.0D0
      WRITE(6,1100)  M,L,N,ALF,UL,UP,A,B
!
      CALL DJTECC ( U,M,L,N,ALF,UL,UP,A,B,K,X2,CX,WK,IERR )
!
      WRITE(6,1200) IERR
      WRITE(6,1300) K
      WRITE(6,1400) (X2(I),I=1,L)
      WRITE(6,1500) CX
!
      STOP
 1000 FORMAT(5X,'*** DJTECC ***',/)
 1100 FORMAT(7X,'**  INPUT  **',/,12X,'M = ',I5,7X,'L = ',I5,/,&
      12X,'N = ',I5,6X,'ALF= ',F5.1,/,11X,'UL = ',F5.1,6X,&
      'UP = ',F5.1,/,12X,'A = ',F5.1,7X,'B = ',F5.1,/)
 1200 FORMAT(7X,'**  OUTPUT  **',/,/,10X,' IERR = ',I5,/)
 1300 FORMAT(11X,'NUMBER OF PASSED TEST (K) = ',I3,/)
 1400 FORMAT(10X,'TEST NO.     1       2       3       4       5   ',&
                 '    6       7       8       9      10   ',/,/,&
             8X,'CHI-SQUARE',/,8X,'VALUE (X2)',10F8.1,/)
 1500 FORMAT(11X,'CHI-SQUARE VALUE FOR PERCENT POINT (CX) = ',F8.1)
      END
```

(c) Output results

```
     *** DJTECC ***

       **  INPUT  **
            M = 10000       L =    10
            N =    10      ALF=   1.0
           UL =  -2.5      UU =   2.0
            A =   0.0       B =   1.0

       **  OUTPUT  **

           IERR =     0

           NUMBER OF PASSED TEST (K) =  10

          TEST NO.     1       2       3       4       5       6       7       8       9      10

        CHI-SQUARE
        VALUE (X2)     9.4    12.6     9.7     7.4     7.5     7.8    15.9     4.9     4.8     3.7

           CHI-SQUARE VALUE FOR PERCENT POINT (CX) =     21.7
```

### 2.3.4 DJTEGU, RJTEGU Gumbel Distribution Random Number Test

(1) **Function**

Performs a frequency one-dimensional test on Gumbel distribution random numbers.

(2) **Usage**

Double precision:

CALL DJTEGU (U, M, LT, N, ALF, UL, UP, A, B, K, X2, CX, WK, IERR)

Single precision:

CALL RJTEGU (U, M, LT, N, ALF, UL, UP, A, B, K, X2, CX, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: {INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer}
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | U | {D/R} | M | Input | Random numbers |
| 2 | M | I | 1 | Input | Total number of random numbers |
| 3 | LT | I | 1 | Input | Test iteration count (See Note (a)) |
| 4 | N | I | 1 | Input | Number of subdivisions |
| 5 | ALF | {D/R} | 1 | Input | Significance level (%) |
| 6 | UL | {D/R} | 1 | Input | Test interval lower limit (See Note (b)) |
| 7 | UP | {D/R} | 1 | Input | Test interval upper limit (See Note (b)) |
| 8 | A | {D/R} | 1 | Input | Location parameter |
| 9 | B | {D/R} | 1 | Input | Scale parameter |
| 10 | K | I | 1 | Output | Number of passed tests |
| 11 | X2 | {D/R} | LT | Output | Test result $\chi^2$ value |
| 12 | CX | {D/R} | 1 | Output | $\chi^2$ value for significance level |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 13 | WK | $\left\{\begin{matrix}\mathrm{D}\\\mathrm{R}\end{matrix}\right\}$ | $2 \times \mathrm{N}$ | Work | Work area |
| 14 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $\mathrm{M} \geq \mathrm{LT}$

(b) $\mathrm{LT} \geq 1$

(c) $\mathrm{N} \geq 2$

(d) $\mathrm{UP} > \mathrm{UL}$

(e) $0.0 < \mathrm{ALF} < 100.0$

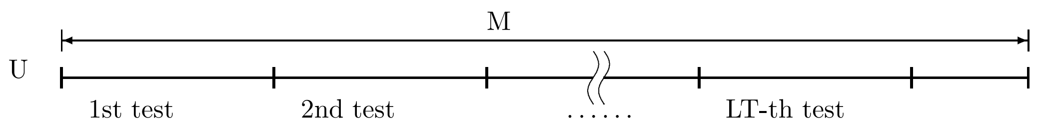
(f) $\mathrm{B} > 0.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | UL was too small, UP was too large, or M was too small. (See Note (b)) | Processing continues. (Test precision gets worse.) |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 3040 | Restriction (e) was not satisfied. | |
| 3050 | Restriction (f) was not satisfied. | |

(6) **Notes**

(a) The following figure shows the relationship between the total number of random numbers M, the test iteration count LT, and the random numbers U used for each test.




$\lfloor \mathrm{M/LT} \rfloor$ random numbers U are taken at a time for each test, where $\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.

(b) If the test range lower limit UL is too small or if the upper limit UP is too large, then an extremely small expected frequency range is tested and test precision worsens.
If $\mathrm{F_{Ti}}$ is the expected frequency in each partial interval, then IERR = 1000 if the following condition occurs in the subroutine:

$$\mathrm{F_{Ti}} < 5 \qquad (\mathrm{i} = 1, \cdots, \mathrm{N})$$

Inequalities for calculating UL and UP, which are used as test criteria, are shown below.

$$\min\left\{\exp\left(\frac{\mathrm{UP}-\mathrm{A}}{\mathrm{B}}\right)\exp\left\lfloor-\exp\left(\frac{\mathrm{UP}-\mathrm{A}}{\mathrm{B}}\right)\right\rfloor,\exp\left(\frac{\mathrm{UL}-\mathrm{A}}{\mathrm{B}}\right)\exp\left\lfloor-\exp\left(\frac{\mathrm{UL}-\mathrm{A}}{\mathrm{B}}\right)\right\rfloor\right\}$$
$$\geq 5\times\frac{\mathrm{B}\times\mathrm{N}}{\lfloor\mathrm{M}/\mathrm{LT}\rfloor\times(\mathrm{UP}-\mathrm{UL})}$$

$\lfloor x\rfloor$ represents the maximum integer that does not exceed $x$.

(7) **Example**

(a) Problem

Perform the test 10 times with 10 subdivisions on 10000 Gumbel distribution random numbers having location parameter 0.0 and scale parameter 1.0.

(b) Main program

```
      PROGRAM BJTEGU
! *** EXEMPLE OF DJTEGU ***
      IMPLICIT REAL(8) (A-H,O-Z)
      PARAMETER ( NMAX=100000 )
      PARAMETER ( DONE = 1.0D0, DMONE = -1.0D0, DZERO = 0.0D0)
      PARAMETER ( DFIVE = 5.0D0, DM100 = -100.0D0, D100 = 100.0D0 )
      PARAMETER ( DM27 = -2.7D0, D27 = 2.7D0 )
      DIMENSION RR(NMAX),X2(NMAX),WK(2*NMAX+4)
!
      N  = 10000
      XA = DONE
      XB = DONE
      IX = 1
      IY = 1
!
      L    = 10
      NDIV = 10
      ALF  = DFIVE
      UL   = DM27
      UP   = D27
!
      CALL DJDBGU(N,XA,XB,IX,IY,RR,KERR)
      IF( KERR .GT. 0 ) THEN
         WRITE(6,6000) KERR
      ELSE
         WRITE(6,6010) N,L,NDIV,ALF,UL,UP,XA,XB
      ENDIF
!
      CALL DJTEGU(RR,N,L,NDIV,ALF,UL,UP,XA,XB,K,X2,CX,WK,IERR)
      WRITE(6,6020) IERR
      IF( IERR .EQ. 0 ) THEN
         WRITE(6,6030) K
         WRITE(6,6040) (X2(I),I=1,L)
         WRITE(6,6050) CX
      ENDIF
      STOP
!
 6000 FORMAT(1X,'KERR = ',I4)
 6010 FORMAT(/,&
      1X,'     ** INPUT **',/,/,&
      1X,'       N    = ',I10,'       L    = ',I10,/,&
      1X,'       NDIV = ',I10,'       ALF  = ',F10.3,/,&
      1X,'       UL   = ',F10.3,'       UP   = ',F10.3,/,&
      1X,'       XA   = ',F10.3,'       XB   = ',F10.3)
 6020 FORMAT(/,&
      1X,'     ** OUTPUT **',/,/,&
      1X,'       IERR = ',I4)
 6030 FORMAT(/,&
      1X,'       NUMBER OF PASSED TEST (K) = ',I3,/)
 6040 FORMAT(&
      1X,'       TEST NO.     1     2     3     4     5',&
         '     6     7     8     9    10',/,/,&
      1X,'       CHI-SQUARE',/,&
      1X,'       VALUE (X2)',10F6.1,/)
 6050 FORMAT(/,&
      1X,'       CHI-SQUARE VALUE FOR PERCENT POINT (CX) = ',F8.1)
      END
```

(c) Output results

```
     ** INPUT **

       N    =      10000       L    =         10
       NDIV =         10       ALF  =      5.000
       UL   =     -2.700       UP   =      2.700
       XA   =      1.000       XB   =      1.000

     ** OUTPUT **
```

```
IERR =    0

NUMBER OF PASSED TEST (K) =  10

TEST NO.      1     2     3     4     5     6     7     8     9    10

CHI-SQUARE
VALUE (X2)   7.1   8.2  10.4   8.0   9.7  14.7  12.1   4.0   4.7  10.9

CHI-SQUARE VALUE FOR PERCENT POINT (CX) =     16.9
```

## 2.3.5 DJTEWE, RJTEWE Weibull Distribution Random Number Tests

(1) **Function**

Performs a frequency one-dimensional test on Weibull distribution random numbers.

(2) **Usage**

Double precision:

CALL DJTEWE (U, M, LT, N, ALF, UP, A, B, K, X2, CX, WK, IERR)

Single precision:

CALL RJTEWE (U, M, LT, N, ALF, UP, A, B, K, X2, CX, WK, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | U | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | M | Input | Random numbers |
| 2 | M | I | 1 | Input | Total number of random numbers |
| 3 | LT | I | 1 | Input | Test iteration count (See Note (a)) |
| 4 | N | I | 1 | Input | Number of subdivisions |
| 5 | ALF | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Significance level (%) |
| 6 | UP | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Test interval upper limit (See Note (b)) |
| 7 | A | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Shape parameter $a$ |
| 8 | B | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Scale parameter $b$ |
| 9 | K | I | 1 | Output | Number of passed tests |
| 10 | X2 | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | LT | Output | Test result $\chi^2$ value |
| 11 | CX | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Output | $\chi^2$ value for significance level |
| 12 | WK | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | $2 \times N$ | Work | Work area |
| 13 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $\mathrm{M} \geq \mathrm{LT}$

(b) $\mathrm{LT} \geq 1$

(c) $\mathrm{N} \geq 2$

(d) $0.0 < \mathrm{ALF} < 100.0$

(e) $\mathrm{UP} > 0.0$

(f) $\mathrm{A} > 0.0$

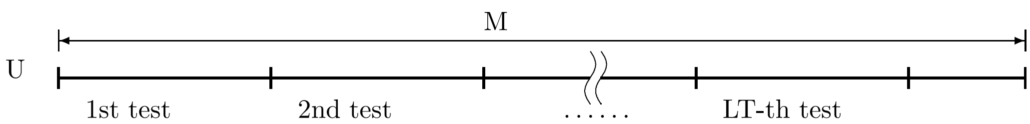
(g) $\mathrm{B} > 0.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | UP was too large or M was too small. (See Note (b)) | Processing continues. (Test precision gets worse.) |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 3040 | Restriction (e) was not satisfied. | |
| 3050 | Restriction (f) was not satisfied. | |
| 3060 | Restriction (g) was not satisfied. | |

(6) **Notes**

(a) The following figure shows the relationship between the total number of random numbers M, the test iteration count LT, and the random numbers U used for each test.




$\lfloor \mathrm{M}/\mathrm{LT} \rfloor$ random numbers U are taken at a time for each test, where $\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.

(b) If the test range upper limit UP is too large, then an extremely small expected frequency range is tested and test precision worsens.
If $\mathrm{F_{Ti}}$ is the expected frequency in each partial interval, then $\mathrm{IERR} = 1000$ if the following condition occurs in this subroutine:

$$\mathrm{F_{Ti}} < 5 \qquad (\mathrm{i} = 1, \cdots, \mathrm{N})$$

The value of UP, which is used as a test criterion, is determined so that the following condition is satisfied.

$$\mathrm{UP} \times \mathrm{e}^{-(\frac{\mathrm{UP}}{\mathrm{B}})^{\mathrm{A}}} = \frac{5 \times \mathrm{N}}{\lfloor \mathrm{M}/\mathrm{LT} \rfloor}$$

$\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.
Sample values of UP are shown below for A=2.0 and B=1.0 .

| $\frac{N}{\lfloor M/LT \rfloor}$ | 0.01 | 0.001 | 0.0001 | 0.00001 |
|---|---|---|---|---|
| UP | 1.9 | 2.5 | 2.9 | 3.3 |

(7) **Example**

(a) Problem

Perform the test 10 times with subdivisions on 10000 Weibull distribution random numbers having the shape parameter 2.0 and the scale parameter 1.0.

(b) Main program

```
      PROGRAM BJTEWE
! *** EXEMPLE OF DJTEWE ***
      IMPLICIT REAL(8) (A-H,O-Z)
      PARAMETER ( NMAX=100000 )
      PARAMETER ( D0 = 0.0D0, D5 = 5.0D0, D100 = 100.0D0 )
      PARAMETER ( D1 = 1.0D0, D2 = 2.0D0, D3 = 3.0D0 )
      DIMENSION RR(NMAX),X2(NMAX),WK(2*NMAX+4)
!
      N  = 10000
      XA = D2
      XB = D1
      IX = 1
      IY = 1
!
      L    = 10
      NDIV = 10
      ALF  = D5
      UP   = 1.9
!
      CALL DJDBWE (N,XA,XB,IX,IY,RR,KERR)
      IF( KERR .GT. 0 ) THEN
         WRITE(6,6000) KERR
      ELSE
         WRITE(6,6010) N,L,NDIV,ALF,UP,XA,XB
      ENDIF
!
      CALL DJTEWE(RR,N,L,NDIV,ALF,UP,XA,XB,K,X2,CX,WK,IERR)
      WRITE(6,6020) IERR
      IF( IERR .EQ. 0 ) THEN
         WRITE(6,6030) K
         WRITE(6,6040) (X2(I),I=1,L)
         WRITE(6,6050) CX
      ENDIF
      STOP
!
 6000 FORMAT(1X,'KERR = ',I4)
 6010 FORMAT(1X,/,&
      1X,'  ** INPUT **',/,/,&
      1X,'    N    = ',I10,'       L    = ',I10,/,&
      1X,'    NDIV = ',I10,'       ALF  = ',F10.3,/,&
      1X,'    UP   = ',F10.3,/,&
      1X,'    XA   = ',F10.3,'       XB   = ',F10.3)
 6020 FORMAT(1X,/,&
      1X,'  ** OUTPUT **',/,/,&
      1X,'    IERR = ',I4)
 6030 FORMAT(1X,/,&
      1X,'    NUMBER OF PASSED TEST (K) = ',I3,/)
 6040 FORMAT(&
      1X,'    TEST NO.      1     2     3     4     5',&
         '     6     7     8     9    10',/,/,&
      1X,'    CHI-SQUARE',/,&
      1X,'    VALUE (X2)',10F6.1,/)
 6050 FORMAT(1X,/,1X,&
      '    CHI-SQUARE VALUE FOR PERCENT POINT (CX) = ',F8.1)
      END
```

(c) Output results

```
 ** INPUT **

   N    =      10000       L    =         10
   NDIV =         10       ALF  =      5.000
   UP   =      1.900
   XA   =      2.000       XB   =      1.000

 ** OUTPUT **

   IERR =    0

   NUMBER OF PASSED TEST (K) =   9
```

```
TEST NO.       1      2      3      4      5      6      7      8      9     10

CHI-SQUARE
VALUE (X2)  10.3   13.5   12.5    9.4   12.3   17.6   10.1    5.1    7.2   11.7

CHI-SQUARE VALUE FOR PERCENT POINT (CX) =     16.9
```

### 2.3.6 DJTEGM, RJTEGM Gamma Distribution Random Number Test

(1) **Function**

Performs a frequency one-dimensional test on gamma distribution random numbers.

(2) **Usage**

Double precision:

CALL DJTEGM (U, M, LT, NDIV, ALF, UL, UP, GAMALP, K, X2, CX, WK, IERR)

Single precision:

CALL RJTEGM (U, M, LT, NDIV, ALF, UL, UP, GAMALP, K, X2, CX, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\\text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | U | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | M | Input | Random numbers |
| 2 | M | I | 1 | Input | Total number of random numbers |
| 3 | LT | I | 1 | Input | Test iteration count (See Note (a))) |
| 4 | NDIV | I | 1 | Input | Number of subdivisions |
| 5 | ALF | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Significance level (%) |
| 6 | UL | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Test interval lower limit (See Note (b)) |
| 7 | UP | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Test interval upper limit (See Note (b)) |
| 8 | GAMALP | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Shape parameter for gamma distribution |
| 9 | K | I | 1 | Output | Number of passed tests |
| 10 | X2 | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | LT | Output | Test result $\chi^2$ value |
| 11 | CX | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Output | $\chi^2$ value for significance level |
| 12 | WK | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | See Contents | Work | Work area<br>**Size**: $2 \times \text{NDIV} + 4$ |
| 13 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $\mathrm{M} \geq \mathrm{LT}$

(b) $\mathrm{LT} \geq 1$

(c) $\mathrm{NDIV} \geq 2$

(d) $\mathrm{UP} > \mathrm{UL}$

(e) $0.0 < \mathrm{ALF} < 100.0$

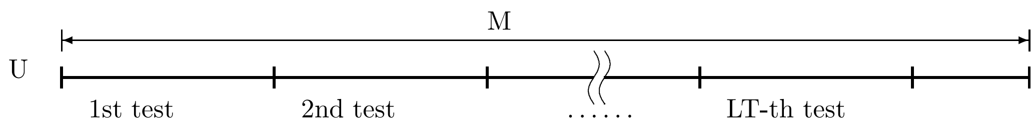
(f) $\mathrm{GAMALP} > 0.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | UL was too small, UP was too large, or M was too small. (See Note (b)) | Processing continues. (Test precision gets worse.) |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 3040 | Restriction (e) was not satisfied. | |
| 3050 | Restriction (f) was not satisfied. | |

(6) **Notes**

(a) The following figure shows the relationship between the total number of random numbers M, the test iteration count LT, and the random numbers U used for each test.




$\lfloor \mathrm{M/LT} \rfloor$ random numbers U are taken at a time for each test, where $\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.

(b) If the test range lower limit UL is too small or if the upper limit UP is too large, then an extremely small expected frequency range is tested and test precision worsens.
If $\mathrm{F_{Ti}}$ is the expected frequency in each partial interval, then IERR = 1000 if the following condition occurs in the subroutine:

$$\mathrm{F_{Ti}} < 5 \qquad (\mathrm{i} = 1, \cdots, \mathrm{NDIV})$$

(7) **Example**

(a) Problem
Perform the test 10 times with 10 subdivisions on 10000 gamma distribution random numbers having the shape parameter $\alpha = 3.0$.

(b) Input data
M = 10000, L = 10, NDIV = 10, ALF = 5.0, UL = 0.0, UP = 9.0 and GAMALP = 3.0.

(c) Main program

```
      PROGRAM BJTEGM
! *** EXEMPLE OF DJTEGM ***
      IMPLICIT REAL(8) (A-H,O-Z)
      PARAMETER ( N=10000 )
      PARAMETER ( DZERO = 0.0D0, D3 = 3.0D0, D5 = 5.0D0, D9 = 9.0D0 )
      DIMENSION RR(N),X2(N),WK(2*N+4)
!
      IX = 1
      IY = 1
!
      L    = 10
      NDIV = 10
      ALF  = D5
!
      UL   = DZERO
!
      GAMALP = D3
      UP   = D9
!
      WRITE(6,5000) N,GAMALP
      CALL DJDBGM(N,GAMALP,IX,IY,RR,KERR)
      WRITE(6,6020) KERR
!
      WRITE(6,6010) N,L,NDIV,ALF,UL,UP,GAMALP
      CALL DJTEGM(RR,N,L,NDIV,ALF,UL,UP,GAMALP,K,X2,CX,WK,IERR)
      WRITE(6,6020) IERR
      IF( IERR .EQ. 0 ) THEN
         WRITE(6,6030) K
         WRITE(6,6040) (X2(I),I=1,L)
         WRITE(6,6050) CX
      ENDIF
      STOP
!
 5000 FORMAT(/,&
      1X,'  ** GERERATION **',/,/,&
      1X,'     ** INPUT **',/,/,&
      1X,'       N    = ',I10,'    GAMALP  = ',F10.3)
 6010 FORMAT(/,&
      1X,'  ** TEST **',/,/,&
      1X,'     ** INPUT **',/,/,&
      1X,'       N    = ',I10,'       L    = ',I10,/,&
      1X,'       NDIV = ',I10,'       ALF  = ',F10.3,/,&
      1X,'       UL   = ',F10.3,'       UP   = ',F10.3,/,&
      1X,'    GAMALP  = ',F10.3)
 6020 FORMAT(/,&
      1X,'     ** OUTPUT **',/,/,&
      1X,'       IERR = ',I4)
 6030 FORMAT(/,&
      1X,'       NUMBER OF PASSED TEST (K) = ',I3,/)
 6040 FORMAT(&
      1X,'       TEST NO.      1     2     3     4     5',&
         '     6     7     8     9    10',/,/,&
      1X,'       CHI-SQUARE',/,&
      1X,'       VALUE (X2)',10F6.1,/)
 6050 FORMAT(/,&
      1X,'       CHI-SQUARE VALUE FOR PERCENT POINT (CX) = ',F8.1)
      END
```

(d) Output results

```
  ** GERERATION **

     ** INPUT **

       N    =      10000    GAMALP  =      3.000

     ** OUTPUT **

       IERR =    0

  ** TEST **

     ** INPUT **

       N    =      10000       L    =         10
       NDIV =         10       ALF  =      5.000
       UL   =      0.000       UP   =      9.000
    GAMALP  =      3.000

     ** OUTPUT **

       IERR =    0

       NUMBER OF PASSED TEST (K) =   9

       TEST NO.      1     2     3     4     5     6     7     8     9    10

       CHI-SQUARE
       VALUE (X2)   3.2   7.9   9.2   9.7   1.9   6.6  20.3   8.6   8.7   6.4


       CHI-SQUARE VALUE FOR PERCENT POINT (CX) =     16.9
```

## 2.3.7 DJTELG, RJTELG Logistic Distribution Random Number Test

(1) **Function**

Performs a frequency one-dimensional test on logistic distribution random numbers.

(2) **Usage**

Double precision:

CALL DJTELG (U, M, LT, NDIV, ALF, UL, UP, XA, XB, K, X2, CX, WK, IERR)

Single precision:

CALL RJTELG (U, M, LT, NDIV, ALF, UL, UP, XA, XB, K, X2, CX, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: {INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer}
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | U | {D / R} | M | Input | Random numbers |
| 2 | M | I | 1 | Input | Total number of random numbers |
| 3 | LT | I | 1 | Input | Test iteration count (See Note (a)) |
| 4 | NDIV | I | 1 | Input | Number of subdivisions |
| 5 | ALF | {D / R} | 1 | Input | Significance level (%) |
| 6 | UL | {D / R} | 1 | Input | Test interval lower limit (See Note (b)) |
| 7 | UP | {D / R} | 1 | Input | Test interval upper limit (See Note (b)) |
| 8 | XA | {D / R} | 1 | Input | Mean value |
| 9 | XB | {D / R} | 1 | Input | Value of parameter $\beta$ (See Note (c)) |
| 10 | K | I | 1 | Output | Number of passed tests |
| 11 | X2 | {D / R} | LT | Output | Test result $\chi^2$ value |
| 12 | CX | {D / R} | 1 | Output | $\chi^2$ value for significance level |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 13 | WK | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | See Contents | Work | Work area<br>**Size**: $2 \times$ NDIV $+ 4$ |
| 14 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) M $\geq$ LT

(b) LT $\geq 1$

(c) NDIV $\geq 2$

(d) UP $>$ UL

(e) $0.0 <$ ALF $< 100.0$

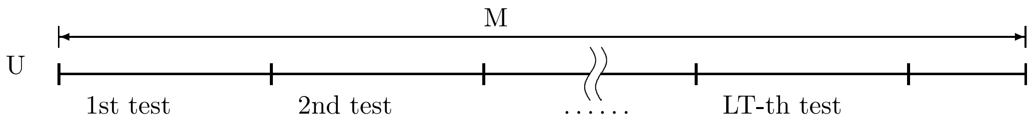
(f) XB $> 0.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | UL was too small, UP was too large, or M was too small. (See Note (b)) | Processing continues. (Test precision gets worse.) |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 3040 | Restriction (e) was not satisfied. | |
| 3050 | Restriction (f) was not satisfied. | |

(6) **Notes**

(a) The following figure shows the relationship between the total number of random numbers M, the test iteration count LT, and the random numbers U used for each test.




$\lfloor$M/LT$\rfloor$ random numbers U are taken at a time for each test, where $\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.

(b) If the test range lower limit UL is too small or if the upper limit UP is too large, then an extremely small expected frequency range is tested and test precision worsens.
If $F_{\mathrm{Ti}}$ is the expected frequency in each partial interval, then IERR = 1000 if the following condition occurs in the subroutine:

$$F_{\mathrm{Ti}} < 5 \qquad (\mathrm{i} = 1, \cdots, \mathrm{NDIV})$$

Expressions for calculating UL and UP, which are used as test criteria, are shown below.

$$(\mathrm{UP}-\mathrm{UL})\times\min\left\{\frac{\exp(-\frac{\mathrm{UL}-\mathrm{XA}}{\mathrm{XB}})}{\left\{1+\exp(-\frac{\mathrm{UL}-\mathrm{XA}}{\mathrm{XB}})\right\}^2},\frac{\exp(-\frac{\mathrm{UP}-\mathrm{XA}}{\mathrm{XB}})}{\left\{1+\exp(-\frac{\mathrm{UP}-\mathrm{XA}}{\mathrm{XB}})\right\}^2}\right\}\geq\frac{5\times\mathrm{NDIV}\times\mathrm{XB}}{\left\lfloor\frac{\mathrm{M}}{\mathrm{LT}}\right\rfloor}$$

$\lfloor x\rfloor$ represents the maximum integer that does not exceed $x$.

(c) Variance is defined as follows.

$$\sigma^2=\frac{\pi^2\beta^2}{3}$$

(7) **Example**

(a) Problem

Perform the test 10 times with 10 subdivisions on 10000 logistic distribution random numbers having mean 0.0 and $\beta=1.0$.

(b) Input data

M = 10000, L = 10, NDIV = 10, ALF = 5.0, UL = $-4.7$, UP = 4.7, XA = 1.0 and XB = 1.0.

(c) Main program

```
      PROGRAM BJTELG
! *** EXEMPLE OF DJTELG ***
      IMPLICIT REAL(8) (A-H,O-Z)
      PARAMETER ( NMAX=100000 )
      PARAMETER ( DONE = 1.0D0, DMONE = -1.0D0, DZERO = 0.0D0)
      PARAMETER ( DFIVE = 5.0D0, DM100 = -100.0D0, D100 = 100.0D0 )
      PARAMETER ( DM47 = -4.7D0, D47 = 4.7D0 )
      DIMENSION RR(NMAX),X2(NMAX),WK(2*NMAX+4)
!
      N  = 10000
      XA = DONE
      XB = DONE
      IX = 1
      IY = 1
!
      L    = 10
      NDIV = 10
      ALF  = DFIVE
      UL   = DM47
      UP   = D47
!
      CALL DJDBLG(N,XA,XB,IX,IY,RR,KERR)
      IF( KERR .GT. 0 ) THEN
         WRITE(6,6000) KERR
      ELSE
         WRITE(6,6010) N,L,NDIV,ALF,UL,UP,XA,XB
      ENDIF
!
      CALL DJTELG(RR,N,L,NDIV,ALF,UL,UP,XA,XB,K,X2,CX,WK,IERR)
      WRITE(6,6020) IERR
      IF( IERR .EQ. 0 ) THEN
         WRITE(6,6030) K
         WRITE(6,6040) (X2(I),I=1,L)
         WRITE(6,6050) CX
      ENDIF
      STOP
!
 6000 FORMAT(1X,'KERR = ',I4)
 6010 FORMAT(/,&
      1X,'     ** INPUT **',/,/,&
      1X,'       N    = ',I10,'       L    = ',I10,/,&
      1X,'       NDIV = ',I10,'       ALF  = ',F10.3,/,&
      1X,'       UL   = ',F10.3,'       UP   = ',F10.3,/,&
      1X,'       XA   = ',F10.3,'       XB   = ',F10.3)
 6020 FORMAT(/,&
      1X,'     ** OUTPUT **',/,/,&
      1X,'       IERR = ',I4)
 6030 FORMAT(/,&
      1X,'       NUMBER OF PASSED TEST (K) = ',I3,/)
 6040 FORMAT(&
      1X,'       TEST NO.      1     2     3     4     5',&
         '     6     7     8     9    10',/,/,&
      1X,'       CHI-SQUARE',/,&
      1X,'       VALUE (X2)',10F6.1,/)
 6050 FORMAT(/,&
      1X,'       CHI-SQUARE VALUE FOR PERCENT POINT (CX) = ',F8.1)
      END
```

(d) Output results

```
     ** INPUT **
```

```
    N    =      10000       L    =         10
    NDIV =         10       ALF  =      5.000
    UL   =     -4.700       UP   =      4.700
    XA   =      1.000       XB   =      1.000

  ** OUTPUT **

    IERR =    0

    NUMBER OF PASSED TEST (K) =  10

    TEST NO.      1     2     3     4     5     6     7     8     9    10

    CHI-SQUARE
    VALUE (X2)   4.8   7.6   7.3   6.7   4.1  10.7   9.8   6.3   6.1   8.1

    CHI-SQUARE VALUE FOR PERCENT POINT (CX) =     16.9
```

## 2.4 DISCRETE DISTRIBUTION RANDOM NUMBER TESTS

### 2.4.1 RJTEBI
### Binomial Distribution Random Number Test

(1) **Function**

Performs a frequency one-dimensional test on binomial distribution random numbers.

(2) **Usage**

Double precision:

Nothing

Single precision:

CALL RJTEBI (NL, M, LT, IUP, ALF, MN, P, K, X2, CX, WK, IERR)

(3) **Arguments**

R:Single precision real  C:Single precision complex  I: $\left\{ \begin{array}{l} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{array} \right\}$

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | NL | I | M | Input | Random numbers |
| 2 | M | I | 1 | Input | Total number of random numbers |
| 3 | LT | I | 1 | Input | Iteration count (See Note (a)) |
| 4 | IUP | I | 1 | Input | Test interval upper limit (See Note (b)) |
| 5 | ALF | R | 1 | Input | Significance (%) |
| 6 | MN | I | 1 | Input | Number of trials |
| 7 | P | R | 1 | Input | Success probability |
| 8 | K | I | 1 | Output | Number of passed tests |
| 9 | X2 | R | LT | Output | Test result $\chi^2$ value |
| 10 | CX | R | 1 | Output | $\chi^2$ value for significance level |
| 11 | WK | R | See Contents | Work | Work area<br>**Size**: $2 \times (\text{IUP} + 1)$ |
| 12 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $\text{M} \geq \text{LT}$

(b) $\text{LT} \geq 1$

(c) $0.0 < \text{ALF} < 100.0$

(d) $\text{MN} \geq 1$

(e) $0.0 < \text{P} < 1.0$

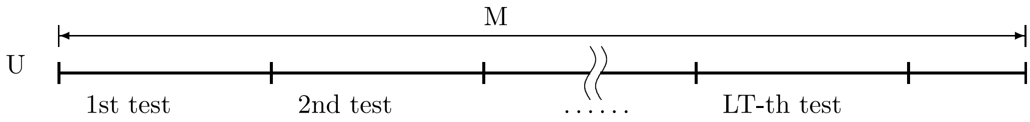
(f) $0 < \text{IUP} \leq \text{MN}$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | IUP was too large or M was too small. (See Note (b)) | Processing continues. (Test precision gets worse.) |
| 3000 | Restriction (a), (b), (c), (d), (e) or (f) was not satisfied. | Processing is aborted. |
| 4000 | NL(I) $< 0$ or NL(I) $>$ MN (I $= 1, \cdots, $ M) | |

(6) **Notes**

(a) The following figure shows the relationship between the total number of random numbers M, the test iteration count LT, and the random numbers NL used for each test.




(b) If the test range upper limit IUP is too large, then an extremely small expected frequency range is tested and test precision worsens.

If $F_{Ti}$ is the expected frequency in each partial interval, then IERR = 1000 if the following condition occurs in the RJTEBI subroutine:

$$F_{Ti} < 5 \qquad (i = 1, \cdots, N)$$

The value of IUP, which is used as a test criterion, is determined so that the following condition is satisfied.

$$\binom{MN}{IUP} P^{IUP}(1-P)^{MN-IUP} = \frac{5}{\lfloor M/LT \rfloor}$$

$\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.

Sample values of IUP are shown below for P=0.5.

| | | $\lfloor M/LT \rfloor$ | | | |
|---|---|---|---|---|---|
| | | 100 | 1, 000 | 10, 000 | 100, 000 |
| MN | 10 | 7 | 9 | 10 | 10 |
| | 50 | 29 | 33 | 36 | 38 |
| | 100 | 54 | 61 | 65 | 69 |

(7) **Example**

(a) Problem

Perform the test 10 times on 10000 binomial distribution random numbers with parmeters $m = 4$ and $p = 0.5$.

(b) Main program

```
      PROGRAM BJTEBI
! *** EXAMPLE OF RJTEBI ***
      PARAMETER ( M=10000, L=10  , MN=4      , IUP=4      )
      DIMENSION  NL(M) , X2(L) , IWK1(MN+2), WK1(MN+2) , WK(IUP+1,2)
!
      WRITE(6,1000)
      IX = 1
      IY = 1
      P  = 0.5
      IWK1(1) = 0
      WK1 (1) = 0.0
!
      CALL RJDBBI ( M,MN,P,IX,IY,NL,IWK1,WK1,IERR )
!
      ALF = 1.00
      WRITE(6,1100) M,L,IUP,ALF,MN,P
!
      CALL RJTEBI ( NL,M,L,IUP,ALF,MN,P,K,X2,CX,WK,IERR )
!
      WRITE(6,1200) IERR
      WRITE(6,1300) K
      WRITE(6,1400) (X2(I),I=1,L)
      WRITE(6,1500) CX
      STOP
 1000 FORMAT(' ',/,/,' *** RJTEBI ***',/)
 1100 FORMAT('    **  INPUT  **',/,10X,'  M = ',I5,5X,'  L = ',I5,/,&
      10X,' IUP= ',I5,5X,' ALF= ',F5.1,/,10X,' MN = ',I5,5X,&
      ' P  = ',F5.3,/)
 1200 FORMAT('    **  OUTPUT  **',/,/,10X,' IERR = ',I5,/)
 1300 FORMAT(10X,' NUMBER OF PASSED TEST (K) = ',I3,/)
 1400 FORMAT(10X,'TEST NO.     1       2       3       4       5   ',&
               '    6       7       8       9      10   ',/,/,&
             8X,'CHI-SQUARE',/,8X,'VALUE (X2)',10F8.1,/)
 1500 FORMAT(10X,' CHI-SQUARE VALUE FOR PERCENT POINT (CX) = ',F8.1)
      END
```

(c) Output results

```
*** RJTEBI ***

   **  INPUT  **
          M = 10000        L =    10
         IUP=     4      ALF=   1.0
         MN =     4      P  = 0.500

   **  OUTPUT  **

         IERR =     0

         NUMBER OF PASSED TEST (K) =  10

        TEST NO.     1       2       3       4       5       6       7       8       9      10

      CHI-SQUARE
      VALUE (X2)     4.6     9.6     2.5     1.9     0.9     1.2     3.8     2.4     2.2     2.5

         CHI-SQUARE VALUE FOR PERCENT POINT (CX) =     13.3
```

## 2.4.2 RJTENG
## Geometric Distribution Random Number Test

(1) **Function**

Performs a frequency one-dimensional test on geometric distribution random numbers.

(2) **Usage**

Double precision:

Nothing

Single precision:

CALL RJTENG (NL, M, LT, IUP, ALF, P, K, X2, CX, IWK, IERR)

(3) **Arguments**

R:Single precision real C:Single precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | NL | I | M | Input | Random numbers |
| 2 | M | I | 1 | Input | Total number of random numbers |
| 3 | LT | I | 1 | Input | Iteration count (See Note (a)) |
| 4 | IUP | I | 1 | Input | Test interval upper limit (See Note (b)) |
| 5 | ALF | R | 1 | Input | Significance (%) |
| 6 | P | R | 1 | Input | Success probability |
| 7 | K | I | 1 | Output | Number of passed tests |
| 8 | X2 | R | LT | Output | Test result $\chi^2$ value |
| 9 | CX | R | 1 | Output | $\chi^2$ value for significance level |
| 10 | IWK | I | $2 \times$ (IUP) | Work | Work area |
| 11 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) M $\geq$ LT

(b) LT $\geq 1$

(c) $0.0 <$ ALF $< 100.0$

(d) $0.0 <$ P $< 1.0$

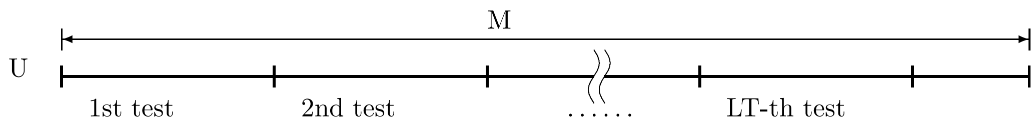
(e) $0 <$ IUP

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | IUP was too large or M was too small. (See Note (b)) | Processing continues. (Test precision gets worse.) |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 3040 | Restriction (e) was not satisfied. | |

(6) **Notes**

(a) The following figure shows the relationship between the total number of random numbers M, the test iteration count LT, and the random numbers NL used for each test.




(b) If the test range upper limit IUP is too large, then an extremely small expected frequency range is tested and test precision worsens.

If $F_{Ti}$ is the expected frequency in each partial interval, then IERR = 1000 if the following condition occurs in the RJTENG subroutine:

$$F_{Ti} < 5 \qquad (i = 1, \cdots, N)$$

The value of IUP, which is used as a test criterion, is determined so that the following condition is satisfied.

$$(1-P)^{IUP-1}\ P = \frac{5}{\lfloor M/LT \rfloor}$$

$\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.

Sample values of IUP are shown below for P=0.5.

| $\lfloor M/LT \rfloor$ | 100 | 1000 | 10000 | 100000 | 1000000 |
|---|---|---|---|---|---|
| IUP | 4 | 7 | 10 | 14 | 17 |

(7) **Example**

(a) Problem

Perform the test 10 times on 10000 geometric distribution random numbers with parameters $m = 4$ and $p = 0.5$.

(b) Main program

```
      PROGRAM BJTENG
! *** EXAMPLE OF RJTENG ***
      PARAMETER ( M=10000, L=10, IUP=6 )
      DIMENSION  NL(M), X2(L), IWK(IUP*2)
!
      IX = 1
      IY = 1
      P  = 0.6E0
!
      CALL RJDBNG ( M, P, IX, IY, NL, IERR )
!
      ALF = 1.0E0
!
      CALL RJTENG ( NL, M, L, IUP, ALF, P, K, X2, CX, IWK, IERR )
!
      WRITE(6,1000)
      WRITE(6,1100) M,L,IUP,ALF,P
      WRITE(6,1200) IERR
      WRITE(6,1300) K
      WRITE(6,1400) (X2(I),I=1,L)
      WRITE(6,1500) CX
      STOP
 1000 FORMAT(1X,/,/,&
             1X,'*** RJTEBI ***',/)
 1100 FORMAT(4X,'**  INPUT  **',/,1X,&
             10X,'M  = ',I5,'     ',' L  = ',I5,/,&
             11X,'IUP= ',I5,'     ',' ALF= ',F5.1,/,&
             11X,'P  = ',F5.1 ,/)
 1200 FORMAT(4X,'**  OUTPUT  **',/,/,&
             11X,'IERR = ',I5,/)
 1300 FORMAT(11X,'NUMBER OF PASSED TEST (K) = ',I3,/)
 1400 FORMAT(11X,&
             'TEST NO.     1       2       3       4       5',&
                  '      6       7       8       9      10   ',/,/,&
            9X ,'CHI-SQUARE',/,&
            9X ,'VALUE (X2)',10F8.1,/)
 1500 FORMAT(9X ,'CHI-SQUARE VALUE FOR PERCENT POINT ',&
                 '(CX) = ',F8.1                                )
      END
```

(c) Output results

```
*** RJTEBI ***

   **  INPUT  **
          M  = 10000      L  =    10
          IUP=     6      ALF=   1.0
          P  =   0.6

   **  OUTPUT  **

          IERR =     0

          NUMBER OF PASSED TEST (K) =  10

          TEST NO.     1       2       3       4       5       6       7       8       9      10

        CHI-SQUARE
        VALUE (X2)    8.4     6.9     7.8     5.6     2.6     6.1     4.7     2.2     4.4     5.3

        CHI-SQUARE VALUE FOR PERCENT POINT (CX) =     15.1
```

### 2.4.3 RJTEPO Poisson Distribution Random Number Test

(1) **Function**

Performs a frequency one-dimensional test on Poisson distribution random numbers.

(2) **Usage**

Double precision:

Nothing

Single precision:

CALL RJTEPO (NL, M, LT, IUP, ALF, AM, K, X2, CX, WK, IERR)

(3) **Arguments**

R:Single precision real  C:Single precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | NL | I | M | Input | Random numbers |
| 2 | M | I | 1 | Input | Total number of random numbers |
| 3 | LT | I | 1 | Input | Test iteration count (See Note (a)) |
| 4 | IUP | I | 1 | Input | Test interval upper limit (See Note (b)) |
| 5 | ALF | R | 1 | Input | Significance level (%) |
| 6 | AM | R | 1 | Input | Mean value |
| 7 | K | I | 1 | Output | Number of passed tests |
| 8 | X2 | R | LT | Output | Test result $\chi^2$ value |
| 9 | CX | R | 1 | Output | $\chi^2$ value for significance level |
| 10 | WK | R | See Contents | Work | Work area<br>**Size**: $2 \times (\mathrm{IUP}+1)$ |
| 11 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $\mathrm{M} \geq \mathrm{LT}$

(b) $\mathrm{LT} \geq 1$

(c) $0.0 < \mathrm{ALF} < 100.0$

(d) $0.0 < \mathrm{AM} <$ LOG (maximum value expressed by the computer)

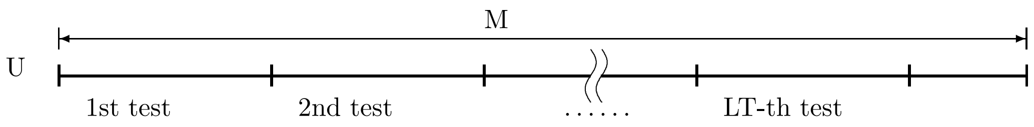
(e) $\mathrm{IUP} > 0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | IUP was too large or M was too small. (See Note (b)) | Processing continues. (Test precision gets worse.) |
| 3000 | Restriction (a), (b), (c), (d) or (e) was not satisfied. | Processing is aborted. |
| 4000 | $\mathrm{NL(i)} < 0 \quad (\mathrm{i} = 1, \cdots, \mathrm{M})$ | |

(6) **Notes**

(a) The following figure shows the relationship between the total number of random numbers M, the test iteration count LT, and the random numbers NL used for each test.




(b) If the test range upper limit IUP is too large, then an extremely small expected frequency range is tested and test precision worsens.

If $F_{Ti}$ is the expected frequency in each partial interval, then IERR = 1000 if the following condition occurs in the RJTEPO subroutine:

$$F_{Ti} < 5 \qquad (i = 1, \cdots, N)$$

The value of IUP, which is used as a test criterion, is determined so that the following condition is satisfied.

$$\frac{AM^{IUP}}{IUP!} = \frac{5 \times e^{AM}}{\lfloor M/LT \rfloor}$$

$\lfloor x \rfloor$ represents the maximum integer that does not exceed $x$.

Sample values of IUP are shown below for three different values of AM.

| | | $\lfloor M/LT \rfloor$ | | | |
|---|---|---|---|---|---|
| | | 100 | 1, 000 | 10, 000 | 100, 000 |
| AM | 1.0 | 3 | 4 | 6 | 7 |
| | 5.0 | 8 | 11 | 13 | 15 |
| | 10.0 | 14 | 18 | 21 | 24 |

(7) **Example**

(a) Problem

Perform the test 10 times on 10000 Poisson distribution random numbers having mean value 1.0.

(b) Main program

```
      PROGRAM BJTEPO
! *** EXAMPLE OF RJTEPO ***
      PARAMETER ( M=10000, L=10  , IUP=4      )
      DIMENSION  NL(M) , X2(L) , IWK1(100), WK1(100) , WK(IUP+1,2)
!
      WRITE(6,1000)
      IX = 1
      IY = 1
      AM = 1.0
      IWK1(1) = 0
      WK1 (1) = 0.0
!
      CALL RJDBPO ( M,AM,IX,IY,NL,IWK1,WK1,IERR )
!
      ALF = 1.0
      WRITE(6,1100) M,L,IUP,ALF,AM
!
      CALL RJTEPO ( NL,M,L,IUP,ALF,AM,K,X2,CX,WK,IERR )
!
      WRITE(6,1200) IERR
      WRITE(6,1300) K
      WRITE(6,1400) (X2(I),I=1,L)
      WRITE(6,1500) CX
      STOP
 1000 FORMAT(' ',/,/,' *** RJTEPO ***',/)
```

```
1100 FORMAT('    **  INPUT  **',/,10X,'  M = ',I5,5X,'  L = ',I5,/,&
     10X,' IUP= ',I5,5X,' ALF= ',F5.1,/,10X,' AM = ',F5.1,/)
1200 FORMAT('    **  OUTPUT  **',/,/,10X,' IERR = ',I5,/)
1300 FORMAT(10X,' NUMBER OF PASSED TEST (K) = ',I3,/)
1400 FORMAT(10X,'TEST NO.     1       2       3       4       5   ',&
                '    6       7       8       9      10   ',/,/,&
            8X,'CHI-SQUARE',/,8X,'VALUE (X2)',10F8.1,/)
1500 FORMAT(10X,' CHI-SQUARE VALUE FOR PERCENT POINT (CX) = ',F8.1)
     END
```

(c) Output results

```
*** RJTEPO ***

   **  INPUT  **
          M = 10000       L =    10
         IUP=     4      ALF=   1.0
         AM =   1.0

   **  OUTPUT  **

         IERR =     0

         NUMBER OF PASSED TEST (K) =  10

        TEST NO.     1       2       3       4       5       6       7       8       9      10

      CHI-SQUARE
      VALUE (X2)     4.7     9.2     7.9     2.7     2.2     5.6     5.1     1.5     1.0     2.4

         CHI-SQUARE VALUE FOR PERCENT POINT (CX) =     13.3
```

# Chapter 3

# PROBABILITY DISTRIBUTIONS

## 3.1 INTRODUCTION

In statistical analysis, processing for estimates or tests associate a variable called a random variable with various types of data. Random variables are broadly divided into **discrete random variables**, which correspond to cases in which the realized values are expressed by using discrete values such as $x_1, x_2, \cdots$ and **continuous random variables**, which correspond to cases in which the realized values take arbitrary values within a continuous interval such as $0 < x < 1$. With a discrete random variable, we can consider the probability ($Pr.\{X = x\}$) that the random variable ($X$) takes a specific value ($x$). However, with a continuous random variable, we consider the probability ($Pr.\{a \le X < b\}$) that the random variable ($X$) takes a value within an arbitrary subinterval ($[a, b)$) in the interval where the realized values of the random variable ($X$) exist. Using the probability $Pr.\{x \le X < x+dx\}$ that the random variable $X$ takes an arbitrary value of the infinitesimal interval $[x, x + dx)$, the "function" $f(x)$ that satisfies

$$\int_x^{x+dx} f(u)du = Pr.\{x \le X < x + dx\} \ (dx \to 0)$$

is called the probability density function (p.d.f.) of the continuous random variable $X$. From the definition of probability, we have

$$\int_{-\infty}^{\infty} f(u)du = 1$$

Normally, the value of $f(x)$ is set to zero for any interval in which the random variable $X$ is not defined. The cumulative distribution function (c.d.f.) $F(x)$ is defined as the integral of the probability density function $f(x)$ as follows:

$$F(x) = \int_{-\infty}^{x} f(u)du$$

For practical use, the function $G(x)$, which is defined by the following expression, is also used as the cumulative distribution function.

$$G(x) = 1 - F(x) = \int_x^{\infty} f(u)du$$

These relationships can also easily be extended to several variables. For stricter definitions of the probability density function and cumulative distribution function or for unified handling of discrete and continuous random variables, refer to specialized technical texts. This library provides functions for computing the probability density function or cumulative distribution function of the following kinds of probability distributions as well as the values of their inverse functions.

- Normal Distribution
- Inverse of Normal Distribution
- Bivariate Normal Distribution
- $\chi^2$ Distribution

- Inverse of $\chi^2$ Distribution
- Noncentral $\chi^2$ Distribution
- Inverse Noncentral $\chi^2$ Distribution
- $t$ Distribution
- Inverse of $t$ Distribution
- Noncentral $t$ Distribution
- Inverse Noncentral $t$ Distribution
- $F$ Distribution
- Inverse of $F$ Distribution
- Gamma Distribution
- Inverse Gamma Distribution
- Beta Distribution
- Inverse Beta Distribution
- Uniform Distribution
- Triangular Distribution
- Pareto Distribution
- Weibull Distribution
- Exponential Distribution
- Gumbel Distribution
- Logarithmic Distribution
- Log-Normal Distribution
- Logistic Distribution
- Cauchy Distribution
- Binomial Distribution and Negative Binomial Distribution
- Geometric Distribution
- Poisson Distribution
- Hypergeometric Distribution
- Negative Hypergeometric Distribution

### 3.1.1 Explanation

(1) **Normal Distribution**

The probability density function of the normal distribution having mean $\mu$ and variance $\sigma^2$ is defined as follows.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} \quad (\sigma > 0)$$

(2) $\chi^2$ **Distribution**

The probability density function of the $\chi^2$ distribution having frequency $\chi^2$ and number of degrees of freedom $\nu$ is defined as follows.

$$f(\chi^2|\nu) = \begin{cases} \frac{1}{2^{\frac{\nu}{2}}\Gamma(\dfrac{\nu}{2})}(\chi^2)^{\frac{\nu}{2}-1}e^{-\frac{\chi^2}{2}} & (\chi^2 > 0) \\ 0 & (\chi^2 \leq 0) \end{cases}$$

The mean and variance of the $\chi^2$ distribution are given by the following equations.

$$E[\chi^2(\nu)] = \nu, \sigma^2[\chi^2(\nu)] = 2\nu$$

(3) **Noncentral** $\chi^2$ **Distribution**

The probability density function of the noncentral $\chi^2$ distribution having frequency $\chi^2$, number of degrees of freedom $\nu$, and noncentrality parameter $\lambda$ is defined as follows.

$$f(x|\nu,\lambda) = \begin{cases} \dfrac{e^{-\frac{(x+\lambda)}{2}}x^{\frac{(\nu-2)}{2}}}{2^{\frac{\nu}{2}}} \displaystyle\sum_{k=0}^{\infty} \frac{\lambda^k x^k}{2^{2k}k!\Gamma(\frac{\nu}{2}+k)} & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

(4) $t$ **Distribution**

The probability density function of the $t$ distribution having frequency $t$ and number of degrees of freedom $\nu$ is defined as follows.

$$f(t|\nu) = \frac{\Gamma(\dfrac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\dfrac{\nu}{2})(1+\dfrac{t^2}{\nu})^{\frac{\nu+1}{2}}}$$

The mean and variance of the $t$ distribution are given by the following equations.

$$E[t(\nu)] = 0, \sigma^2[t(\nu)] = \frac{\nu}{\nu-2} \quad (\nu > 2)$$

(5) **Noncentral** $t$ **Distribution** The probability density function of the noncentral $t$ distribution having frequency $t$, number of degrees of freedom $\nu$, and noncentrality parameter $\delta$ is defined as follows.

$$f(t|\nu,\delta) = \frac{\nu^{\frac{\nu}{2}}e^{-\frac{\delta^2}{2}}}{\sqrt{\pi}\Gamma(\dfrac{\nu}{2})(\nu+t^2)^{\frac{(\nu+1)}{2}}} \sum_{k=0}^{\infty} \Gamma(\frac{\nu+k+1}{2})\frac{\delta^k}{k!}(\frac{2t^2}{\nu+t^2})^{\frac{k}{2}}$$

(6) $F$ **Distribution**

The probability density function of the $F$ distribution having frequency $F$ and numbers of degrees of freedom $\nu_1$ and $\nu_2$ is defined as follows.

$$f(x|\nu_1,\nu_2) = \frac{\nu_1^{\frac{\nu_1}{2}} \cdot \nu_2^{\frac{\nu_2}{2}} \cdot x^{\frac{\nu_1}{2}-1}}{B(\frac{\nu_1}{2},\frac{\nu_2}{2})(\nu_1 x+\nu_2)^{\frac{\nu_1+\nu_2}{2}}} = \frac{1}{B(\frac{\nu_1}{2},\frac{\nu_2}{2})}(\frac{\nu_1}{\nu_2})^{\frac{\nu_1}{2}}(1+\frac{\nu_1}{\nu_2}x)^{-\frac{\nu_1+\nu_2}{2}}x^{\frac{\nu_1}{2}-1}$$

The mean and variance of the $F$ distribution are given by the following equations.

$$E[F] = \frac{\nu_2}{\nu_2-2}\ (\nu_2>2), \sigma^2[F] = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}\ (\nu_2>4)$$

(7) **Gamma Distribution**

The probability density function of the gamma distribution having parameters $\alpha$ and $\beta$ is defined as follows.

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} & (x>0;\alpha,\beta>0) \\ 0 & (x\leq 0;\alpha,\beta>0) \end{cases}$$

The mean and variance of the gamma distribution are given by the following equations.

$$E[x] = \frac{\alpha}{\beta}, \sigma^2[x] = \frac{\alpha}{\beta^2}$$

(8) **Beta Distribution**

The probability density function of the beta distribution having parameters $a$ and $b$ is defined as follows.

$$f(x;a,b) = \begin{cases} \dfrac{1}{B(a,b)}x^{a-1}(1-x)^{b-1} & (0<x<1;a,b>0) \\ 0 & (x\leq 0, x\geq 1;a,b>0) \end{cases}$$

(9) **Uniform Distribution**

The probability density function of the uniform distribution within the interval $(a,b)$ is defined as follows.

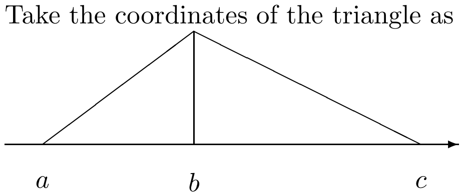
$$f(x) = \begin{cases} \dfrac{1}{b-a} & (a\leq x\leq b) \\ 0 & (x<a, x>b) \end{cases}$$

The mean and variance of the uniform distribution are given by the following equations.

$$E[x] = \frac{a+b}{2}, \sigma^2[x] = \frac{(b-a)^2}{12}$$

(10) **Triangular Distribution**

Take the coordinates of the triangle as shown in the following figure.




$a$:$x$ coordinate of the left end of the triangular distribution

$b$:$x$ coordinate of the apex of the triangular distribution

$c$:$x$ coordinate of the right end of the triangular distribution

The probability density function of the triangular distribution at this time is defined as follows.

$$f(x) = \begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)} & (a \leq x \leq b) \\ \dfrac{2(c-x)}{(c-a)(c-b)} & (b < x \leq c) \\ 0 & (x < a, x > c) \end{cases}$$

(11) **Pareto Distribution**

The probability density function of the Pareto distribution having parameters $a$ and $b$ ($a > 1$ , $b > 0$) is defined as follows.

$$f(x;a,b) = \begin{cases} (a-1)(\dfrac{x}{b})^{-a}\dfrac{1}{b} & (x > b; a > 1, b > 0) \\ 0 & (x \leq b; a > 1, b > 0) \end{cases}$$

(12) **Weibull Distribution**

The probability density function of the Weibull distribution having parameters $a$ and $b$ ($a > 0$ , $b > 0$) is defined as follows.

$$f(x;a,b) = \begin{cases} a\left(\dfrac{x}{b}\right)^{a-1} e^{-(\frac{x}{b})^a}\dfrac{1}{b} & (0 < x; a, b > 0) \\ 0 & (x \leq 0; a, b > 0) \end{cases}$$

(13) **Exponential Distribution**

A gamma distribution in which the parameter $\alpha$ is 1 is called the exponential distribution. In the exponential distribution, $\lambda$ is used in place of the parameter $\beta$. The probability density function is defined as follows.

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & (x > 0; \lambda > 0) \\ 0 & (x \leq 0; \lambda > 0) \end{cases}$$

The mean and variance of the exponential distribution are given by the following equations.

$$E[x] = \frac{1}{\lambda}, \sigma^2[x] = \frac{1}{\lambda^2}$$

(14) **Gumbel Distribution**

The probability density function of the Gumbel distribution having parameters $a$ and $b$ ($b > 0$) is defined as follows.

$$f(x;a,b) = \frac{1}{b}e^{\frac{x-a}{b}}e^{-e^{\frac{x-a}{b}}}$$

(15) **Logarithmic Distribution**

The probability density function of the Logarithmic distribution parameters within the interval $(a, b)$ is defined as follows.

$$f(x;a,b) = \frac{\log x}{b(\log b - 1) - a(\log a - 1)}$$

(16) **Log-Normal Distribution**

The probability density function of the log-normal distribution having mean $e^{\mu}\sqrt{e^{\sigma^2}}$ and variance $e^{2\mu}e^{\sigma^2}(e^{\sigma^2} - 1)$ is defined as follows.

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \quad (\sigma > 0)$$

(17) **Logistic Distribution**

The probability density function of the logistic distribution having mean $\alpha$ and variance $\sigma^2$ is defined as follows.

$$f(x) = \frac{e^{-\frac{x-\alpha}{\beta}}}{\beta\left\{1+e^{-\frac{x-\alpha}{\beta}}\right\}^2}$$
$$(-\infty < x < \infty, -\infty < \alpha < \infty, \beta > 0) \tag{3.1}$$
$$\sigma^2 = \frac{\pi^2\beta^2}{3}$$

(18) **Cauchy Distribution**

The probability density function of Cauchy distribution having parameters $\alpha$ and $\beta$ is defined as follows.

$$f(x;\alpha,\beta) = \frac{1}{\pi}\left[\frac{\beta}{\beta^2+(x-\alpha)^2}\right] \quad (\beta > 0)$$

(19) **Binomial Distribution and Negative Binomial Distribution**

Given the probability that an event will occur $p$, the number of trials $n$, and the number of occurrences $m$, the binomial distribution probability in $m$ occurrences is defined as follows.

$$P_{BIN}(X=m;p,n) = \begin{pmatrix} n \\ m \end{pmatrix} p^m \cdot q^{n-m} \quad (q = 1-p)$$

The mean and variance of the binomial distribution are given by the following equations.

$$E[m] = np, \sigma^2[m] = np(1-p)$$

Given the probability of success in one trial $p$ and the number of successes $n$ and number of failures $m$ in repeated trials, the negative binomial distribution probability in $m$ failures is defined as follows.

$$P_{NB}(X=m;p,n) = \begin{pmatrix} n+m-1 \\ m \end{pmatrix} p^n \cdot q^m \quad (q = 1-p)$$

The mean and variance of the negative binomial distribution are given by the following equations.

$$E[m] = \frac{n}{p}, \sigma^2[m] = \frac{n(1-p)}{p^2}$$

(20) **Geometric Distribution**

Given the probability of success in $m$ trial $p$, the D1DDBP or R1DDBP obtains the values of the geometric distribution probability $P_{NB}(X=m;p)$ is defined by the following equations.

$$P_{NB}(X=m;p) = q^{m-1}p \qquad (q = 1-p)$$

The mean and variance of the Geometric distribution are given by the following equations.

$$E[m] = \frac{1}{p}, \sigma^2[m] = \frac{q}{p^2}$$

(21) **Poisson Distribution**
Given the mean $\lambda$ and random variable $k$, the value of the probability $Pr.\{X = k\}$ of a Poisson distribution is defined by the following expressions.

$$Pr.\{X = k\} = e^{-\lambda}\frac{\lambda^k}{k!} \qquad (k = 0, 1, 2, \cdots; \lambda > 0)$$

(22) **Hypergeometric Distribution**
Assume that there is a lot of size $N$ in which $M$ of the $N$ articles are inferior goods and $N - M$ articles are of good quality. When an arbitrary sample of size $n$ is extracted from this lot, the hypergeometric distribution probability $Pr.\{X = k\}$ corresponding to the probability distribution in which $k$ inferior goods appear is defined as follows.

$$Pr.\{X = k\} = \begin{cases} \dfrac{\begin{pmatrix} M \\ k \end{pmatrix}\begin{pmatrix} N-M \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} & k = 0, 1, 2, \cdots, \min\{M, n\} \\ 0 & Otherwise \end{cases}$$

The mathematical expectation and variance of the hypergeometric distribution are given by the following equations.

$$E(X) = np, \sigma^2(X) = \frac{N-k}{N-1}np(1-p) \quad (p = \frac{M}{N})$$

(23) **Negative Hypergeometric Distribution**
Assume that there is a lot of size $NN$ in which $M$ of the $NN$ articles are inferior goods and $NN - M$ articles are of good quality. Sampling from this lot is continued until $n$ inferior goods are extracted. The negative hypergeometric distribution probability $Pr.\{X = k\}$ is defined as the probability of such occurrences that exactly $k$ goods has been extracted at this time. The probability probability $Pr.\{X = k\}$ is defined as follows.

$$Pr.\{X = k\} = \begin{cases} \dfrac{\begin{pmatrix} M \\ NN-1 \end{pmatrix}\begin{pmatrix} NN-M \\ k-n \end{pmatrix}}{\begin{pmatrix} NN \\ k-1 \end{pmatrix}} \times \dfrac{M-n+1}{NN-k+1} & \\ = \dfrac{\begin{pmatrix} k-1 \\ n-1 \end{pmatrix}\begin{pmatrix} NN-k \\ M-n \end{pmatrix}}{\begin{pmatrix} NN \\ M \end{pmatrix}} & \text{When} k = n, n+1, n+2, \cdots, NN-M+n \\ 0 & \text{Otherwise.} \end{cases}$$

$$F(k) = \sum_{i=n}^{k} Pr.\{X = i\} = \frac{\sum_{i=0}^{k} \begin{pmatrix} i-1 \\ n-1 \end{pmatrix}\begin{pmatrix} NN-i \\ M-n \end{pmatrix}}{\begin{pmatrix} NN \\ M \end{pmatrix}}$$

### 3.1.2 Reference Bibliography

(1) Feller, W. , “An introduction to probability theory and its applications: I (3rd ed.) II (2nd ed.)”, John Wiley & Sons, New York (1968, 1970)

(2) Abramowitz, M. and Stegur, I. A. , eds. , “Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables”, Dover Publications, Inc. (1965).

## 3.2 CONTINUOUS DISTRIBUTIONS

### 3.2.1 D1CDNO, R1CDNO<br>Normal Distribution

(1) **Function**

For a normal distribution having mean $\mu$ and variance $\sigma^2$, the D1CDNO or R1CDNO obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (\sigma > 0)$$

(b) cumulative distribution function; c.d.f.

$$P(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad (\sigma > 0)$$

(c) c.d.f.

$$Q(x) = 1 - P(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad (\sigma > 0)$$

(2) **Usage**

Double precision:

CALL D1CDNO (XE, XV, XI, XO, ISW, IERR)

Single precision:

CALL R1CDNO (XE, XV, XI, XO, ISW, IERR)

## (3) Arguments

D:Double precision real  Z:Double precision complex  I: $\begin{Bmatrix} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{Bmatrix}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | XE | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | 1 | Input | Value of mean $\mu$ |
| 2 | XV | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | 1 | Input | Value of variance $\sigma^2$ |
| 3 | XI | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | 1 | Input | Value of random variable $x$ |
| 4 | XO | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | 1 | Output | Value of the probability density function $f(x)$ of the normal distribution or of the cumulative distribution function $P(x)$ or $Q(x)$ of the normal distribution. |
| 5 | ISW | I | 1 | Input | ISW=0: Obtain the value of the probability density function $f(x)$ for XO<br>ISW=1: Obtain the value of the cumulative distribution function $P(x)$ for XO<br>ISW=2:Obtain the value of the cumulative distribution function $Q(x)$ for XO |
| 6 | IERR | I | 1 | Output | Error indicator |

## (4) Restrictions

(a) ISW $\in \{0, 1, 2\}$

(b) XV $> 0.0$

## (5) Error indicator

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |

## (6) Notes

(a) From the relational expression $P(x) + Q(x) = 1$, it is possible to obtain either $P(x)$ or $Q(x)$ from the other. However, there may be times when cancellation of significant digits occurs, preventing good precision from being obtained.

(b) A random variable that obeys a normal distribution having parameters $\mu$ and $\sigma$ is usually represented by $N(\mu, \sigma^2)$. When $\mu = 0$ and $\sigma = 1$, the random variable $N(0, 1)$ is called the standard random variable, and the probability distribution is called the standard normal distribution.

(7) **Example**

(a) Problem

Let $\mu = 5.0$, $\sigma^2 = 2.5$ and $x = 3.0$ and obtain the values of the probability density function $f(x)$ and the cumulative distribution functions $P(x)$ and $Q(x)$.

(b) Input data

XE = 5.0, XV = 2.5 and XI = 3.0.

(c) Main program

```
      PROGRAM B1CDNO
! *** EXAMPLE OF D1CDNO ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER ISW
      REAL(8) XE,XV,XI,XO
!
      XE=5.0D0
      XV=2.5D0
      XI=3.0D0
      WRITE(6,1000)
      WRITE(6,2000) XE
      WRITE(6,2010) XV
      WRITE(6,2020) XI
      WRITE(6,3000)
      ISW=0
      CALL D1CDNO(XE,XV,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) XO
!
      ISW=1
      CALL D1CDNO(XE,XV,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDNO(XE,XV,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDNO ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'XE = ',F4.1)
 2010 FORMAT(9X,'XV = ',F4.1)
 2020 FORMAT(9X,'XI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'VALUE OF P.D.F = ',D17.10)
 5010 FORMAT(9X,'VALUE OF C.D.F(1) = ',D17.10)
 5020 FORMAT(9X,'VALUE OF C.D.F(2) = ',D17.10)
      END
```

(d) Output results

```
    *** D1CDNO ***
     ** INPUT **
        XE =  5.0
        XV =  2.5
        XI =  3.0


     ** OUTPUT **
        IERR =    0
        VALUE OF P.D.F =  0.1133716522D+00
        IERR =    0
        VALUE OF C.D.F(1) =  0.1029516054D+00
        IERR =    0
        VALUE OF C.D.F(2) =  0.8970483946D+00
```

### 3.2.2 D1CDIN, R1CDIN Inverse of Normal Distribution

(1) **Function**

When given the cumulative distribution function (c.d.f.) $P(x)$ or $Q(x)$ of the normal distribution having mean $\mu$ and variance $\sigma^2$, the D1CDIN or R1CDIN obtains the random variable value $x$ at that time. $P(x)$ and $Q(x)$ are defined by the following equations.

$$P(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad (\sigma > 0)$$

$$Q(x) = 1 - P(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad (\sigma > 0)$$

(2) **Usage**

Double precision:
CALL D1CDIN (XE, XV, XI, XO, ISW, IERR)
Single precision:
CALL R1CDIN (XE, XV, XI, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | XE | $\begin{Bmatrix}\text{D}\\ \text{R}\end{Bmatrix}$ | 1 | Input | Value of mean $\mu$ |
| 2 | XV | $\begin{Bmatrix}\text{D}\\ \text{R}\end{Bmatrix}$ | 1 | Input | Value of variance $\sigma^2$ |
| 3 | XI | $\begin{Bmatrix}\text{D}\\ \text{R}\end{Bmatrix}$ | 1 | Input | Value of the cumulative distribution function $P(x)$ or $Q(x)$ of the normal distribution. |
| 4 | XO | $\begin{Bmatrix}\text{D}\\ \text{R}\end{Bmatrix}$ | 1 | Output | Value of random variable $x$ |
| 5 | ISW | I | 1 | Input | ISW=1: Input the value of $P(x)$ for XI<br>ISW=2: Input the value of $Q(x)$ for XI |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{1, 2\}$

(b) XV $> 0.0$

(c) $0.0 \leq$ XI $\leq 1.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | XI = 0.0 or XI = 1.0 | The positive maximum value or negative minimum value is set for XO. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3500 | The maximum number of iterations was reached before the specified precision was obtained. | The value at that time is returned. |

(6) **Notes**

(a) A random variable that obeys a normal distribution having parameters $\mu$ and $\sigma$ is usually represented by $N(\mu, \sigma^2)$. When $\mu = 0$ and $\sigma = 1$, the random variable $N(0, 1)$ is called the standard random variable, and the probability distribution is called the standard normal distribution.

(7) **Example**

(a) Problem
For $\mu = 5.0$ and $\sigma^2 = 2.5$, obtain the values of $x$ for which $P(x) = 0.2$ and $Q(x) = 0.2$ occur, respectively.

(b) Input data
XE = 5.0, XV = 2.5 and XI = 0.2.

(c) Main program

```
      PROGRAM B1CDIN
! *** EXAMPLE OF D1CDIN ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER ISW
      REAL(8) XE,XV,XI,XO
!
      XE=5.0D0
      XV=2.5D0
      XI=0.2D0
      WRITE(6,1000)
      WRITE(6,2000) XE
      WRITE(6,2010) XV
      WRITE(6,2020) XI
      WRITE(6,3000)
!
      ISW=1
      CALL D1CDIN(XE,XV,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDIN(XE,XV,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDIN ***',/,&
```

```
     6X,'** INPUT **')
2000 FORMAT(9X,'XE = ',F4.1)
2010 FORMAT(9X,'XV = ',F4.1)
2020 FORMAT(9X,'XI = ',F4.1)
3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
4000 FORMAT(9X,'IERR = ',I4)
5010 FORMAT(9X,'VALUE X CORRESPONDING TO P(X)=XI: ',D17.10)
5020 FORMAT(9X,'VALUE X CORRESPONDING TO Q(X)=XI: ',D17.10)
     END
```

(d) Output results

```
 *** D1CDIN ***
  ** INPUT **
     XE =  5.0
     XV =  2.5
     XI =  0.2


  ** OUTPUT **
     IERR =    0
     VALUE X CORRESPONDING TO P(X)=XI:  0.3669279987D+01
     IERR =    0
     VALUE X CORRESPONDING TO Q(X)=XI:  0.6330720013D+01
```

### 3.2.3 D1CDBN, R1CDBN Bivariate Normal Distribution

(1) **Function**

For a bivariate normal distribution having means $\mu_x, \mu_y$, variances ${\sigma_x}^2, {\sigma_y}^2$ and correlation coefficient $\rho$, the D1CDBN or R1CDBN obtains the values of the following functions.

(a) probability density function (p.d.f.)

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{Q}{2(1-\rho^2)}} \quad (\sigma_x, \sigma_y > 0)$$

where,

$$Q = \frac{(x-\mu_x)^2}{{\sigma_x}^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{{\sigma_y}^2} \quad (\sigma_x, \sigma_y > 0)$$

(b) cumulative distribution function; c.d.f.

$$B(h,k;\rho) = \int_{-\infty}^{h} \int_{-\infty}^{k} f(x,y)\ dxdy$$

(c) c.d.f.

$$L(h,k;\rho) = \int_{h}^{\infty} \int_{k}^{\infty} f(x,y)\ dxdy$$

(2) **Usage**

Double precision:

CALL D1CDBN (XE, YE, XV, YV, XH, YK, RHO, XO, ISW, IERR)

Single precision:

CALL R1CDBN (XE, YE, XV, YV, XH, YK, RHO, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\\text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | XE | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 1 | Input | Value of mean $\mu_x$. |
| 2 | YE | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 1 | Input | Value of mean $\mu_y$. |
| 3 | XV | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 1 | Input | Value of variance ${\sigma_x}^2$. |
| 4 | YV | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 1 | Input | Value of variance ${\sigma_y}^2$. |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 5 | XH | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | 1 | Input | ISW=0:Value of random variable $x$<br>ISW=1:Upper bound $h$ of the integration range of random variable $x$<br>ISW=2:Lower bound $h$ of the integration range of random variable $x$ |
| 6 | YK | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | 1 | Input | ISW=0:Value of random variable $y$<br>ISW=1:Upper bound $k$ of the integration range of random variable $y$<br>ISW=2:Lower bound $k$ of the integration range of random variable $y$ |
| 7 | RHO | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | 1 | Input | Value of correlation coefficient $\rho$. |
| 8 | XO | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | 1 | Output | Value of the probability density function $f(x, y)$ or of the cumulative distribution function $B(h, k; \rho)$ or $L(h, k; \rho)$ of the bivariate normal distribution. |
| 9 | ISW | I | 1 | Input | Processing switch<br>ISW=0: Obtain the value of the probability density function $f(x, y)$ for XO<br>ISW=1:Obtain the value of the cumulative distribution function $B(h, k; \rho)$ for XO<br>ISW=2:Obtain the value of the cumulative distribution function $L(h, k; \rho)$ for XO |
| 10 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $\text{ISW} \in \{0, 1, 2\}$

(b) $\text{XV}, \text{YV} > 0.0$

(c) $-1.0 \leq \rho \leq 1.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3500 | The maximum number of iterations was reached before the specified precision was obtained. | The value at that time is returned. |

(6) **Notes**
None

(7) **Example**

(a) Problem

Let $\mu_x = 1.0, \mu_y = 2.0, {\sigma_x}^2 = 3.0, {\sigma_y}^2 = 4.0,\ x = h = 5.0, y = k = 6.0$ and $\rho = 0.7$ and obtain the values of the probability density function $f(x, y)$ and the cumulative distribution functions $B(h, k; \rho)$ and $L(h, k; \rho)$.

(b) Input data

XE = 1.0, YE = 2.0, XV = 3.0, YV = 4.0, XH = 5.0, YK = 6.0 and RHO = 0.7.

(c) Main program

```
      PROGRAM B1CDBN
! *** EXAMPLE OF D1CDBN ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER ISW
      REAL(8) XE,YE,XV,YV,XH,YK,RHO,XO
!
      XE=1.0D0
      YE=2.0D0
      XV=3.0D0
      YV=4.0D0
      XH=5.0D0
      YK=6.0D0
      RHO=0.7D0
      WRITE(6,1000)
      WRITE(6,2000) XE,YE
      WRITE(6,2010) XV,YV
      WRITE(6,2020) XH,YK
      WRITE(6,2030) RHO
      WRITE(6,3000)
      ISW=0
      CALL D1CDBN(XE,YE,XV,YV,XH,YK,RHO,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) XO
!
      ISW=1
      CALL D1CDBN(XE,YE,XV,YV,XH,YK,RHO,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDBN(XE,YE,XV,YV,XH,YK,RHO,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDBN ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'XE = ',F4.1,'  YE = ',F4.1)
 2010 FORMAT(9X,'XV = ',F4.1,'  YV = ',F4.1)
 2020 FORMAT(9X,'XH = ',F4.1,'  YK = ',F4.1)
 2030 FORMAT(9X,'RHO = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'VALUE OF P.D.F = ',D17.10)
 5010 FORMAT(9X,'VALUE OF C.D.F(1) = ',D17.10)
 5020 FORMAT(9X,'VALUE OF C.D.F(2) = ',D17.10)
      END
```

(d) Output results

```
     *** D1CDBN ***
      ** INPUT **
         XE =  1.0  YE =  2.0
         XV =  3.0  YV =  4.0
         XH =  5.0  YK =  6.0
         RHO =  0.7


      ** OUTPUT **
         IERR =    0
         VALUE OF P.D.F =  0.3870186645D-02
         IERR =    0
         VALUE OF C.D.F(1) =  0.9711871129D+00
         IERR =    0
         VALUE OF C.D.F(2) =  0.4397912483D-02
```

### 3.2.4 D1CDCH, R1CDCH $\chi^2$ Distribution

(1) **Function**

For a $\chi^2$ distribution having frequency $\chi^2$ and number of degrees of freedom $\nu$, the D1CDCH or R1CDCH obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(\chi^2|\nu) = \begin{cases} \dfrac{1}{2^{\frac{\nu}{2}}\Gamma(\dfrac{\nu}{2})}(\chi^2)^{\frac{\nu}{2}-1}e^{-\frac{\chi^2}{2}} & (\chi^2 > 0) \\ 0 & (\chi^2 \le 0) \end{cases}$$

(b) cumulative distribution function; c.d.f.

$$P(\chi^2|\nu) = \int_0^{\chi^2} \frac{1}{2^{\frac{\nu}{2}}\Gamma(\dfrac{\nu}{2})}(t)^{\frac{\nu}{2}-1}e^{-\frac{t}{2}}dt \quad (0 \le \chi^2 < \infty)$$

(c) c.d.f.

$$Q(\chi^2|\nu) = 1 - P(\chi^2|\nu) = \int_{\chi^2}^{\infty} \frac{1}{2^{\frac{\nu}{2}}\Gamma(\dfrac{\nu}{2})}(t)^{\frac{\nu}{2}-1}e^{-\frac{t}{2}}dt \quad (0 \le \chi^2 < \infty)$$

(2) **Usage**

Double precision:

CALL D1CDCH (N, XI, XO, ISW, IERR)

Single precision:

CALL R1CDCH (N, XI, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: { INTEGER(4) as for 32bit Integer, INTEGER(8) as for 64bit Integer }
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Value of number of degrees of freedom $\nu$ |
| 2 | XI | {D, R} | 1 | Input | Value of frequency $\chi^2$ |
| 3 | XO | {D, R} | 1 | Output | Value of the probability density function $f(\chi^2\|\nu)$ of the $\chi^2$ distribution or of the cumulative distribution function $P(\chi^2\|\nu)$ or $Q(\chi^2\|\nu)$ of the $\chi^2$ distribution. |
| 4 | ISW | I | 1 | Input | ISW=0: Obtain the value of the probability density function $f(\chi^2\|\nu)$ for XO<br>ISW=1: Obtain the value of the cumulative distribution function $P(\chi^2\|\nu)$ for XO<br>ISW=2: Obtain the value of the cumulative distribution function $Q(\chi^2\|\nu)$ for XO |
| 5 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{0, 1, 2\}$

(b) N $\geq 1$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | XI $\leq 0.0$ | 0.0 or 1.0 is set for XO. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |

(6) **Notes**

(a) From the relational expression $P(\chi^2|\nu)+Q(\chi^2|\nu) = 1$, it is possible to obtain either $P(\chi^2|\nu)$ or $Q(\chi^2|\nu)$ from the other. However, there may be times when cancellation of significant digits occurs, preventing good precision from being obtained.

(b) The mean and variance of a $\chi^2$ distribution having $n$ degrees of freedom are given by the following equations.

$$E[\chi^2(n)] = n, \sigma^2[\chi^2(n)] = 2n$$

(c) If $u$ is the random variable that obeys the standard normal distribution $N(0, 1)$, then the distribution of $u^2$ is the $\chi^2$ distribution having 1 degree of freedom.

(d) If $X_i(i = 1, \cdots, n)$ are the random variables of an arbitrary sample of size $n$ extracted from a normal population $(N(\mu, \sigma^2))$ having mean $\mu$ and variance $\sigma^2$, the following expressions

$$\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} \text{ and } \frac{n(\overline{X} - \mu)^2}{\sigma^2}$$

obey $\chi^2$ distributions having $n - 1$ and 1 degrees of freedom respectively, and they are mutually independent.

(7) **Example**

(a) Problem

Let $\chi^2 = 5.0$ and $\nu = 2$ and obtain the values of the probability density function $f(\chi^2|\nu)$ and the cumulative distribution functions $P(\chi^2|\nu)$ and $Q(\chi^2|\nu)$.

(b) Input data

XI = 5.0 and N = 2.

(c) Main program

```
      PROGRAM B1CDCH
! *** EXAMPLE OF D1CDCH ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER N,ISW
      REAL(8) XI,XO
!
      N=2
      XI=5.0
      WRITE(6,1000)
      WRITE(6,2000) N
      WRITE(6,2010) XI
      WRITE(6,3000)
      ISW=0
      CALL D1CDCH(N,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) XO
!
      ISW=1
      CALL D1CDCH(N,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDCH(N,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDCH ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'N  = ',I3)
 2010 FORMAT(9X,'XI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'VALUE OF P.D.F = ',D17.10)
 5010 FORMAT(9X,'VALUE OF C.D.F(1) = ',D17.10)
 5020 FORMAT(9X,'VALUE OF C.D.F(2) = ',D17.10)
      END
```

(d) Output results

```
    *** D1CDCH ***
     ** INPUT **
        N  =   2
        XI =  5.0


     ** OUTPUT **
        IERR =    0
        VALUE OF P.D.F =  0.4104249931D-01
        IERR =    0
        VALUE OF C.D.F(1) =  0.9179150014D+00
        IERR =    0
        VALUE OF C.D.F(2) =  0.8208499862D-01
```

## 3.2.5 D1CDIC, R1CDIC Inverse of $\chi^2$ Distribution

(1) **Function**

When given the cumulative distribution function (c.d.f.) $P(\chi^2|\nu)$ or $Q(\chi^2|\nu)$ of the $\chi^2$ distribution for which the number of degrees of freedom is $\nu$, the D1CDIC or R1CDIC obtains the frequency $\chi^2$ at that time. $P(\chi^2|\nu)$ and $Q(\chi^2|\nu)$ are defined by the following equations.

$$P(\chi^2|\nu) = \int_0^{\chi^2} \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}(t)^{\frac{\nu}{2}-1}e^{-\frac{t}{2}}dt \quad (0 \leq \chi^2 < \infty)$$

$$Q(\chi^2|\nu) = 1 - P(\chi^2|\nu) = \int_{\chi^2}^{\infty} \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}(t)^{\frac{\nu}{2}-1}e^{-\frac{t}{2}}dt \quad (0 \leq \chi^2 < \infty)$$

(2) **Usage**

Double precision:
CALL D1CDIC (N, XI, XO, ISW, IERR)
Single precision:
CALL R1CDIC (N, XI, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Value of number of degrees of freedom $\nu$ |
| 2 | XI | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Value of the cumulative distribution function $P(\chi^2\|\nu)$ or $Q(\chi^2\|\nu)$ of the $\chi^2$ distribution. |
| 3 | XO | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Output | Value of frequency $\chi^2$ |
| 4 | ISW | I | 1 | Input | ISW=1: Input the value of $P(\chi^2\|\nu)$ for XI<br>ISW=2: Input the value of $Q(\chi^2\|\nu)$ for XI |
| 5 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{1, 2\}$

(b) N $\geq 1$

(c) $0.0 \leq$ XI $\leq 1.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | XI=0.0 or XI=1.0 | 0.0 or the positive maximum value is set for XO. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3500 | The maximum number of iterations was reached before the specified precision was obtained. | The value at that time is returned. |
| 4000 | An error occurred in subroutine 3.2.4 $\left\{\begin{matrix}\text{D1CDCH}\\\text{R1CDCH}\end{matrix}\right\}$. | Processing is aborted. |

(6) **Notes**

(a) The mean and variance of a $\chi^2$ distribution having $n$degrees of freedom are given by the following equations.

$$E[\chi^2(n)] = n, \sigma^2[\chi^2(n)] = 2n$$

(b) If $u$ is the random variable that obeys the standard normal distribution $N(0,1)$, then the distribution of $u^2$ is the $\chi^2$ distribution having 1 degree of freedom.

(c) If $X_i (i = 1, \cdots, n)$ are the random variables of an arbitrary sample of size $n$ extracted from a normal population ($N(\mu, \sigma^2)$) having mean $\mu$ and variance $\sigma^2$, the following expressions

$$\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} \text{ and } \frac{n(\overline{X} - \mu)^2}{\sigma^2}$$

obey $\chi^2$ distributions having $n-1$ and 1 degrees of freedom respectively, and they are mutually independent.

(7) **Example**

(a) Problem
For $\nu = 2$, obtain the values of $\chi^2$ for which $P(\chi^2|\nu) = 0.2$ and $Q(\chi^2|\nu) = 0.2$ occur, respectively.

(b) Input data
XI = 0.2 and N = 2.

(c) Main program

```
      PROGRAM B1CDIC
! *** EXAMPLE OF D1CDIC ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER N,ISW
      REAL(8) XI,XO
!
      N=2
      XI=0.2
      WRITE(6,1000)
      WRITE(6,2000) N
      WRITE(6,2010) XI
      WRITE(6,3000)
!
      ISW=1
      CALL D1CDIC(N,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
```

```
!
      ISW=2
      CALL D1CDIC(N,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDIC ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'N  = ',I3)
 2010 FORMAT(9X,'XI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5010 FORMAT(9X,'VALUE X CORRESPONDING TO P(X,N)=XI: ',D17.10)
 5020 FORMAT(9X,'VALUE X CORRESPONDING TO Q(X,N)=XI: ',D17.10)
      END
```

(d) Output results

```
     *** D1CDIC ***
      ** INPUT **
         N  =   2
         XI =  0.2


      ** OUTPUT **
         IERR =    0
         VALUE X CORRESPONDING TO P(X,N)=XI:  0.4462871101D+00
         IERR =    0
         VALUE X CORRESPONDING TO Q(X,N)=XI:  0.3218875795D+01
```

### 3.2.6 D1CDNC, R1CDNC<br>Noncentral $\chi^2$ Distribution

(1) **Function**

For a noncentral $\chi^2$ distribution in which the frequency $\chi^2$ value is $x$, the number of degrees of freedom is $\nu$, and the noncentrality parameter is $\lambda$, the D1CDNC or R1CDNC obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(x|\nu,\lambda) = \begin{cases} \dfrac{e^{-\frac{(x+\lambda)}{2}} x^{\frac{(\nu-2)}{2}}}{2^{\frac{\nu}{2}}} \displaystyle\sum_{k=0}^{\infty} \frac{\lambda^k x^k}{2^{2k} k! \Gamma(\frac{\nu}{2}+k)} & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

(b) cumulative distribution function; c.d.f.

$$P(x|\nu,\lambda) = \begin{cases} \displaystyle\sum_{k=0}^{\infty} \frac{e^{-\frac{\lambda}{2}}(\frac{\lambda}{2})^k}{k!} \int_0^x \frac{t^{\frac{(\nu+2k)}{2}-1} e^{-\frac{t}{2}}}{2^{\frac{\nu+2k}{2}} \Gamma(\dfrac{\nu+2k}{2})} dt & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

(c) c.d.f.

$$Q(x|\nu,\lambda) = \begin{cases} 1 - P(x|\nu,\lambda) & (x > 0) \\ 1 & (x \leq 0) \end{cases}$$

(2) **Usage**

Double precision:

CALL D1CDNC (N, XL, XI, XO, ISW, IERR)

Single precision:

CALL R1CDNC (N, XL, XI, XO, ISW, IERR)

### (3) Arguments

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Value of number of degrees of freedom $\nu$. |
| 2 | XL | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Value of noncentrality parameter $\lambda$. |
| 3 | XI | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Value of frequency $\chi^2$. |
| 4 | XO | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Output | Value of the probability density function $f(x\|\nu,\lambda)$ or of the cumulative distribution function $P(x\|\nu,\lambda)$ or $Q(x\|\nu,\lambda)$ of the $\chi^2$ distribution. |
| 5 | ISW | I | 1 | Input | Processing switch<br>ISW=0:Obtain the value of the probability density function $f(x\|\nu,\lambda)$ for XO<br>ISW=1:Obtain the value of the cumulative distribution function $P(x\|\nu,\lambda)$ for XO<br>ISW=2:Obtain the value of the cumulative distribution function $Q(x\|\nu,\lambda)$ for XO |
| 6 | IERR | I | 1 | Output | Error indicator |

### (4) Restrictions

(a) ISW $\in \{0,1,2\}$

(b) N $\geq 1$

(c) XL $\geq 0.0$

### (5) Error indicator

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | XI $\leq 0.0$ | 0.0 or 1.0 is set for XO. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) or (c) was not satisfied. | |
| 3500 | The maximum number of iterations was reached before the specified precision was obtained. | The value at that time is returned. |

(6) **Notes**

(a) If the number of degrees of freedom $\nu$ or the noncentrality parameter $\lambda$ is greater than or equal to 1000, the function value may not be obtained.

(b) From the relational expression $P(x|\nu,\lambda) + Q(x|\nu,\lambda) = 1$, it is possible to obtain either $P(x|\nu,\lambda)$ or $Q(x|\nu,\lambda)$ from the other. However, there may be times when cancellation of significant digits occurs, preventing good precision from being obtained.

(c) The mean and variance of a noncentral $\chi^2$ distribution having degrees of freedom $\nu$ and noncentrality parameter $\lambda$ are given by the following equations.

$$E[\chi^{'2}(\nu,\lambda)] = a \quad , \quad \sigma^2[\chi^{'2}(\nu,\lambda)] = 2a(1+b)$$

Here, $a$ and $b$ are as follows.

$$a = \nu + \lambda \quad , \quad b = \frac{\lambda}{\nu+\lambda}$$

(d) A noncentral $\chi^2$ distribution having noncentrality parameter $\lambda = 0.0$ matches a $\chi^2$ distribution.

(e) If $X_i (i = 1, \cdots, n)$ are the random variables extracted from a normal population $(N_i(\mu_i, {\sigma_i}^2 = 1))$ having mean $\mu_i$ and variance ${\sigma_i}^2 = 1$, respectively, and if $Z$ is defined as follows,

$$Z = \sum_{i=0}^{n} {X_i}^2$$

then $Z$ obeys a noncentral $\chi^2$ distribution having degrees of freedom $n$ and noncentrality parameter $\lambda$, where $\lambda$ is given by the following equation.

$$\lambda = \sum_{i=0}^{n} {\mu_i}^2$$

(7) **Example**

(a) Problem

Let $\nu = 2$, $\lambda = 1.0$, and $x = 5.0$ and obtain the values of the probability density function $f(x|\nu,\lambda)$, and of the cumulative distribution functions $P(x|\nu,\lambda)$ and $Q(x|\nu,\lambda)$.

(b) Input data

N = 2, XL = 1.0 and XI = 5.0.

(c) Main program

```
      PROGRAM B1CDNC
! *** EXAMPLE OF D1CDNC ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER N,IERR
      INTEGER ISW
      REAL(8) XL,XI,XO
!
      N=2
      XL=1.0
      XI=5.0
      WRITE(6,1000)
      WRITE(6,2000) N,XL
      WRITE(6,2010) XI
      WRITE(6,3000)
      ISW=0
      CALL D1CDNC(N,XL,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) XO
!
      ISW=1
      CALL D1CDNC(N,XL,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDNC(N,XL,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
```

```
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDNC ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'N = ',I4,'  XL = ',F4.1)
 2010 FORMAT(9X,'XI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'VALUE OF P.D.F = ',D17.10)
 5010 FORMAT(9X,'VALUE OF C.D.F(1) = ',D17.10)
 5020 FORMAT(9X,'VALUE OF C.D.F(2) = ',D17.10)
      END
```

(d) Output results

```
     *** D1CDNC ***
      ** INPUT **
         N =    2  XL =  1.0
         XI =  5.0


      ** OUTPUT **
         IERR =    0
         VALUE OF P.D.F =  0.6719602550D-01
         IERR =    0
         VALUE OF C.D.F(1) =  0.8107099626D+00
         IERR =    0
         VALUE OF C.D.F(2) =  0.1892900374D+00
```

### 3.2.7 D1CDIX, R1CDIX
### Inverse Noncentral $\chi^2$ Distribution

(1) **Function**

Given the cumulative distribution function (c.d.f.) $P(x|\nu,\lambda)$ or $Q(x|\nu,\lambda)$ of a noncentral $\chi^2$ distribution for which the number of degrees of freedom is $\nu$ and the noncentrality parameter is $\lambda$, the D1CDIX or R1CDIX obtains the value $x$ of the frequency $\chi^2$ at that time. $P(x|\nu,\lambda)$ and $Q(x|\nu,\lambda)$ are defined by the following equations.

$$P(x|\nu,\lambda) = \sum_{k=0}^{\infty} \frac{e^{-\frac{\lambda}{2}}(\frac{\lambda}{2})^k}{k!} \int_0^x \frac{t^{\frac{(\nu+2k)}{2}-1} e^{-\frac{t}{2}}}{2^{\frac{\nu+2k}{2}} \Gamma(\frac{\nu+2k}{2})} dt \quad (0 \le x < \infty)$$

$$Q(x|\nu,\lambda) = 1 - P(x|\nu,\lambda) \quad (0 \le x < \infty)$$

(2) **Usage**

Double precision:
  CALL D1CDIX (N, XL, XI, XO, ISW, IERR)
Single precision:
  CALL R1CDIX (N, XL, XI, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Value of number of degrees of freedom $\nu$. |
| 2 | XL | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Value of noncentrality parameter $\lambda$. |
| 3 | XI | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Value of the cumulative distribution function $P(x|\nu,\lambda)$ or $Q(x|\nu,\lambda)$. |
| 4 | XO | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Output | Value of frequency $\chi^2$. |
| 5 | ISW | I | 1 | Input | Processing switch<br>ISW=1:Enter the value of the cumulative distribution function $P(x|\nu,\lambda)$ for XI<br>ISW=2:Enter the value of the cumulative distribution function $Q(x|\nu,\lambda)$ for XI |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{1, 2\}$

(b) N $\geq 1$

(c) $\lambda \geq 0.0$

(d) $0.0 \leq$ XI $\leq 1.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | XI = 0.0 or XI = 1.0 | 0.0 or the maximum value is set for XO. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) or (c) was not satisfied. | |
| 3020 | Restriction (d) was not satisfied. | |
| 3500 | The upper bound could not be found by the bisection method. | The maximum value is set for XO. |
| 3600 | The maximum number of iterations was reached before the specified precision was obtained. | The value at that time is returned. |
| 4000 | An error occurred in subroutine 3.2.6 $\left\{\begin{matrix}\text{D1CDNC}\\ \text{R1CDNC}\end{matrix}\right\}$. | Processing is aborted. |

(6) **Notes**

(a) The mean and variance of a noncentral $\chi^2$ distribution having degrees of freedom $\nu$ and noncentrality parameter $\lambda$ are given by the following equations.

$$E[{\chi'}^2(\nu, \lambda)] = a \quad , \quad \sigma^2[{\chi'}^2(\nu, \lambda)] = 2a(1 + b)$$

Here, $a$ and $b$ are as follows.

$$a = \nu + \lambda \quad , \quad b = \frac{\lambda}{\nu + \lambda}$$

(b) A noncentral $\chi^2$ distribution having noncentrality parameter $\lambda = 0.0$ matches a $\chi^2$ distribution.

(c) If $X_i (i = 1, \cdots, n)$ are the random variables extracted from a normal population $(N_i(\mu_i, {\sigma_i}^2 = 1))$ having mean $\mu_i$ and variance ${\sigma_i}^2 = 1$, respectively, and if $Z$ is defined as follows,

$$Z = \sum_{i=0}^{n} {X_i}^2$$

then $Z$ obeys a noncentral $\chi^2$ distribution having degrees of freedom $n$ and noncentrality parameter $\lambda$, where $\lambda$ is given by the following equation.

$$\lambda = \sum_{i=0}^{n} {\mu_i}^2$$

(7) **Example**

(a) Problem

Let $\nu = 2$ and $\lambda = 1.0$ and obtain the value of $x$ for which the cumulative distribution functions satisfy $P(x|\nu, \lambda) = 0.7$ and $Q(x|\nu, \lambda) = 0.7$.

(b) Input data

N = 2, XL = 1.0 and XI = 0.7.

(c) Main program

```
      PROGRAM B1CDIX
! *** EXAMPLE OF D1CDIX ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER N,IERR
      INTEGER ISW
      REAL(8) XL,XI,XO
!
      N=2
      XL=1.0
      XI=0.7
      WRITE(6,1000)
      WRITE(6,2000) N,XL
      WRITE(6,2010) XI
      WRITE(6,3000)
!
      ISW=1
      CALL D1CDIX(N,XL,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDIX(N,XL,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDIX ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'N = ',I4,'  XL = ',F4.1)
 2010 FORMAT(9X,'XI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5010 FORMAT(9X,&
      'VALUE OF X CORRESPONDING TO P(X;N,XL) = XI:',D17.10)
 5020 FORMAT(9X,&
      'VALUE OF X CORRESPONDING TO Q(X;N,XL) = XI:',D17.10)
      END
```

(d) Output results

```
     *** D1CDIX ***
      ** INPUT **
         N =    2  XL =  1.0
         XI =  0.7


      ** OUTPUT **
         IERR =    0
         VALUE OF X CORRESPONDING TO P(X;N,XL) = XI: 0.3685399605D+01
         IERR =    0
         VALUE OF X CORRESPONDING TO Q(X;N,XL) = XI: 0.1144032357D+01
```

### 3.2.8 D1CDTB, R1CDTB
$t$ **Distribution**

(1) **Function**

For a $t$ distribution having frequency $t$ and number of degrees of freedom $\nu$, the D1CDTB or R1CDTB obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(t|\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})(1+\frac{t^2}{\nu})^{\frac{\nu+1}{2}}}$$

(b) cumulative distribution function; c.d.f.

$$P(t|\nu) = \int_{-\infty}^{t} \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})(1+\frac{x^2}{\nu})^{\frac{\nu+1}{2}}} dx$$

(c) c.d.f.

$$Q(t|\nu) = \int_{t}^{\infty} \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})(1+\frac{x^2}{\nu})^{\frac{\nu+1}{2}}} dx$$

(2) **Usage**

Double precision:

CALL D1CDTB (N, TI, TO, ISW, IERR)

Single precision:

CALL R1CDTB (N, TI, TO, ISW, IERR)

### (3) **Arguments**

| | | |
|---|---|---|
| D:Double precision real | Z:Double precision complex | I: $\left\{ \begin{array}{l} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{array} \right\}$ |
| R:Single precision real | C:Single precision complex | |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Value of number of degrees of freedom $\nu$ |
| 2 | TI | $\left\{\begin{array}{c} \text{D} \\ \text{R} \end{array}\right\}$ | 1 | Input | Value of frequency $t$ |
| 3 | TO | $\left\{\begin{array}{c} \text{D} \\ \text{R} \end{array}\right\}$ | 1 | Output | Value of the probability density function $f(t\|\nu)$ of the $t$ distribution or of the cumulative distribution function $P(t\|\nu)$ or $Q(t\|\nu)$ of the $t$ distribution. |
| 4 | ISW | I | 1 | Input | ISW=0: Obtain the value of the probability density function $f(t\|\nu)$ for TO<br>ISW=1: Obtain the value of the cumulative distribution function $P(t\|\nu)$ for TO<br>ISW=2: Obtain the value of the cumulative distribution function $Q(t\|\nu)$ for TO |
| 5 | IERR | I | 1 | Output | Error indicator |

### (4) **Restrictions**

(a) N$\geq$1

(b) ISW$\in${0, 1, 2}

### (5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3100 | Restriction (b) was not satisfied. | |
| 4000 | An overflow occurred during the calculation (when (ISW=0). | |

(6) **Notes**

(a) If $u$ is the random variable that obeys the standard normal distribution $N(0,1)$, $\chi^2$ is the random variable that obeys the $\chi^2$ distribution having $\nu$ degrees of freedom, and $u$ and $\chi^2$ are mutually independent, then the distribution of the random variable $t$ given by the following equation obeys $t$ distribution having $\nu$ degrees of freedom.

$$t = \frac{u}{\sqrt{\frac{\chi^2}{\nu}}}$$

(7) **Example**

(a) Problem
Let $t$=5.0 and $\nu$=2 and obtain the values of the probability density function $f(t|\nu)$ and the cumulative distribution functions $P(t|\nu)$ and $Q(t|\nu)$.

(b) Input data
TI=5.0 and N=2.

(c) Main program

```
      PROGRAM B1CDTB
! *** EXAMPLE OF D1CDTB ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR,N,ISW
      REAL(8) TI,TO
!
      N=2
      TI=5.0
      WRITE(6,1000)
      WRITE(6,2000) N
      WRITE(6,2010) TI
      WRITE(6,3000)
      ISW=0
      CALL D1CDTB(N,TI,TO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) TO
!
      ISW=1
      CALL D1CDTB(N,TI,TO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) TO
!
      ISW=2
      CALL D1CDTB(N,TI,TO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) TO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDTB ***',/,&
             6X,'** INPUT **')
 2000 FORMAT(9X,'N  = ',I3)
 2010 FORMAT(9X,'TI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'VALUE OF P.D.F. = ',D17.10)
 5010 FORMAT(9X,'VALUE OF C.D.F.(1) = ',D17.10)
 5020 FORMAT(9X,'VALUE OF C.D.F.(2) = ',D17.10)
      END
```

(d) Output results

```
     *** D1CDTB ***
      ** INPUT **
         N  =   2
         TI =  5.0


      ** OUTPUT **
         IERR =    0
         VALUE OF P.D.F. =  0.7127781101D-02
         IERR =    0
         VALUE OF C.D.F.(1) =  0.9811252243D+00
         IERR =    0
         VALUE OF C.D.F.(2) =  0.1887477568D-01
```

## 3.2.9 D1CDIT, R1CDIT
## Inverse of $t$ Distribution

(1) **Function**

When given the cumulative distribution function (c.d.f.) $P(t|\nu)$ or $Q(t|\nu)$ of the $t$ distribution for which the number of degrees of freedom is $\nu$, the D1CDIT or R1CDIT obtains the frequency t at that time. $P(t|\nu)$ and $Q(t|\nu)$ are defined by the following equations.

$$P(t|\nu) = \int_{-\infty}^{t} \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})(1+\frac{x^2}{\nu})^{\frac{\nu+1}{2}}} dx$$

$$Q(t|\nu) = \int_{t}^{\infty} \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})(1+\frac{x^2}{\nu})^{\frac{\nu+1}{2}}} dx$$

(2) **Usage**

Double precision:
CALL D1CDIT (N, TI, TO, ISW, IERR)
Single precision:
CALL R1CDIT (N, TI, TO, ISW, IERR)

(3) **Arguments**

D:Double precision real　Z:Double precision complex　I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real　C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Value of number of degrees of freedom $\nu$ |
| 2 | TI | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Value of the cumulative distribution function $P(t\|\nu)$ or $Q(t\|\nu)$ of the $t$ distribution. |
| 3 | TO | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Output | Value of frequency $t$ |
| 4 | ISW | I | 1 | Input | ISW=1: Input the value of $P(t\|\nu)$ for TI<br>ISW=2: Input the value of $Q(t\|\nu)$ for TI |
| 5 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) 0.0≤TI≤1.0
(b) N≥1
(c) ISW∈{1, 2}

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | TI=0.0 was specified. | Processing continues with the negative minimum value or positive maximum value set for TO. |
| 1100 | TI=1.0 was specified. | Processing continues with the positive maximum value or negative minimum value set for TO. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3100 | Restriction (b) was not satisfied. | |
| 3200 | Restriction (c) was not satisfied. | |

(6) **Notes**

(a) To obtain the percentage point for a two-tailed probability of a $t$ distribution, assign a value that is 1/2 of the two-tailed probability for P.

(b) If $u$ is the random variable that obeys the standard normal distribution $N(0,1)$, $\chi^2$ is the random variable that obeys the $\chi^2$ distribution having $\nu$ degrees of freedom, and $u$ and $\chi^2$ are mutually independent, then the distribution of the random variable $t$ given by the following equation obeys $t$ distribution having $\nu$ degrees of freedom.

$$t = \frac{u}{\sqrt{\frac{\chi^2}{\nu}}}$$

(7) **Example**

(a) Problem

For $\nu$=2, obtain the values of $t$ for which $P(t|\nu)$=0.2 and $Q(t|\nu)$=0.2 occur, respectively.

(b) Input data

TI=0.2 and N=2.

(c) Main program

```
      PROGRAM B1CDIT
! *** EXAMPLE OF D1CDIT ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR,N,ISW
      REAL(8) TI,TO
!
      N=2
      TI=0.2
      WRITE(6,1000)
      WRITE(6,2000) N
      WRITE(6,2010) TI
      WRITE(6,3000)
      ISW=1
      CALL D1CDIT(N,TI,TO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) TO
!
      ISW=2
      CALL D1CDIT(N,TI,TO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) TO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDIT ***',/,&
             6X,'** INPUT **')
 2000 FORMAT(9X,'N  = ',I3)
 2010 FORMAT(9X,'TI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
```

```
5010 FORMAT(9X,'VALUE T CORRESPONDING TO P(X,N)=TI:',D17.10)
5020 FORMAT(9X,'VALUE T CORRESPONDING TO Q(X,N)=TI:',D17.10)
     END
```

(d) Output results

```
*** D1CDIT ***
 ** INPUT **
    N  =   2
    TI =  0.2

 ** OUTPUT **
    IERR =    0
    VALUE T CORRESPONDING TO P(X,N)=TI:-0.1060660155D+01
    IERR =    0
    VALUE T CORRESPONDING TO Q(X,N)=TI: 0.1060660155D+01
```

### 3.2.10 D1CDNT, R1CDNT Noncentral $t$ Distribution

(1) **Function**

For a noncentral $t$ distribution in which the frequency is $t$, the number of degrees of freedom is $\nu$, and the noncentrality parameter is $\delta$, the D1CDNT or R1CDNT obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(t|\nu,\delta) = \frac{\nu^{\frac{\nu}{2}} e^{-\frac{\delta^2}{2}}}{\sqrt{\pi}\Gamma(\frac{\nu}{2})(\nu+t^2)^{\frac{(\nu+1)}{2}}} \sum_{k=0}^{\infty} \Gamma(\frac{\nu+k+1}{2}) \frac{\delta^k}{k!} (\frac{2t^2}{\nu+t^2})^{\frac{k}{2}}$$

(b) cumulative distribution function; c.d.f.

$$P(t|\nu,\delta) = \int_{-\infty}^{t} \frac{\nu^{\frac{\nu}{2}} e^{-\frac{\delta^2}{2}}}{\sqrt{\pi}\Gamma(\frac{\nu}{2})(\nu+x^2)^{\frac{(\nu+1)}{2}}} \sum_{k=0}^{\infty} \Gamma(\frac{\nu+k+1}{2}) \frac{\delta^k}{k!} (\frac{2x^2}{\nu+x^2})^{\frac{k}{2}} dx$$

(c) c.d.f.

$$Q(t|\nu,\delta) = 1 - P(t|\nu,\delta)$$

(2) **Usage**

Double precision:
  CALL D1CDNT (N, DEL, XI, XO, ISW, IERR)
Single precision:
  CALL R1CDNT (N, DEL, XI, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{ \begin{array}{l} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{array} \right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Value of number of degrees of freedom $\nu$. |
| 2 | DEL | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 1 | Input | Value of noncentrality parameter $\delta$ |
| 3 | XI | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 1 | Input | Value of frequency $t$ |
| 4 | XO | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 1 | Output | Value of the probability density function $f(t\|\nu,\delta)$ or of the cumulative distribution function $P(t\|\nu,\delta)$ or $Q(t\|\nu,\delta)$ of the $t$ distribution. |
| 5 | ISW | I | 1 | Input | Processing switch<br>ISW=0:Obtain the value of the probability density function $f(t\|\nu,\delta)$ for XO<br>ISW=1:Obtain the value of the cumulative distribution function $P(t\|\nu,\delta)$ for XO<br>ISW=2:Obtain the value of the cumulative distribution function $Q(t\|\nu,\delta)$ for XO |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{0, 1, 2\}$

(b) N $\geq 1$

(c) DEL $\geq 0.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | XI $\leq 0.0$ | 0.0 or 1.0 is set for XO. |
| 2000 | When ISW $= 0$, a solution having sufficient precision could not be found. | The value of probability density function at that time is returned. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) or (c) was not satisfied. | |

(6) **Notes**

(a) From the relational expression $P(t|\nu,\delta) + Q(t|\nu,\delta) = 1$, it is possible to obtain either $P(t|\nu,\delta)$ or $Q(t|\nu,\delta)$ from the other. However, there may be times when cancellation of significant digits occurs, preventing good precision from being obtained.

(b) A noncentral $t$ distribution having noncentrality parameter $\delta = 0.0$ matches a $t$ distribution.

(7) **Example**

(a) Problem
Let $\nu = 2$, $\delta = 1.0$ and $x = 5.0$ and obtain the values of the probability density function $f(t|\nu,\delta)$ and of the cumulative distribution functions $P(t|\nu,\delta)$ and $Q(t|\nu,\delta)$.

(b) Input data
N = 2, DEL = 1.0 and XI = 5.0.

(c) Main program

```
      PROGRAM B1CDNT
! *** EXAMPLE OF D1CDNT ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER N,IERR
      INTEGER ISW
      REAL(8) DEL,XI,XO
!
      N=2
      DEL=1.0
      XI=5.0
      WRITE(6,1000)
      WRITE(6,2000) N,DEL
      WRITE(6,2010) XI
      WRITE(6,3000)
      ISW=0
      CALL D1CDNT(N,DEL,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) XO
!
      ISW=1
      CALL D1CDNT(N,DEL,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDNT(N,DEL,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDNT ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'N = ',I4,'  DEL = ',F4.1)
 2010 FORMAT(9X,'XI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'VALUE OF P.D.F = ',D17.10)
 5010 FORMAT(9X,'VALUE OF C.D.F(1) = ',D17.10)
 5020 FORMAT(9X,'VALUE OF C.D.F(2) = ',D17.10)
      END
```

(d) Output results

```
     *** D1CDNT ***
      ** INPUT **
         N =    2  DEL =  1.0
         XI =  5.0


      ** OUTPUT **
         IERR =    0
         VALUE OF P.D.F =  0.2533236257D-01
         IERR =    0
         VALUE OF C.D.F(1) =  0.9301737669D+00
         IERR =    0
         VALUE OF C.D.F(2) =  0.6982623314D-01
```

### 3.2.11 D1CDIS, R1CDIS
### Inverse Noncentral $t$ Distribution

(1) **Function**

Given the cumulative distribution function (c.d.f.) $P(t|\nu,\delta)$ or $Q(t|\nu,\delta)$ of a noncentral $t$ distribution for which the number of degrees of freedom is $\nu$ and the noncentrality parameter is $\delta$, the D1CDIS or R1CDIS obtains the value of the frequency $t$ at that time. $P(t|\nu,\delta)$ and $Q(t|\nu,\delta)$ are defined by the following equations.

$$P(t|\nu,\delta) = \int_{-\infty}^{t} \frac{\nu^{\frac{\nu}{2}} e^{-\frac{\delta^2}{2}}}{\sqrt{\pi}\Gamma(\frac{\nu}{2})(\nu+x^2)^{\frac{(\nu+1)}{2}}} \sum_{k=0}^{\infty} \Gamma(\frac{\nu+k+1}{2}) \frac{\delta^k}{k!} (\frac{2x^2}{\nu+x^2})^{\frac{k}{2}} dx$$

$$Q(t|\nu,\delta) = 1 - P(t|\nu,\delta)$$

(2) **Usage**

Double precision:
CALL D1CDIS (N, DEL, XI, XO, ISW, IERR)
Single precision:
CALL R1CDIS (N, DEL, XI, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Value of number of degrees of freedom $\nu$. |
| 2 | DEL | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Value of noncentrality parameter $\delta$. |
| 3 | XI | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Value of the cumulative distribution function $P(t\|\nu,\delta)$ or $Q(t\|\nu,\delta)$ of the $t$ distribution. |
| 4 | XO | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Output | Value of frequency $t$. |
| 5 | ISW | I | 1 | Input | Processing switch<br>ISW=1:Input the value of the cumulative distribution function $P(t\|\nu,\delta)$ for XI<br>ISW=2:Input the value of the cumulative distribution function $Q(t\|\nu,\delta)$ for XI |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{0, 1, 2\}$

(b) N $\geq 1$

(c) $\delta \geq 0.0$

(d) $0.0 \leq$ XI $\leq 1.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | XI = 0.0 or XI = 1.0 | The minimum value or the maximum value is set for XO. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) or (c) was not satisfied. | |
| 3020 | Restriction (d) was not satisfied. | |
| 3510 | The lower bound could not be found by the bisection method. | The minimum value is set for XO. |
| 3520 | The upper bound could not be found by the bisection method. | The maximum value is set for XO. |
| 3600 | The maximum number of iterations was reached before the specified precision was obtained. | The value at that time is returned. |
| 4000 | An error occurred in subroutine 3.2.10 $\left\{\begin{array}{l}\text{D1CDNT}\\\text{R1CDNT}\end{array}\right\}$. | Processing is aborted. |

(6) **Notes**

(a) A noncentral $t$ distribution having noncentrality parameter $\delta = 0.0$ matches a $t$ distribution.

(7) **Example**

(a) Problem
Let $\nu = 2$ and $\delta = 1.0$ and obtain the value of $t$ for which the cumulative distribution functions satisfy $P(t|\nu, \delta) = 0.7$ and $Q(t|\nu, \delta) = 0.7$.

(b) Input data
N = 2, DEL = 1.0 and XI = 0.7.

(c) Main program

```
      PROGRAM B1CDIS
! *** EXAMPLE OF D1CDIS ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER N,IERR
      INTEGER ISW
      REAL(8) DEL,XI,XO
!
      N=2
      DEL=1.0
      XI=0.7
      WRITE(6,1000)
      WRITE(6,2000) N,DEL
      WRITE(6,2010) XI
      WRITE(6,3000)
!
      ISW=1
      CALL D1CDIS(N,DEL,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDIS(N,DEL,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDIS ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'N = ',I4,'  DEL = ',F4.1)
 2010 FORMAT(9X,'XI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5010 FORMAT(9X,&
      'VALUE OF X CORRESPONDING TO P(X;N,DEL) = XI:',D17.10)
 5020 FORMAT(9X,&
      'VALUE OF X CORRESPONDING TO Q(X;N,DEL) = XI:',D17.10)
      END
```

(d) Output results

```
    *** D1CDIS ***
     ** INPUT **
        N =    2  DEL =  1.0
        XI =  0.7


     ** OUTPUT **
        IERR =    0
        VALUE OF X CORRESPONDING TO P(X;N,DEL) = XI: 0.1963296797D+01
        IERR =    0
        VALUE OF X CORRESPONDING TO Q(X;N,DEL) = XI: 0.5208679536D+00
```

## 3.2.12 D1CDFB, R1CDFB
## *F* Distribution

(1) **Function**

For a $F$ distribution having frequency $F$ and numbers of degrees of freedom $\nu_1$ and $\nu_2$, the D1CDFB or R1CDFB obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(F|\nu_1,\nu_2) = \frac{\nu_1^{\frac{\nu_1}{2}} \cdot \nu_2^{\frac{\nu_2}{2}} \cdot x^{\frac{\nu_1}{2}-1}}{B(\frac{\nu_1}{2},\frac{\nu_2}{2})(\nu_1 x+\nu_2)^{\frac{\nu_1+\nu_2}{2}}}$$

(b) cumulative distribution function; c.d.f.

$$P(F|\nu_1,\nu_2) = \int_0^F \frac{\nu_1^{\frac{\nu_1}{2}} \cdot \nu_2^{\frac{\nu_2}{2}} \cdot x^{\frac{\nu_1}{2}-1}}{B(\frac{\nu_1}{2},\frac{\nu_2}{2})(\nu_1 x+\nu_2)^{\frac{\nu_1+\nu_2}{2}}} dx$$

(c) c.d.f.

$$Q(F|\nu_1,\nu_2) = \int_F^\infty \frac{\nu_1^{\frac{\nu_1}{2}} \cdot \nu_2^{\frac{\nu_2}{2}} \cdot x^{\frac{\nu_1}{2}-1}}{B(\frac{\nu_1}{2},\frac{\nu_2}{2})(\nu_1 x+\nu_2)^{\frac{\nu_1+\nu_2}{2}}} dx$$

(2) **Usage**

Double precision:

CALL D1CDFB (N1, N2, FI, FO, ISW, IERR)

Single precision:

CALL R1CDFB (N1, N2, FI, FO, ISW, IERR)

## (3) **Arguments**

D:Double precision real  Z:Double precision complex  I: { INTEGER(4) as for 32bit Integer ; INTEGER(8) as for 64bit Integer }
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N1 | I | 1 | Input | Number of degrees of freedom $\nu_1$ |
| 2 | N2 | I | 1 | Input | Number of degrees of freedom $\nu_2$ |
| 3 | FI | {D, R} | 1 | Input | Value of frequency $F$ |
| 4 | FO | {D, R} | 1 | Output | Value of the probability density function $f(F\|\nu_1, \nu_2)$ of the $F$ distribution or of the cumulative distribution function $P(F\|\nu_1, \nu_2)$ or $Q(F\|\nu_1, \nu_2)$ of the $F$ distribution. |
| 5 | ISW | I | 1 | Input | ISW=0: Obtain the value of the probability density function $f(F\|\nu_1, \nu_2)$ for FO<br>ISW=1: Obtain the value of the cumulative distribution function $P(F\|\nu_1, \nu_2)$ for FO<br>ISW=2: Obtain the value of the cumulative distribution function $Q(F\|\nu_1, \nu_2)$ for FO |
| 6 | IERR | I | 1 | Output | Error indicator |

## (4) **Restrictions**

(a) F>0.0

(b) N1≥1, N2≥1

(c) ISW∈{0, 1, 2}

## (5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 2000 | N1>2000, N2>2000 (when ISW=0) | Processing terminated, but the precision of the solution is low. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3100 | Restriction (b) was not satisfied. | |
| 3200 | Restriction (c) was not satisfied. | |
| 4000 | $\|B(\frac{n_1}{2}, \frac{n_2}{2})\|$ <(Positive minimum value) (when ISW=0) | |
| 4100 | An error occurred in the step for obtaining the beta function value. (when ISW=0) | |
| 4200 | An overflow occurred during the calculation. (when ISW=0) | |

(6) **Notes**

(a) If $\chi_1^2$ and $\chi_2^2$ are the random variables that obey the $\chi^2$ distribution having $\nu_1$ and $\nu_2$ degrees of freedom respectively and $\chi_1^2$ and $\chi_2^2$ are mutually independent, then the distribution of the random variable $F$ given by the following equation obeys $F$ distribution having $\nu_1$ and $\nu_2$ degrees of freedom.

$$F = \frac{\frac{\chi_1^2}{\nu_1}}{\frac{\chi_2^2}{\nu_2}}$$

(7) **Example**

(a) Problem
Let $F$=5.0, $\nu_1$=2 and $\nu_2$=2 and obtain the values of the probability density function $f(F|\nu_1,\nu_2)$ and the cumulative distribution functions $P(F|\nu_1,\nu_2)$ and $Q(F|\nu_1,\nu_2)$.

(b) Input data
FI=5.0, N1=2 and N2=2.

(c) Main program

```
      PROGRAM B1CDFB
! *** EXAMPLE OF D1CDFB ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR,N1,N2,ISW
      REAL(8) FI,FO
!
      N1=2
      N2=2
      FI=5.0
      WRITE(6,1000)
      WRITE(6,2000) N1
      WRITE(6,2010) N2
      WRITE(6,2020) FI
      WRITE(6,3000)
      ISW=0
      CALL D1CDFB(N1,N2,FI,FO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) FO
!
      ISW=1
      CALL D1CDFB(N1,N2,FI,FO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) FO
!
      ISW=2
      CALL D1CDFB(N1,N2,FI,FO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) FO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDFB ***',/,&
             6X,'** INPUT **')
 2000 FORMAT(9X,'N1 = ',I3)
 2010 FORMAT(9X,'N2 = ',I3)
 2020 FORMAT(9X,'FI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'VALUE OF P.D.F = ',D17.10)
 5010 FORMAT(9X,'VALUE OF C.D.F.(1) = ',D17.10)
 5020 FORMAT(9X,'VALUE OF C.D.F.(2) = ',D17.10)
      END
```

(d) Output results

```
*** D1CDFB ***
 ** INPUT **
    N1 =   2
    N2 =   2
    FI =  5.0


 ** OUTPUT **
    IERR =    0
    VALUE OF P.D.F =  0.2777777778D-01
    IERR =    0
    VALUE OF C.D.F.(1) =  0.8333333333D+00
    IERR =    0
    VALUE OF C.D.F.(2) =  0.1666666667D+00
```

## 3.2.13 D1CDIF, R1CDIF Inverse of $F$ Distribution

(1) **Function**

When given the cumulative distribution function (c.d.f.) $P(F|\nu_1,\nu_2)$ or $Q(F|\nu_1,\nu_2)$ of the $F$ distribution for which the numbers of degrees of freedom are $\nu_1$ and $\nu_2$, the D1CDIF or R1CDIF obtains the frequency $F$ at that time. $P(F|\nu_1,\nu_2)$ and $Q(F|\nu_1,\nu_2)$ are defined by the following equations.

$$P(F|\nu_1,\nu_2) = \int_0^F \frac{\nu_1^{\frac{\nu_1}{2}} \cdot \nu_2^{\frac{\nu_2}{2}} \cdot x^{\frac{\nu_1}{2}-1}}{B(\frac{\nu_1}{2},\frac{\nu_2}{2})(\nu_1 x+\nu_2)^{\frac{\nu_1+\nu_2}{2}}}dx$$

$$Q(F|\nu_1,\nu_2) = \int_F^\infty \frac{\nu_1^{\frac{\nu_1}{2}} \cdot \nu_2^{\frac{\nu_2}{2}} \cdot x^{\frac{\nu_1}{2}-1}}{B(\frac{\nu_1}{2},\frac{\nu_2}{2})(\nu_1 x+\nu_2)^{\frac{\nu_1+\nu_2}{2}}}dx$$

(2) **Usage**

Double precision:

CALL D1CDIF (N1, N2, FI, FO, ISW, IERR)

Single precision:

CALL R1CDIF (N1, N2, FI, FO, ISW, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N1 | I | 1 | Input | Number of degrees of freedom $\nu_1$ |
| 2 | N2 | I | 1 | Input | Number of degrees of freedom $\nu_2$ |
| 3 | FI | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Value of the cumulative distribution function $P(F|\nu_1,\nu_2)$ or $Q(F|\nu_1,\nu_2)$ of the $F$ distribution. |
| 4 | FO | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Output | Value of frequency $F$ |
| 5 | ISW | I | 1 | Input | ISW=1: Assign the value of $P(F|\nu_1,\nu_2)$ for FI<br>ISW=2: Assign the value of $Q(F|\nu_1,\nu_2)$ for FI |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) 0.0≤FI≤1.0

(b) N1≥1, N2≥1

(c) ISW∈{1, 2}

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | FI=0.0 was specified. | Processing continues with 0.0 or the positive maximum value set for FO. |
| 1100 | FI=1.0 was specified. | Processing continues with the positive maximum value or 0.0 set for FO. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3100 | Restriction (b) was not satisfied. | |
| 3200 | Restriction (c) was not satisfied. | |

(6) **Notes**

(a) If $\chi_1^2$ and $\chi_2^2$ are the random variables that obey the $\chi^2$ distribution having $\nu_1$ and $\nu_2$ degrees of freedom respectively and $\chi_1^2$ and $\chi_2^2$ are mutually independent, then the distribution of the random variable $F$ given by the following equation obeys $F$ distribution having $\nu_1$ and $\nu_2$ degrees of freedom.

$$F = \frac{\frac{\chi_1^2}{\nu_1}}{\frac{\chi_2^2}{\nu_2}}$$

(7) **Example**

(a) Problem

For $\nu_1$=2 and $\nu_2$=2 obtain the values of $F$ for which $P(F|\nu_1,\nu_2)$=0.2 and $Q(F|\nu_1,\nu_2)$=0.2 occur, respectively.

(b) Input data

FI=0.2, N1=2 and N2=2.

(c) Main program

```
      PROGRAM B1CDIF
! *** EXAMPLE OF D1CDIF ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR,N1,N2,ISW
      REAL(8) FI,FO
!
      N1=2
      N2=2
      FI=0.2
      WRITE(6,1000)
      WRITE(6,2000) N1
      WRITE(6,2010) N2
      WRITE(6,2020) FI
      WRITE(6,3000)
      ISW=1
      CALL D1CDIF(N1,N2,FI,FO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) FO
!
      ISW=2
      CALL D1CDIF(N1,N2,FI,FO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) FO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDIF ***',/,&
              6X,'** INPUT **')
 2000 FORMAT(9X,'N1 = ',I3)
 2010 FORMAT(9X,'N2 = ',I3)
 2020 FORMAT(9X,'FI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5010 FORMAT(9X,'VALUE F CORRESPONDING TO P(X,N)=FI:',D17.10)
 5020 FORMAT(9X,'VALUE F CORRESPONDING TO Q(X,N)=FI:',D17.10)
      END
```

(d) Output results

```
*** D1CDIF ***
 ** INPUT **
    N1 =   2
    N2 =   2
    FI =  0.2


 ** OUTPUT **
    IERR =    0
    VALUE F CORRESPONDING TO P(X,N)=FI: 0.2500000047D+00
    IERR =    0
    VALUE F CORRESPONDING TO Q(X,N)=FI: 0.3999999925D+01
```

### 3.2.14 D1CDGM, R1CDGM Gamma Distribution

(1) **Function**

For a gamma distribution having parameters $\alpha$ and $\beta$, the D1CDGM or R1CDGM obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(x;\alpha,\beta)=\begin{cases}\dfrac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} & (x>0;\alpha,\beta>0)\\ 0 & (x\le 0;\alpha,\beta>0)\end{cases}$$

(b) cumulative distribution function; c.d.f.

$$P(x;\alpha,\beta)=\int_0^x \frac{\beta^{\alpha}}{\Gamma(\alpha)}t^{\alpha-1}e^{-\beta t}dt \quad (\alpha,\beta>0)$$

(c) c.d.f.

$$Q(x;\alpha,\beta)=1-P(x;\alpha,\beta)=\int_x^{\infty} \frac{\beta^{\alpha}}{\Gamma(\alpha)}t^{\alpha-1}e^{-\beta t}dt \quad (\alpha,\beta>0)$$

(2) **Usage**

Double precision:

CALL D1CDGM (A, B, XI, XO, ISW, IERR)

Single precision:

CALL R1CDGM (A, B, XI, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{matrix}\text{D}\\ \text{R}\end{matrix}\right\}$ | 1 | Input | Value of shape parameter $\alpha$ |
| 2 | B | $\left\{\begin{matrix}\text{D}\\ \text{R}\end{matrix}\right\}$ | 1 | Input | Value of scale parameter $\beta$ |
| 3 | XI | $\left\{\begin{matrix}\text{D}\\ \text{R}\end{matrix}\right\}$ | 1 | Input | Value of random variable $x$ |
| 4 | XO | $\left\{\begin{matrix}\text{D}\\ \text{R}\end{matrix}\right\}$ | 1 | Output | Value of the probability density function $f(x;\alpha,\beta)$ of the gamma distribution or of the cumulative distribution function $P(x;\alpha,\beta)$ or $Q(x;\alpha,\beta)$ of the gamma distribution. |
| 5 | ISW | I | 1 | Input | ISW=0: Obtain the value of the probability density function $f(x;\alpha,\beta)$ for XO<br>ISW=1: Obtain the value of the cumulative distribution function $P(x;\alpha,\beta)$ for XO<br>ISW=2: Obtain the value of the cumulative distribution function $Q(x;\alpha,\beta)$ for XO |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{0, 1, 2\}$

(b) A $> 0.0$

(c) B $> 0.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | XI $\leq 0.0$ | 0.0 or 1.0 is set for XO. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |

(6) **Notes**

(a) From the relational expression $P(x;\alpha,\beta)+Q(x;\alpha,\beta)=1$, it is possible to obtain either $P(x;\alpha,\beta)$ or $Q(x;\alpha,\beta)$ from the other. However, there may be times when cancellation of significant digits occurs, preventing good precision from being obtained.

(b) The mean and variance of a gamma distribution having parameters $\alpha$ and $\beta$ are given by the following equations.

$$E[x]=\frac{\alpha}{\beta},\sigma^2[x]=\frac{\alpha}{\beta^2}$$

(c) A gamma distribution for which $\alpha = 1$ is an exponential distribution. Also, a distribution for which the value of $\alpha$ is limited to a positive integer is called an Erlang distribution.

(7) **Example**

(a) Problem
Let $\alpha = 5.0$, $\beta = 2.0$ and $x = 3.0$ and obtain the values of the probability density function $f(x;\alpha,\beta)$ and the cumulative distribution functions $P(x;\alpha,\beta)$ and $Q(x;\alpha,\beta)$.

(b) Input data
A = 5.0, B = 2.0 and XI = 3.0.

(c) Main program

```
      PROGRAM B1CDGM
! *** EXAMPLE OF D1CDGM ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER ISW
      REAL(8) XI,XO
!
      A=5.0D0
      B=2.0D0
      XI=3.0D0
      WRITE(6,1000)
      WRITE(6,2000) A
      WRITE(6,2010) B
      WRITE(6,2020) XI
      WRITE(6,3000)
      ISW=0
      CALL D1CDGM(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) XO
!
      ISW=1
      CALL D1CDGM(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDGM(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDGM ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'A  = ',F4.1)
 2010 FORMAT(9X,'B  = ',F4.1)
 2020 FORMAT(9X,'XI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'VALUE OF P.D.F = ',D17.10)
 5010 FORMAT(9X,'VALUE OF C.D.F(1) = ',D17.10)
 5020 FORMAT(9X,'VALUE OF C.D.F(2) = ',D17.10)
      END
```

(d) Output results

```
     *** D1CDGM ***
      ** INPUT **
         A  =  5.0
         B  =  2.0
         XI =  3.0


      ** OUTPUT **
```

```
IERR =    0
VALUE OF P.D.F =  0.2677052351D+00
IERR =    0
VALUE OF C.D.F(1) =  0.7149434997D+00
IERR =    0
VALUE OF C.D.F(2) =  0.2850565003D+00
```

## 3.2.15 D1CDIG, R1CDIG Inverse Gamma Distribution

(1) **Function**

Given the (cumulative distribution function; c.d.f.) $P(x;\alpha,\beta)$ or $Q(x;\alpha,\beta)$ of a gamma distribution having parameters $\alpha$ and $\beta$, the D1CDIG or R1CDIG obtains the value $x$ of the random variable at that time. $P(x;\alpha,\beta)$ and $Q(x;\alpha,\beta)$ are defined by the following equations.

$$P(x;\alpha,\beta) = \int_0^x \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} dt \quad (\alpha,\beta > 0)$$

$$Q(x;\alpha,\beta) = 1 - P(x;\alpha,\beta) = \int_x^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} dt \quad (\alpha,\beta > 0)$$

(2) **Usage**

Double precision:
CALL D1CDIG (A, B, XI, XO, ISW, IERR)
Single precision:
CALL R1CDIG (A, B, XI, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: {INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer}

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | {D, R} | 1 | Input | Value of shape parameter $\alpha$. |
| 2 | B | {D, R} | 1 | Input | Value of scale parameter $\beta$. |
| 3 | XI | {D, R} | 1 | Input | Value of the cumulative distribution function $P(x;\alpha,\beta)$ or $Q(x;\alpha,\beta)$ of the gamma distribution. |
| 4 | XO | {D, R} | 1 | Output | Value of random variable $x$. |
| 5 | ISW | I | 1 | Input | Processing switch<br>ISW=1:Input the value of the cumulative distribution function $P(x;\alpha,\beta)$ for XI<br>ISW=2:Input the value of the cumulative distribution function $Q(x;\alpha,\beta)$ for XI |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{1, 2\}$

(b) $\mathrm{A}, \mathrm{B} > 0.0$

(c) $0.0 \leq \mathrm{XI} \leq 1.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | XI = 0.0 or XI = 1.0 | If XI = 0.0<br>When ISW=1: 0.0 is set for XO.<br>When ISW=2:The maximum value is set for XO.<br>If XI = 1.0<br>When ISW=1:The maximum value is set for XO.<br>When ISW=2:0.0 is set for XO. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3500 | The upper bound could not be found by the bisection method. | The maximum value is set for XO. |
| 3600 | The maximum number of iterations was reached before the specified precision was obtained. | The value at that time is returned. |
| 4000 | An error occurred in subroutine 3.2.14 $\left\{\begin{matrix}\text{D1CDGM}\\ \text{R1CDGM}\end{matrix}\right\}$. | Processing is aborted. |

(6) **Notes**

(a) The mean and variance of a gamma distribution having parameters $\alpha$ and $\beta$ are given by the following equations.

$$E[x] = \frac{\alpha}{\beta}, \sigma^2[x] = \frac{\alpha}{\beta^2}$$

(b) A gamma distribution with $\alpha = 1$ is an exponential distribution. A distribution for which the value of $\alpha$ is limited to positive integers is called an Erlang distribution.

(7) **Example**

(a) Problem

For $\alpha = 5.0$ and $\beta = 2.0$, obtain the values of $x$ for which the cumulative distribution function $P(x;\alpha,\beta) = 0.7$, or $Q(x;\alpha,\beta) = 0.7$, respectively.

(b) Input data

A = 5.0, B = 2.0 and XI = 0.7.

(c) Main program

```
      PROGRAM B1CDIG
! *** EXAMPLE OF D1CDIG ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER ISW
      REAL(8) A,B,XI,XO
!
      A=5.0D0
      B=2.0D0
      XI=0.7D0
      WRITE(6,1000)
      WRITE(6,2000) A,B
      WRITE(6,2010) XI
      WRITE(6,3000)
!
      ISW=1
      CALL D1CDIG(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDIG(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDIG ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'A = ',F4.1,'  B = ',F4.1)
 2010 FORMAT(9X,'XI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5010 FORMAT(9X,&
      'VALUE OF X CORRESPONDING TO P(X;ALPHA,BETA) = XI:',D17.10)
 5020 FORMAT(9X,&
      'VALUE OF X CORRESPONDING TO Q(X;ALPHA,BETA) = XI:',D17.10)
      END
```

(d) Output results

```
     *** D1CDIG ***
      ** INPUT **
         A =  5.0  B =  2.0
         XI =  0.7


      ** OUTPUT **
         IERR =    0
         VALUE OF X CORRESPONDING TO P(X;ALPHA,BETA) = XI: 0.2945180657D+01
         IERR =    0
         VALUE OF X CORRESPONDING TO Q(X;ALPHA,BETA) = XI: 0.1816804541D+01
```

### 3.2.16 D1CDBT, R1CDBT Beta Distribution

(1) **Function**

For a beta distribution having the two positive numbers $a$ and $b$ as parameters, the D1CDBT or R1CDBT obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(x;a,b) = \begin{cases} \dfrac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & (0 < x < 1; a, b > 0) \\ 0 & (x \le 0, x \ge 1; a, b > 0) \end{cases}$$

(b) cumulative distribution function; c.d.f.

$$P(x;a,b) = \begin{cases} 0 & (x \le 0; a, b > 0) \\ \dfrac{1}{B(a,b)} \displaystyle\int_0^x t^{a-1}(1-t)^{b-1} dt & (0 < x < 1; a, b > 0) \\ 1 & (x \ge 1; a, b > 0) \end{cases}$$

(c) c.d.f.

$$\begin{aligned} Q(x;a,b) &= 1 - P(x;a,b) \\ &= \begin{cases} 1 & (x \le 0; a, b > 0) \\ \dfrac{1}{B(a,b)} \displaystyle\int_x^1 t^{a-1}(1-t)^{b-1} dt & (0 < x < 1; a, b > 0) \\ 0 & (x \ge 1; a, b > 0) \end{cases} \end{aligned}$$

(2) **Usage**

Double precision:

CALL D1CDBT (A, B, XI, XO, ISW, IERR)

Single precision:

CALL R1CDBT (A, B, XI, XO, ISW, IERR)

## (3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\begin{Bmatrix}D\\R\end{Bmatrix}$ | 1 | Input | Value of shape parameter $a$. |
| 2 | B | $\begin{Bmatrix}D\\R\end{Bmatrix}$ | 1 | Input | Value of shape parameter $b$. |
| 3 | XI | $\begin{Bmatrix}D\\R\end{Bmatrix}$ | 1 | Input | Value of random variable $x$ |
| 4 | XO | $\begin{Bmatrix}D\\R\end{Bmatrix}$ | 1 | Output | Value of the probability density function $f(x;a,b)$ or of the cumulative distribution function $P(x;a,b)$ or $Q(x;a,b)$ of the beta distribution |
| 5 | ISW | I | 1 | Input | Processing switch<br>ISW=0:Obtain the value of the probability density function $f(x;a,b)$ for XO<br>ISW=1:Obtain the value of the cumulative distribution function $P(x;a,b)$ for XO<br>ISW=2:Obtain the value of the cumulative distribution function $Q(x;a,b)$ for XO |
| 6 | IERR | I | 1 | Output | Error indicator |

## (4) **Restrictions**

(a) ISW $\in \{0, 1, 2\}$

(b) $\mathrm{A}, \mathrm{B} > 0.0$

## (5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | XI $\leq 0.0$ or XI $\geq 1.0$ | 0.0 or 1.0 is set for XO. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3500 | The maximum number of iterations was reached before the specified precision was obtained. | The value at that time is returned. |

(6) **Notes**

(a) From the relational expression $P(x; a, b) + Q(x; a, b) = 1$, it is possible to obtain either $P(x; a, b)$ or $Q(x; a, b)$ from the other. However, there may be times when cancellation of significant digits occurs, preventing good precision from being obtained.

(b) The mean and variance of a beta distribution having parameters $a$ and $b$ are given by the following equations.

$$E[x] = \frac{a}{a+b}, \quad \sigma^2[x] = \frac{ab}{(a+b)^2(a+b+1)}$$

(c) A beta distribution with $a = b = 1$ is a uniform distribution on the interval (0, 1).

(7) **Example**

(a) Problem
Let $a = 5.0$, $b = 2.0$ and $x = 0.3$ and obtain the values of the probability density function $f(x; a, b)$ and the cumulative distribution functions $P(x; a, b)$ and $Q(x; a, b)$.

(b) Input data
A = 5.0, B = 2.0 and XI = 0.3.

(c) Main program

```
      PROGRAM B1CDBT
! *** EXAMPLE OF D1CDBT ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER ISW
      REAL(8) A,B,XI,XO
!
      A=5.0D0
      B=2.0D0
      XI=0.3D0
      WRITE(6,1000)
      WRITE(6,2000) A,B
      WRITE(6,2010) XI
      WRITE(6,3000)
      ISW=0
      CALL D1CDBT(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) XO
!
      ISW=1
      CALL D1CDBT(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDBT(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDBT ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'A = ',F4.1,'  B = ',F4.1)
 2010 FORMAT(9X,'XI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'VALUE OF P.D.F = ',D17.10)
 5010 FORMAT(9X,'VALUE OF C.D.F(1) = ',D17.10)
 5020 FORMAT(9X,'VALUE OF C.D.F(2) = ',D17.10)
      END
```

(d) Output results

```
     *** D1CDBT ***
      ** INPUT **
         A =  5.0  B =  2.0
         XI =  0.3


      ** OUTPUT **
         IERR =    0
         VALUE OF P.D.F =  0.1701000000D+00
         IERR =    0
         VALUE OF C.D.F(1) =  0.1093500000D-01
         IERR =    0
         VALUE OF C.D.F(2) =  0.9890650000D+00
```

### 3.2.17 D1CDIB, R1CDIB
### Inverse Beta Distribution

(1) **Function**

Given the (cumulative distribution function; c.d.f.) $P(x;\alpha,\beta)$ or $Q(x;\alpha,\beta)$ of a beta distribution having positive parameters $a$ and $b$, the D1CDIB or R1CDIB obtains the value $x$ of the random variable at that time. $P(x;\alpha,\beta)$ and $Q(x;\alpha,\beta)$ are defined by the following equations.

$$P(x;a,b) = \begin{cases} 0 & (x \le 0; a,b > 0) \\ \dfrac{1}{B(a,b)} \displaystyle\int_0^x t^{a-1}(1-t)^{b-1}dt & (0 < x < 1; a,b > 0) \\ 1 & (x \ge 1; a,b > 0) \end{cases}$$

$$Q(x;a,b) = 1 - P(x;a,b)$$

$$= \begin{cases} 1 & (x \le 0; a,b > 0) \\ \dfrac{1}{B(a,b)} \displaystyle\int_x^\infty t^{a-1}(1-t)^{b-1}dt & (0 < x < 1; a,b > 0) \\ 0 & (x \ge 1; a,b > 0) \end{cases}$$

(2) **Usage**

Double precision:

CALL D1CDIB (A, B, XI, XO, ISW, IERR)

Single precision:

CALL R1CDIB (A, B, XI, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\\text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 1 | Input | Value of shape parameter $a$. |
| 2 | B | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 1 | Input | Value of shape parameter $b$. |
| 3 | XI | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 1 | Input | Value of the cumulative distribution function $P(x;a,b)$ or $Q(x;a,b)$ of the beta distribution. |
| 4 | XO | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 1 | Output | Value of random variable $x$. |
| 5 | ISW | I | 1 | Input | Processing switch<br>ISW=1:Input the value of the cumulative distribution function $P(x;a,b)$ for XI<br>ISW=2:Input the value of the cumulative distribution function $Q(x;a,b)$ for XI |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $\text{ISW} \in \{1,2\}$

(b) $\text{A}, \text{B} > 0.0$

(c) $0.0 \leq \text{XI} \leq 1.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | $\text{XI} = 0.0$ or $\text{XI} = 1.0$ | 0.0 or 1.0 is set for XO. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3600 | The maximum number of iterations was reached before the specified precision was obtained. | The value at that time is returned. |
| 4000 | An error occurred in the subroutine 3.2.16 $\left\{\begin{array}{c}\text{D1CDBT}\\\text{R1CDBT}\end{array}\right\}$. | Processing is aborted. |

(6) **Notes**

(a) The mean and variance of a beta distribution having parameters $a$ and $b$ are given by the following equations.

$$E[x] = \frac{a}{a+b}, \quad \sigma^2[x] = \frac{ab}{(a+b)^2(a+b+1)}$$

(b) A beta distribution with $a = b = 1$ is a uniform distribution on the interval (0, 1).

(7) **Example**

(a) Problem

For $a = 5.0$ and $b = 2.0$, obtain the values of $x$ for which the cumulative distribution function $P(x; \alpha, \beta) = 0.7$, or $Q(x; \alpha, \beta) = 0.7$, respectively.

(b) Input data

A = 5.0, B = 2.0 and XI = 0.7.

(c) Main program

```
      PROGRAM B1CDIB
! *** EXAMPLE OF D1CDIB ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER ISW
      REAL(8) A,B,XI,XO
!
      A=5.0D0
      B=2.0D0
      XI=0.7D0
      WRITE(6,1000)
      WRITE(6,2000) A,B
      WRITE(6,2010) XI
      WRITE(6,3000)
!
      ISW=1
      CALL D1CDIB(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDIB(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDIB ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'A = ',F4.1,'  B = ',F4.1)
 2010 FORMAT(9X,'XI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5010 FORMAT(9X,&
      'VALUE OF X CORRESPONDING TO P(X;A,B) = XI:',D17.10)
 5020 FORMAT(9X,&
      'VALUE OF X CORRESPONDING TO Q(X;A,B) = XI:',D17.10)
      END
```

(d) Output results

```
    *** D1CDIB ***
     ** INPUT **
        A =  5.0  B =  2.0
        XI =  0.7


     ** OUTPUT **
        IERR =    0
        VALUE OF X CORRESPONDING TO P(X;A,B) = XI: 0.8181965287D+00
        IERR =    0
        VALUE OF X CORRESPONDING TO Q(X;A,B) = XI: 0.6396423096D+00
```

### 3.2.18 D1CDUF, R1CDUF<br>Uniform Distribution

(1) **Function**

For a uniform distribution within the interval $(a, b)$, the D1CDUF or R1CDUF obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(x; a, b) = \begin{cases} \dfrac{1}{b-a} & (a \leq x \leq b) \\ 0 & (x < a, x > b) \end{cases}$$

(b) cumulative distribution function; c.d.f.

$$F(x; a, b) = \begin{cases} 0 & (x < a) \\ \dfrac{x-a}{b-a} & (a \leq x \leq b) \\ 1 & (x > b) \end{cases}$$

(2) **Usage**

Double precision:

CALL D1CDUF (XL, XU, XI, XO, ISW, IERR)

Single precision:

CALL R1CDUF (XL, XU, XI, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: {INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer}

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | XL | {D / R} | 1 | Input | Lower bound $a$ of the interval for the random variable $x$ |
| 2 | XU | {D / R} | 1 | Input | Upper bound $b$ of the interval for the random variable $x$ |
| 3 | XI | {D / R} | 1 | Input | Value of random variable $x$ |
| 4 | XO | {D / R} | 1 | Output | Value of the probability density function $f(x;a,b)$ or the cumulative distribution function $F(x;a,b)$ of the uniform distribution. |
| 5 | ISW | I | 1 | Input | ISW=0: Obtain the value of the probability density function $f(x;a,b)$ for XO<br>ISW=1: Obtain the value of the cumulative distribution function $F(x;a,b)$ for XO |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) XL≤XU

(b) ISW∈{0, 1}

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3100 | Restriction (b) was not satisfied. | |

(6) **Notes**
None

(7) **Example**

(a) Problem
Let $a$=0.0, $b$=1.0 and $x$=0.5 and obtain the values of the probability density function $f(x;a,b)$ and the cumulative distribution function $F(x;a,b)$.

(b) Input data
XL=0.0, XU=1.0 and XI=0.5.

(c) Main program

```
      PROGRAM B1CDUF
! *** EXAMPLE OF D1CDUF ***
      IMPLICIT REAL(8) (A-H,O-Z)
```

```
      INTEGER ISW,IERR
      REAL(8) XL,XU,XI,XO
!
      XL=0.0D0
      XU=1.0D0
      XI=0.5D0
!
      WRITE(6,1000)
      WRITE(6,2000) XL,XU,XI
      WRITE(6,3000)
!
      ISW=0
      CALL D1CDUF(XL,XU,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=1
      CALL D1CDUF(XL,XU,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDUF ***',/,&
             6X,'** INPUT **')
 2000 FORMAT(9X,'XL = ',F4.1,/,&
             9X,'XU = ',F4.1,/,&
             9X,'XI = ',F4.1)
 3000 FORMAT(' ',/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5010 FORMAT(9X,'VALUE OF P.D.F. =',D17.10)
 5020 FORMAT(9X,'VALUE OF C.D.F. =',D17.10)
      END
```

(d) Output results

```
     *** D1CDUF ***
      ** INPUT **
         XL =  0.0
         XU =  1.0
         XI =  0.5

      ** OUTPUT **
         IERR =    0
         VALUE OF P.D.F. = 0.1000000000D+01
         IERR =    0
         VALUE OF C.D.F. = 0.5000000000D+00
```

### 3.2.19 D1CDTR, R1CDTR Triangular Distribution

(1) **Function**

For a triangular distribution, the D1CDTR or R1CDTR obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(x;a,b,c) = \begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)} & (a \le x \le b) \\ \dfrac{2(c-x)}{(c-a)(c-b)} & (b < x \le c) \\ 0 & (x < a, x > c) \end{cases}$$

(b) cumulative distribution function; c.d.f.

$$F(x;a,b,c) = \begin{cases} 0 & (x < a) \\ \dfrac{(x-a)^2}{(b-a)(c-a)} & (a \le x \le b) \\ 1 - \dfrac{(c-x)^2}{(c-a)(c-b)} & (b < x \le c) \\ 1 & (x > c) \end{cases}$$

(2) **Usage**

Double precision:

CALL D1CDTR (A, B, C, XI, XO, ISW, IERR)

Single precision:

CALL R1CDTR (A, B, C, XI, XO, ISW, IERR)

## (3) Arguments

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | $x$ coordinate of the left end of the triangular distribution. (See Notes (a)) |
| 2 | B | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | $x$ coordinate of the apex of the triangular distribution. (See Notes (a)) |
| 3 | C | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | $x$ coordinate of the right end of the triangular distribution. (See Notes (a)) |
| 4 | XI | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Value of the random variable $x$ |
| 5 | XO | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Output | Value of the probability density function $f(x;a,b,c)$ or the cumulative distribution function $F(x;a,b,c)$ of the triangular distribution. |
| 6 | ISW | I | 1 | Input | ISW=0: Obtain the value of the probability density function $f(x;a,b,c)$ for XO<br>ISW=1: Obtain the value of the cumulative distribution function $F(x;a,b,c)$ for XO |
| 7 | IERR | I | 1 | Output | Error indicator |

## (4) Restrictions

(a) A≤B≤C

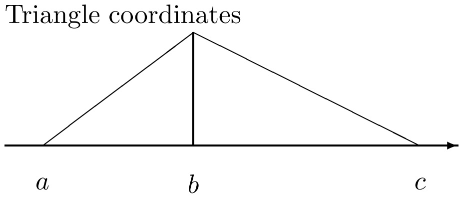
(b) ISW∈{0, 1}

## (5) Error indicator

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3100 | Restriction (b) was not satisfied. | |

## (6) Notes

(a) Triangle coordinates



$a$: $x$ coordinate of the left end of the triangular distribution
$b$: $x$ coordinate of the apex of the triangular distribution
$c$: $x$ coordinate of the right end of the triangular distribution

## (7) Example

(a) Problem
Let $a$=0.0, $b$=1.0, $c$=2.0 and $x$=0.5 and obtain the values of the probability density function $f(x; a, b, c)$ and the cumulative distribution function $F(x; a, b, c)$.

(b) Input data
A=0.0, B=1.0, C=2.0 and XI=0.5.

(c) Main program

```
      PROGRAM B1CDTR
! *** EXAMPLE OF D1CDTR ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER ISW,IERR
      REAL(8) A,B,C,XI,XO
!
      A=0.0D0
      B=1.0D0
      C=2.0D0
      XI=0.5D0
!
      WRITE(6,1000)
      WRITE(6,2000) A,B,C,XI
      WRITE(6,3000)
!
      ISW=0
      CALL D1CDTR(A,B,C,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=1
      CALL D1CDTR(A,B,C,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDTR ***',/,&
             6X,'** INPUT **')
 2000 FORMAT(9X,'A  = ',F4.1,/,&
             9X,'B  = ',F4.1,/,&
             9X,'C  = ',F4.1,/,&
             9X,'XI = ',F4.1)
 3000 FORMAT(' ',/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5010 FORMAT(9X,'VALUE OF P.D.F. =',D17.10)
 5020 FORMAT(9X,'VALUE OF C.D.F. =',D17.10)
      END
```

(d) Output results

```
     *** D1CDTR ***
      ** INPUT **
         A  =  0.0
         B  =  1.0
         C  =  2.0
         XI =  0.5

      ** OUTPUT **
         IERR =    0
         VALUE OF P.D.F. = 0.5000000000D+00
         IERR =    0
         VALUE OF C.D.F. = 0.1250000000D+00
```

### 3.2.20 D1CDPA, R1CDPA Pareto Distribution

(1) **Function**

For a Pareto distribution having parameters $a$ and $b$ ($a > 1$, $b > 0$), the D1CDPA or R1CDPA obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(x;a,b) = \begin{cases} (a-1)(\frac{x}{b})^{-a}\frac{1}{b} & (x > b; a > 1, b > 0) \\ 0 & (x \le b; a > 1, b > 0) \end{cases}$$

(b) cumulative distribution function; c.d.f.

$$P(x;a,b) = \int_b^x f(t;a,b)\ dt$$
$$= \begin{cases} 1-(\frac{x}{b})^{1-a} & (x > b; a > 1, b > 0) \\ 0 & (x \le b; a > 1, b > 0) \end{cases}$$

(c) c.d.f.

$$Q(x;a,b) = 1 - P(x;a,b)$$
$$= \begin{cases} (\frac{x}{b})^{1-a} & (x > b; a > 1, b > 0) \\ 1 & (x \le b; a > 1, b > 0) \end{cases}$$

(2) **Usage**

Double precision:
CALL D1CDPA (A, B, XI, XO, ISW, IERR)
Single precision:
CALL R1CDPA (A, B, XI, XO, ISW, IERR)

### (3) Arguments

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{matrix}D\\R\end{matrix}\right\}$ | 1 | Input | Value of parameter $a$. |
| 2 | B | $\left\{\begin{matrix}D\\R\end{matrix}\right\}$ | 1 | Input | Value of parameter $b$. |
| 3 | XI | $\left\{\begin{matrix}D\\R\end{matrix}\right\}$ | 1 | Input | ISW=0:Value of random variable $x$<br>ISW=1 or 2:Integration range of random variable $x$ |
| 4 | XO | $\left\{\begin{matrix}D\\R\end{matrix}\right\}$ | 1 | Output | Value of the probability density function $f(x;a,b)$<br>or of the cumulative distribution function $P(x;a,b)$<br>or $Q(x;a,b)$ of the Pareto distribution. |
| 5 | ISW | I | 1 | Input | Processing switch<br>ISW=0:Obtain the value of the probability density function $f(x;a,b)$ for XO<br>ISW=1: Obtain the value of the cumulative distribution function $P(x;a,b)$ for XO<br>ISW=2:Obtain the value of the cumulative distribution function $Q(x;a,b)$ |
| 6 | IERR | I | 1 | Output | Error indicator |

### (4) Restrictions

(a) $\mathrm{ISW} \in \{0, 1, 2\}$

(b) $\mathrm{A} > 1.0, \mathrm{B} > 0.0$

### (5) Error indicator

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | $\mathrm{XI} \leq \mathrm{B}$. | When ISW = 0, XO = 0.0 is set.<br>When ISW = 1, XO = 0.0 is set.<br>When ISW = 2, XO = 1.0 is set. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |

### (6) Notes

None

(7) **Example**

(a) Problem

Let $a = 5.0, b = 2.0$, and $x = 3.0$ and obtain the values of the probability density function $f(x; a, b)$ and the cumulative distribution functions $P(x; a, b)$ and $Q(x; a, b)$.

(b) Input data

A = 5.0, B = 2.0 and X = 3.0.

(c) Main program

```
      PROGRAM B1CDPA
! *** EXAMPLE OF D1CDPA ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER ISW
      REAL(8) A,B,XI,XO
!
      A=5.0D0
      B=2.0D0
      XI=3.0D0
      WRITE(6,1000)
      WRITE(6,2000) A,B
      WRITE(6,2010) XI
      WRITE(6,3000)
      ISW=0
      CALL D1CDPA(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) XO
!
      ISW=1
      CALL D1CDPA(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDPA(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDPA ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'A = ',F4.1,'  B = ',F4.1)
 2010 FORMAT(9X,'XI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'VALUE OF P.D.F = ',D17.10)
 5010 FORMAT(9X,'VALUE OF C.D.F(1) = ',D17.10)
 5020 FORMAT(9X,'VALUE OF C.D.F(2) = ',D17.10)
      END
```

(d) Output results

```
     *** D1CDPA ***
      ** INPUT **
         A =  5.0  B =  2.0
         XI =  3.0

      ** OUTPUT **
         IERR =    0
         VALUE OF P.D.F =  0.2633744856D+00
         IERR =    0
         VALUE OF C.D.F(1) =  0.8024691358D+00
         IERR =    0
         VALUE OF C.D.F(2) =  0.1975308642D+00
```

### 3.2.21 D1CDWE, R1CDWE<br>Weibull Distribution

(1) **Function**

For a Weibull distribution having parameters $a$ and $b$ ($a > 0$, $b > 0$), the D1CDWE or R1CDWE obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(x;a,b) = \begin{cases} a\left(\dfrac{x}{b}\right)^{a-1} e^{-\left(\frac{x}{b}\right)^a}\dfrac{1}{b} & (0 < x; a, b > 0) \\ 0 & (x \leq 0; a, b > 0) \end{cases}$$

(b) cumulative distribution function; c.d.f.

$$\begin{aligned} P(x;a,b) &= \int_0^x f(t;a,b)\ dt \\ &= \begin{cases} 1 - e^{-\left(\frac{x}{b}\right)^a} & (0 < x; a, b > 0) \\ 0 & (x \leq 0; a, b > 0) \end{cases} \end{aligned}$$

(c) c.d.f.

$$\begin{aligned} Q(x;a,b) &= 1 - P(x;a,b) \\ &= \begin{cases} e^{-\left(\frac{x}{b}\right)^a} & (0 < x; a, b > 0) \\ 1 & (x \leq 0; a, b > 0) \end{cases} \end{aligned}$$

(2) **Usage**

Double precision:

CALL D1CDWE (A, B, XI, XO, ISW, IERR)

Single precision:

CALL R1CDWE (A, B, XI, XO, ISW, IERR)

## (3) Arguments

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: $\begin{Bmatrix} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{Bmatrix}$

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | 1 | Input | Value of shape parameter $a$. |
| 2 | B | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | 1 | Input | Value of scale parameter $b$. |
| 3 | XI | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | 1 | Input | ISW=0: Value of random variable $x$<br>ISW=1 or 2: Integration range of random variable $x$ |
| 4 | XO | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | 1 | Output | Value of the probability density function $f(x;a,b)$ or of the cumulative distribution function $P(x;a,b)$ or $Q(x;a,b)$ of the Weibull distribution |
| 5 | ISW | I | 1 | Input | Processing switch<br>ISW=0:Obtain the value of the probability density function $f(x;a,b)$ for XO<br>ISW=1:Obtain the value of the cumulative distribution function $P(x;a,b)$ for XO<br>ISW=2:Obtain the value of the cumulative distribution function $Q(x;a,b)$ for XO |
| 6 | IERR | I | 1 | Output | Error indicator |

## (4) Restrictions

(a) ISW $\in \{0, 1, 2\}$

(b) $\mathrm{A} > 0.0, \mathrm{B} > 0.0$

## (5) Error indicator

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | XI $\leq 0.0$ | When ISW = 0, XO = 0.0 is set.<br>When ISW = 1, XO = 0.0 is set.<br>When ISW = 2, XO = 1.0 is set. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |

## (6) Notes

None

(7) **Example**

(a) Problem

Let $a = 5.0, b = 2.0$, and $x = 2.0$ and obtain the values of the probability density function $f(x; a, b)$ and the cumulative distribution functions $P(x; a, b)$ and $Q(x; a, b)$.

(b) Input data

A = 5.0, B = 2.0 and XI = 2.0.

(c) Main program

```
      PROGRAM B1CDWE
! *** EXAMPLE OF D1CDWE ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER ISW
      REAL(8) A,B,XI,XO
!
      A=5.0D0
      B=2.0D0
      XI=2.0D0
      WRITE(6,1000)
      WRITE(6,2000) A,B
      WRITE(6,2010) XI
      WRITE(6,3000)
      ISW=0
      CALL D1CDWE(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) XO
!
      ISW=1
      CALL D1CDWE(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDWE(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDWE ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'A = ',F4.1,'  B = ',F4.1)
 2010 FORMAT(9X,'XI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'VALUE OF P.D.F = ',D17.10)
 5010 FORMAT(9X,'VALUE OF C.D.F(1) = ',D17.10)
 5020 FORMAT(9X,'VALUE OF C.D.F(2) = ',D17.10)
      END
```

(d) Output results

```
     *** D1CDWE ***
      ** INPUT **
         A =  5.0  B =  2.0
         XI =  2.0


      ** OUTPUT **
         IERR =    0
         VALUE OF P.D.F =  0.9196986029D+00
         IERR =    0
         VALUE OF C.D.F(1) =  0.6321205588D+00
         IERR =    0
         VALUE OF C.D.F(2) =  0.3678794412D+00
```

### 3.2.22 D1CDEX, R1CDEX Exponential Distribution

(1) **Function**

For a exponential distribution having parameters $\lambda$ , the D1CDEX or R1CDEX obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & (x > 0; \lambda > 0) \\ 0 & (x \leq 0; \lambda > 0) \end{cases}$$

(b) cumulative distribution function; c.d.f.

$$P(x;\lambda) = \int_0^x \lambda e^{-\lambda t} dt \quad (\lambda > 0)$$

(c) c.d.f.

$$Q(x;\lambda) = 1 - P(x;\lambda) = \int_x^\infty \lambda e^{-\lambda t} dt \quad (\lambda > 0)$$

(2) **Usage**

Double precision:
CALL D1CDEX (B, XI, XO, ISW, IERR)
Single precision:
CALL R1CDEX (B, XI, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\\text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | B | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Value of scale parameter $\lambda$ |
| 2 | XI | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Value of random variable $x$ |
| 3 | XO | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Output | Value of the probability density function $f(x;\lambda)$ of the gamma distribution or of the cumulative distribution function $P(x;\lambda)$ or $Q(x;\lambda)$ of the exponential distribution. |
| 4 | ISW | I | 1 | Input | ISW=0: Obtain the value of the probability density function $f(x;\lambda)$ for XO<br>ISW=1: Obtain the value of the cumulative distribution function $P(x;\lambda)$ for XO<br>ISW=2: Obtain the value of the cumulative distribution function $Q(x;\lambda)$ for XO |
| 5 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{0, 1, 2\}$

(b) B $> 0.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | XI $\leq 0.0$ | 0.0 or 1.0 is set for XO. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |

(6) **Notes**

(a) The mean and variance of a gamma distribution having parameters $\lambda$ are given by the following equations.

$$E[x] = \frac{1}{\lambda}, \sigma^2[x] = \frac{1}{\lambda^2}$$

(b) A exponential distribution is gamma distribution for which $\alpha = 1$.

(7) **Example**

(a) Problem
Let $\lambda = 2.0$ and $x = 1.0$ and obtain the values of the probability density function $f(x;\lambda)$ and the cumulative distribution functions $P(x;\lambda)$ and $Q(x;\lambda)$.

(b) Input data
B $= 2.0$ and XI $= 1.0$.

(c) Main program

```
      PROGRAM B1CDEX
! *** EXAMPLE OF B1CDEX ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER ISW,IERR
      REAL(8) XI,XO
      REAL(8) B
!
      B=2.0D0
      XI=1.0D0
      WRITE(6,1000)
      WRITE(6,2000) B
      WRITE(6,2010) XI
      WRITE(6,3000)
      ISW=0
      CALL D1CDEX(B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) XO
!
      ISW=1
      CALL D1CDEX(B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDEX(B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(/,5X,'*** D1CDEX ***',/,/,&
             6X,'** INPUT **',/)
 2000 FORMAT(8X,'B = ',F4.1,/)
 2010 FORMAT(8X,'XI = ',F4.1,/)
 3000 FORMAT(6X,'** OUTPUT **',/)
 4000 FORMAT(8X,'IERR = ',I4,/)
 5000 FORMAT(8X,'VALUE OF P.D.F = ',D17.10,/)
 5010 FORMAT(8X,'VALUE OF C.D.F(1) = ',D17.10,/)
 5020 FORMAT(8X,'VALUE OF C.D.F(2) = ',D17.10,/)
      END
```

(d) Output results

```
*** D1CDEX ***
 ** INPUT **
   B =  2.0
   XI =  1.0
 ** OUTPUT **
   IERR =    0
   VALUE OF P.D.F =  0.2706705665D+00
   IERR =    0
   VALUE OF C.D.F(1) =  0.8646647168D+00
   IERR =    0
   VALUE OF C.D.F(2) =  0.1353352832D+00
```

### 3.2.23 D1CDGU, R1CDGU Gumbel Distribution

(1) **Function**

For a Gumbel distribution having the parameters $a$ and $b$ ($b > 0$), the D1CDGU or R1CDGU obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(x; a, b) = \frac{1}{b} e^{\frac{x-a}{b}} e^{-e^{\frac{x-a}{b}}}$$

(b) cumulative distribution function; c.d.f.

$$P(x; a, b) = \int_{-\infty}^{x} f(t; a, b) dt$$

(c) c.d.f.

$$Q(x; a, b) = 1 - P(x; a, b) = \int_{x}^{\infty} f(t; a, b) dt$$

(2) **Usage**

Double precision:

CALL D1CDGU (A, B, XI, XO, ISW, IERR)

Single precision:

CALL R1CDGU (A, B, XI, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{ \begin{array}{l} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{array} \right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{ \begin{array}{c} \text{D} \\ \text{R} \end{array} \right\}$ | 1 | Input | Value of the location parameter $a$. |
| 2 | B | $\left\{ \begin{array}{c} \text{D} \\ \text{R} \end{array} \right\}$ | 1 | Input | Value of the scale parameter $b$. |
| 3 | XI | $\left\{ \begin{array}{c} \text{D} \\ \text{R} \end{array} \right\}$ | 1 | Input | ISW=0:Value of random variable $x$<br>ISW=1 or 2:Integration range of random variable $x$ |
| 4 | XO | $\left\{ \begin{array}{c} \text{D} \\ \text{R} \end{array} \right\}$ | 1 | Output | Value of the probability density function $f(x;a,b)$ or of the cumulative distribution function $P(x;a,b)$ or $Q(x;a,b)$ of the Gumbel distribution |
| 5 | ISW | I | 1 | Input | Processing switch<br>ISW=0 : Obtain the value of the probability density function $f(x;a,b)$ for XO<br>ISW=1:Obtain the value of the cumulative distribution function $P(x;a,b)$ for XO<br>ISW=2:Obtain the value of the cumulative distribution function $Q(x;a,b)$ for XO |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) B>0.0

(b) ISW $\in \{0, 1, 2\}$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |

(6) **Notes**

None

(7) **Example**

(a) Problem

Let $a$=1.0, $b$=2.0 and $x$=1.5 and obtain the values of the probability density function $f(x;a,b)$ and the cumulative distribution function $P(x;a,b)$ and $Q(x;a,b)$.

(b) Input data

XL=1.0, XU=2.0 and XI=1.5.

(c) Main program

```
      PROGRAM B1CDGU
! *** EXAMPLE OF D1CDGU ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER ISW
      REAL(8) A,B,XI,XO
!
      A=1.0D0
      B=2.0D0
      XI=1.5D0
      WRITE(6,1000)
      WRITE(6,2000) A,B
      WRITE(6,2010) XI
      WRITE(6,3000)
      ISW=0
      CALL D1CDGU(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) XO
!
      ISW=1
      CALL D1CDGU(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDGU(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(1X,'*** D1CDGU ***',/,&
      /,1X,'** INPUT **',/)
 2000 FORMAT(1X,'A = ',F4.1,'  B = ',F4.1)
 2010 FORMAT(1X,'XI = ',F4.1)
 3000 FORMAT(1X,/,1X,'** OUTPUT **')
 4000 FORMAT(1X,/,1X,'IERR = ',I4)
 5000 FORMAT(1X,'VALUE OF P.D.F = ',D17.10)
 5010 FORMAT(1X,'VALUE OF C.D.F(1) = ',D17.10)
 5020 FORMAT(1X,'VALUE OF C.D.F(2) = ',D17.10)
      END
```

(d) Output results

```
*** D1CDGU ***

** INPUT **

A =  1.0  B =  2.0
XI =  1.5

** OUTPUT **

IERR =    0
VALUE OF P.D.F =  0.1777863737D+00

IERR =    0
VALUE OF C.D.F(1) =  0.7230796659D+00

IERR =    0
VALUE OF C.D.F(2) =  0.2769203341D+00
```

### 3.2.24 D1CDLD, R1CDLD Logarithmic Distribution

(1) **Function**

For a logarithmic distribution within the interval $(a, b)$, the D1CDLD or R1CDLD obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(x; a, b) = \frac{\log x}{b(\log b - 1) - a(\log a - 1)}$$

(b) cumulative distribution function; c.d.f.

$$P(x; a, b) = \int_a^x f(t; a, b)dt$$

(c) c.d.f.

$$Q(x; a, b) = 1 - P(x; a, b) = \int_x^b f(t; a, b)dt$$

(2) **Usage**

Double precision:
CALL D1CDLD (XL, XU, XI, XO, ISW, IERR)
Single precision:
CALL R1CDLD (XL, XU, XI, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{ \begin{array}{l} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{array} \right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | XL | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Lower bound $a$ of the interval for the variable $x$ |
| 2 | XU | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Upper bound $b$ of the interval for the variable $x$ |
| 3 | XI | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Value of the variable $x$ |
| 4 | XO | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Output | Value of the probability density function $f(x)$ or the cumulative distribution function $F(x)$ of the logarithmic distribution. |
| 5 | ISW | I | 1 | Input | ISW=0: Obtain the value of the probability density function $f(x)$ for XO<br>ISW=1: Obtain the value of the cumulative distribution function $P(x)$ for XO ISW=2: Obtain the value of the cumulative distribution function $Q(x)$ for XO |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) 0.0<XL<XU

(b) XL<XI<XU

(c) ISW $\in \{0, 1, 2\}$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |

(6) **Notes**
None

(7) **Example**

(a) Problem
Let $a$=0.0, $b$=2.0 and $x$=1.5 and obtain the values of the probability density function $f(x; a, b)$ and the cumulative distribution function $P(x; a, b)$ and $Q(x; a, b)$.

(b) Input data

XL=1.0, XU=2.0 and XI=1.5.

(c) Main program

```
      PROGRAM B1CDLD
! *** EXAMPLE OF D1CDLD ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER ISW,IERR
      REAL(8) XL,XU,XI,XO
!
      XL=1.0D0
      XU=2.0D0
      XI=1.5D0
!
      WRITE(6,1000)
      WRITE(6,2000) XL,XU,XI
      WRITE(6,3000)
!
      ISW=0
      CALL D1CDLD(XL,XU,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=1
      CALL D1CDLD(XL,XU,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      ISW=2
      CALL D1CDLD(XL,XU,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(1X,'*** D1CDLD ***',/,&
             /,1X,'** INPUT **',/)
 2000 FORMAT(1X,'XL = ',F4.1,/,&
             1X,'XU = ',F4.1,/,&
             1X,'XI = ',F4.1)
 3000 FORMAT(1X,/,1X,'** OUTPUT **')
 4000 FORMAT(1X,/,1X,'IERR = ',I4)
 5010 FORMAT(1X,'VALUE OF P.D.F. =',D17.10)
 5020 FORMAT(1X,'VALUE OF C.D.F. =',D17.10)
      END
```

(d) Output results

```
*** D1CDLD ***

** INPUT **

XL =  1.0
XU =  2.0
XI =  1.5

** OUTPUT **

IERR =    0
VALUE OF P.D.F. = 0.1049627302D+01

IERR =    0
VALUE OF C.D.F. = 0.2800912285D+00

IERR =    0
VALUE OF C.D.F. = 0.7199087715D+00
```

### 3.2.25 D1CDLN, R1CDLN Log-Normal Distribution

(1) **Function**

For a log-normal distribution having mean $e^{\mu}\sqrt{e^{\sigma^2}}$ and variance $e^{2\mu}e^{\sigma^2}(e^{\sigma^2}-1)$, the D1CDLN or R1CDLN obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \quad (\sigma > 0)$$

(b) cumulative distribution function; c.d.f.

$$P(x;\mu,\sigma) = \int_0^x \frac{1}{t\sigma\sqrt{2\pi}}e^{-\frac{(\ln t-\mu)^2}{2\sigma^2}}dt \quad (\sigma > 0)$$

(c) c.d.f.

$$Q(x;\mu,\sigma) = 1 - P(x;\mu,\sigma) = \int_x^\infty \frac{1}{t\sigma\sqrt{2\pi}}e^{-\frac{(\ln t-\mu)^2}{2\sigma^2}}dt \quad (\sigma > 0)$$

(2) **Usage**

Double precision:

CALL D1CDLN (XE, XV, XI, XO, ISW, IERR)

Single precision:

CALL R1CDLN (XE, XV, XI, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | XE | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Value of parameter $\mu$ |
| 2 | XV | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Value of parameter $\sigma^2$ |
| 3 | XI | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Value of random variable $x$ |
| 4 | XO | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Output | Value of the probability density function $f(x)$ of the log-normal distribution or of the cumulative distribution function $P(x;\mu,\sigma)$ or $Q(x;\mu,\sigma)$ of the normal distribution. |
| 5 | ISW | I | 1 | Input | ISW=0: Obtain the value of the probability density function $f(x;\mu,\sigma)$ for XO<br>ISW=1: Obtain the value of the cumulative distribution function $P(x;\mu,\sigma)$ for XO<br>ISW=2:Obtain the value of the cumulative distribution function $Q(x;\mu,\sigma)$ for XO |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{0,1,2\}$

(b) XV $> 0.0$

(c) X $> 0.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |

(6) **Notes**

(a) From the relational expression $P(x;\mu,\sigma)+Q(x;\mu,\sigma)=1$, it is possible to obtain either $P(x;\mu,\sigma)$ or $Q(x;\mu,\sigma)$ from the other. However, there may be times when cancellation of significant digits occurs, preventing good precision from being obtained.

(b) When the random variable $x$ obeys a log-normal distribution having mean $e^{\mu}\sqrt{e^{\sigma^2}}$ and variance $e^{2\mu}e^{\sigma^2}(e^{\sigma^2}-1)$, the random variable $\ln x$ obeys a normal distribution $N(\mu,\sigma^2)$.

(7) **Example**

(a) Problem

Let $\mu = 5.0$, $\sigma^2 = 2.5$ and $x = 3.0$ and obtain the values of the probability density function $f(x;\mu,\sigma)$ and the cumulative distribution functions $P(x;\mu,\sigma)$ and $Q(x;\mu,\sigma)$.

(b) Input data

XE = 5.0, XV = 2.5 and XI = 3.0.

(c) Main program

```
      PROGRAM B1CDLN
! *** EXAMPLE OF D1CDLN ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER ISW
      REAL(8) XE,XV,XI,XO
!
      XE=5.0D0
      XV=2.5D0
      XI=EXP(3.0D0)
      WRITE(6,1000)
      WRITE(6,2000) XE
      WRITE(6,2010) XV
      WRITE(6,2020) XI
      WRITE(6,3000)
      ISW=0
      CALL D1CDLN(XE,XV,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) XO
!
      ISW=1
      CALL D1CDLN(XE,XV,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDLN(XE,XV,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1CDLN ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'XE = ',F4.1)
 2010 FORMAT(9X,'XV = ',F4.1)
 2020 FORMAT(9X,'XI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'VALUE OF P.D.F = ',D17.10)
 5010 FORMAT(9X,'VALUE OF C.D.F(1) = ',D17.10)
 5020 FORMAT(9X,'VALUE OF C.D.F(2) = ',D17.10)
      END
```

(d) Output results

```
     *** D1CDLN ***
      ** INPUT **
         XE =  5.0
         XV =  2.5
         XI = 20.1


      ** OUTPUT **
         IERR =    0
         VALUE OF P.D.F =  0.5644442201D-02
         IERR =    0
         VALUE OF C.D.F(1) =  0.1029516054D+00
         IERR =    0
         VALUE OF C.D.F(2) =  0.8970483946D+00
```

### 3.2.26 D1CDLG, R1CDLG Logistic Distribution

(1) **Function**

For a logistic distribution having mean $\alpha$ and variance $\left(\sigma^2 = \dfrac{\pi^2\beta^2}{3}\right)$ the D1CDLG or R1CDLG obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(x;\alpha,\beta) = \frac{e^{-\frac{x-\alpha}{\beta}}}{\beta\left\{1+e^{-\frac{x-\alpha}{\beta}}\right\}^2} \quad (\beta > 0)$$

$$\sigma^2 = \frac{\pi^2\beta^2}{3}$$

(b) cumulative distribution function; c.d.f.

$$P(x;\alpha,\beta) = \frac{1}{1+e^{-\frac{x-\alpha}{\beta}}} \quad (\beta > 0)$$

(c) c.d.f.

$$Q(x;\alpha,\beta) = 1 - P(x;\alpha,\beta) = \frac{1}{1+e^{\frac{x-\alpha}{\beta}}} \quad (\beta > 0)$$

(2) **Usage**

Double precision:

CALL D1CDLG (XA, XB, XI, XO, ISW, IERR)

Single precision:

CALL R1CDLG (XA, XB, XI, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: { INTEGER(4) as for 32bit Integer ; INTEGER(8) as for 64bit Integer }
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | XA | {D, R} | 1 | Input | Value of mean $\alpha$ |
| 2 | XB | {D, R} | 1 | Input | Value of parameter $\beta$ |
| 3 | XI | {D, R} | 1 | Input | Value of random variable $x$ |
| 4 | XO | {D, R} | 1 | Output | Value of the probability density function $f(x)$ of the logistic distribution or of the cumulative distribution function $P(x;\alpha,\beta)$ or $Q(x;\alpha,\beta)$ of the logistic distribution. |
| 5 | ISW | I | 1 | Input | ISW=0: Obtain the value of the probability density function $f(x;\alpha,\beta)$ for XO ISW=1:Obtain the value of the cumulative distribution function $P(x;\alpha,\beta)$ ISW=2:Obtain the value of the cumulative distribution function $Q(x;\alpha,\beta)$ |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{0, 1, 2\}$

(b) XB $> 0.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |

(6) **Notes**

(a) From the relational expression $P(x;\alpha,\beta) + Q(x;\alpha,\beta) = 1$, it is possible to obtain either $P(x;\alpha,\beta)$ or $Q(x;\alpha,\beta)$ from the other. However, there may be times when cancellation of significant digits occurs, preventing good precision from being obtained.

(7) **Example**

(a) Problem
Let $\alpha = 1.0$, $\beta = 1.0$ and $x = 3.0$ and obtain the values of the probability density function $f(x;\alpha,\beta)$ and the cumulative distribution functions $P(x;\alpha,\beta)$ and $Q(x;\alpha,\beta)$.

(b) Input data
XA = 1.0, XB = 1.0 and XI = 3.0.

(c) Main program

```
      PROGRAM B1CDLG
! *** EXAMPLE OF D1CDLG ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER ISW,IERR
      REAL(8) XA,XB,XI,XO
!
      XA=1.0D0
      XB=1.0D0
      XI=3.0D0
      WRITE(6,6000)
      WRITE(6,6010) XA
      WRITE(6,6020) XB
      WRITE(6,6030) XI
      WRITE(6,6040)
!
      ISW=0
      CALL D1CDLG(XA,XB,XI,XO,ISW,IERR)
      WRITE(6,6050) IERR
      WRITE(6,6060) XO
!
      ISW=1
      CALL D1CDLG(XA,XB,XI,XO,ISW,IERR)
      WRITE(6,6050) IERR
      WRITE(6,6070) XO
!
      ISW=2
      CALL D1CDLG(XA,XB,XI,XO,ISW,IERR)
      WRITE(6,6050) IERR
      WRITE(6,6080) XO
      STOP
!
 6000 FORMAT(/,&
      1X,'    *** D1CDLG ***',/,/,&
      1X,'     ** INPUT **',/)
 6010 FORMAT(1X,'        XA = ',F4.1)
 6020 FORMAT(1X,'        XB = ',F4.1)
 6030 FORMAT(1X,'        XI = ',F4.1)
 6040 FORMAT(/,&
      1X,'     ** OUTPUT **',/)
 6050 FORMAT(1X,'        IERR = ',I4)
 6060 FORMAT(1X,'        VALUE OF P.D.F    = ',D17.10,/)
 6070 FORMAT(1X,'        VALUE OF C.D.F(1) = ',D17.10,/)
 6080 FORMAT(1X,'        VALUE OF C.D.F(2) = ',D17.10,/)
      END
```

(d) Output results

```
    *** D1CDLG ***

     ** INPUT **

        XA =  1.0
        XB =  1.0
        XI =  3.0

     ** OUTPUT **

        IERR =    0
        VALUE OF P.D.F    =  0.1049935854D+00

        IERR =    0
        VALUE OF C.D.F(1) =  0.8807970780D+00

        IERR =    0
        VALUE OF C.D.F(2) =  0.1192029220D+00
```

### 3.2.27 D1CDCC, R1CDCC Cauchy Distribution

(1) **Function**

For Cauchy distribution having parameters $\alpha$ and $\beta$, the D1CDCC or R1CDCC obtains the values of the following functions.

(a) probability density function; p.d.f.

$$f(x;\alpha,\beta) = \frac{1}{\pi}\left[\frac{\beta}{\beta^2+(x-\alpha)^2}\right] \quad (\beta > 0)$$

(b) cumulative distribution function; c.d.f.

$$\begin{aligned} P(x;\alpha,\beta) &= \int_{-\infty}^{x} f(t;\alpha,\beta)\ dt \\ &= \frac{1}{2}+\frac{1}{\pi}\tan^{-1}\frac{(x-\alpha)}{\beta} \quad (\beta > 0) \end{aligned}$$

(c) c.d.f.

$$\begin{aligned} Q(x;\alpha,\beta) &= 1 - P(x;\alpha,\beta) \\ &= \frac{1}{2}-\frac{1}{\pi}\tan^{-1}\frac{(x-\alpha)}{\beta} \quad (\beta > 0) \end{aligned}$$

(2) **Usage**

Double precision:

CALL D1CDCC (A, B, XI, XO, ISW, IERR)

Single precision:

CALL R1CDCC (A, B, XI, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{ \begin{array}{l} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{array} \right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{ \begin{array}{c} D \\ R \end{array} \right\}$ | 1 | Input | Value of parameter $\alpha$ |
| 2 | B | $\left\{ \begin{array}{c} D \\ R \end{array} \right\}$ | 1 | Input | Value of parameter $\beta$ |
| 3 | XI | $\left\{ \begin{array}{c} D \\ R \end{array} \right\}$ | 1 | Input | Value of random variable $x$ |
| 4 | XO | $\left\{ \begin{array}{c} D \\ R \end{array} \right\}$ | 1 | Output | Value of the probability density function $f(x;\alpha,\beta)$ of Cauchy distribution or of the cumulative distribution function $P(x;\alpha,\beta)$ or $Q(x;\alpha,\beta)$ of Cauchy distribution. |
| 5 | ISW | I | 1 | Input | ISW=0: Obtain the value of the probability density function $f(x;\alpha,\beta)$ for XO<br>ISW=1: Obtain the value of the cumulative distribution function $P(x;\alpha,\beta)$ for XO<br>ISW=2: Obtain the value of the cumulative distribution function $Q(x;\alpha,\beta)$ for XO |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{0, 1, 2\}$

(b) B $> 0.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |

(6) **Notes**

None

(7) **Example**

(a) Problem

Let $\alpha = 2.0$, $\beta = 1.0$ and $x = 3.0$ and obtain the values of the probability density function $f(x; \alpha, \beta)$ and the cumulative distribution functions $P(x; \alpha, \beta)$ and $Q(x; \alpha, \beta)$.

(b) Input data

A = 2.0, B = 1.0 and XI = 3.0.

(c) Main program

```
      PROGRAM B1CDCC
! *** EXAMPLE OF B1CDCC ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER ISW,IERR
      REAL(8) XI,XO
      REAL(8) A,B
!
      A=2.0D0
      B=1.0D0
      XI=3.0D0
      WRITE(6,1000)
      WRITE(6,2000)A
      WRITE(6,2010)B
      WRITE(6,2020)XI
      WRITE(6,3000)
      ISW=0
      CALL D1CDCC(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) XO
!
      ISW=1
      CALL D1CDCC(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      ISW=2
      CALL D1CDCC(A,B,XI,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) XO
!
      STOP
!
 1000 FORMAT(/,5X,'*** D1CDCC ***',/,/,&
             6X,'** INPUT **',/)
 2000 FORMAT(8X,'A = ',F4.1,/)
 2010 FORMAT(8X,'B = ',F4.1,/)
 2020 FORMAT(8X,'XI = ',F4.1,/)
 3000 FORMAT(6X,'** OUTPUT **',/)
 4000 FORMAT(8X,'IERR = ',I4,/)
 5000 FORMAT(8X,'VALUE OF P.D.F = ',D17.10,/)
 5010 FORMAT(8X,'VALUE OF C.D.F(1) = ',D17.10,/)
 5020 FORMAT(8X,'VALUE OF C.D.F(2) = ',D17.10,/)
      END
```

(d) Output results

```
     *** D1CDCC ***

      ** INPUT **

        A =  2.0

        B =  1.0

        XI =  3.0

      ** OUTPUT **

        IERR =    0

        VALUE OF P.D.F =  0.1591549431D+00

        IERR =    0

        VALUE OF C.D.F(1) =  0.7500000000D+00

        IERR =    0

        VALUE OF C.D.F(2) =  0.2500000000D+00
```

## 3.3 DISCRETE DISTRIBUTIONS

### 3.3.1 D1DDBP, R1DDBP
### Binomial Distribution and Negative Binomial Distribution

(1) **Function**

(a) Binomial Distribution (1)

Given the probability that an event will occur $p$, the number of trials $n$, and the number of occurrences $m$, the D1DDBP or R1DDBP obtains the values of the binomial distribution probability $P_{BIN}(X = m; p, n)$ and the cumulative distribution function (c.d.f.) $P_{BIN}(X \le m; p, n)$ in $m$ occurrences, which are defined by the following equations.

$$P_{BIN}(X = m; p, n) = \begin{pmatrix} n \\ m \end{pmatrix} p^m \cdot q^{n-m} \quad (q = 1 - p)$$

$$P_{BIN}(X \le m; p, n) = \sum_{i=m}^{n} \begin{pmatrix} n \\ i \end{pmatrix} p^i \cdot q^{n-i}$$

(b) Binomial Distribution (2)

Given the probability of success in one trial $p$, the number of trials of independent events $n$, and the maximum value for the number of failures $m$, the D1DDBP or R1DDBP obtains the value of the probability $Q_{BIN}(X = m; p, n)$ of at least $(n-m)$ successes in $n$ trials, which is defined by the following equation.

$$\begin{aligned} Q_{BIN}(X = m; p, n) &= P_{BIN}(X \ge n - m; p, n) \\ &= 1 - \sum_{r=0}^{n-m-1} \begin{pmatrix} n \\ r \end{pmatrix} p^r \cdot q^{n-r} \quad (q = 1 - p) \end{aligned}$$

(c) Negative Binomial Distribution

Given the probability of success in one trial $p$ and the number of successes $n$ and number of failures $m$ in repeated trials, the D1DDBP or R1DDBP obtains the values of the negative binomial distribution probability $P_{NB}(X = m; p, n)$ and cumulative distribution function (c.d.f.) $P_{NB}(X \le m; p, n)$ in $m$ failures, which are defined by the following equations.

$$P_{NB}(X = m; p, n) = \begin{pmatrix} n + m - 1 \\ m \end{pmatrix} p^n \cdot q^m \qquad (q = 1 - p)$$

$$P_{NB}(X \le m; p, n) = \sum_{i=0}^{m} \begin{pmatrix} n + i - 1 \\ i \end{pmatrix} p^n \cdot q^i$$

(2) **Usage**

Double precision:

CALL D1DDBP (N, M, PI, PO, ISW, IERR)

Single precision:

CALL R1DDBP (N, M, PI, PO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Variable $n$ |
| 2 | M | I | 1 | Input | Variable $m$ |
| 3 | PI | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Probability $p$ |
| 4 | PO | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Output | Value of binomial distribution probability $P_{BIN}(X = m; p, n)$, binomial distribution cumulative distribution function $P_{BIN}(X \leq m; p, n)$, negative binomial distribution probability $P_{NB}(X = m; p, n)$, negative binomial distribution cumulative distribution function $P_{NB}(X \leq m; p, n)$ or probability $Q_{BIN}(X = m; p, n)$ |
| 5 | ISW | I | 1 | Input | Processing switch<br>ISW=1: Obtain the value of the binomial distribution probability $P_{BIN}(X = m; p, n)$ for PO<br>ISW=2: Obtain the value of the binomial distribution cumulative distribution function $P_{BIN}(X \leq m; p, n)$ for PO<br>ISW=3: Obtain the value of the negative binomial distribution probability $P_{NB}(X = m; p, n)$ for PO<br>ISW=4: Obtain the value of the negative binomial distribution cumulative distribution function $P_{NB}(X \leq m; p, n)$ for PO<br>ISW=5: Obtain the value of the probability $Q_{BIN}(X = m; p, n)$ for PO |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) 0.0<PI<1.0

(b) N≥1

(c) 0≤M≤N (when ISW∈{1, 2, 5})
M≥0 (when ISW∈{3, 4})

(d) ISW∈{1, 2, 3, 4, 5}

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3100 | Restriction (b) was not satisfied. | |
| 3200 | Restriction (c) was not satisfied. | |
| 3300 | Restriction (d) was not satisfied. | |

(6) **Notes**

None

(7) **Example**

(a) Problem

- Let the probability that an event will occur $p$=0.3, the number of trials $n$=10 and the number of occurrences $m$=3 and obtain the values of the binomial distribution probability $P_{BIN}(X = m; p, n)$ and cumulative distribution function $P_{BIN}(X \leq m; p, n)$.
- Let the probability of success in one trial $p$=0.3, the number of successes in repeated trials $n$=10 and the number of failures in repeated trials $m$=3 and obtain the values of the negative binomial distribution probability $P_{NB}(X = m; p, n)$ and cumulative distribution function $P_{NB}(X \leq m; p, n)$.
- Let the probability of success in one trial $p$=0.3, the number of trials of independent events $n$=10 and the maximum value of the number of failures $m$=3 and obtain the probability $Q_{BIN}(X = m; p, n)$.

(b) Input data

PI=0.3, N=10 and M=3.

(c) Main program

```
      PROGRAM B1DDBP
! *** EXAMPLE OF D1DDBP ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR,N,M,ISW
      REAL(8) PI,PO
!
      N=10
      M=3
      PI=0.3D0
      WRITE(6,1000)
      WRITE(6,2000) N
      WRITE(6,2010) M
      WRITE(6,2020) PI
      WRITE(6,3000)
      ISW=1
      CALL D1DDBP(N,M,PI,PO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) PO
!
      ISW=2
      CALL D1DDBP(N,M,PI,PO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) PO
!
      ISW=3
      CALL D1DDBP(N,M,PI,PO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5020) PO
!
      ISW=4
      CALL D1DDBP(N,M,PI,PO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5030) PO
!
      ISW=5
      CALL D1DDBP(N,M,PI,PO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5040) PO
      STOP
```

```
!
 1000 FORMAT(' ',/,5X,'*** D1DDBP ***',/,&
             6X,'** INPUT **')
 2000 FORMAT(9X,'N  = ',I3)
 2010 FORMAT(9X,'M  = ',I3)
 2020 FORMAT(9X,'PI = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'VALUE OF P.D.F. OF BINOMIAL DISTRIBUTION= ',D17.10)
 5010 FORMAT(9X,'VALUE OF C.D.F. OF BINOMIAL DISTRIBUTION= ',D17.10)
 5020 FORMAT(9X,&
       'VALUE OF P.D.F. OF NEGATIVE BINOMIAL DISTRIBUTION= ',D17.10)
 5030 FORMAT(9X,&
       'VALUE OF C.D.F. OF NEGATIVE BINOMIAL DISTRIBUTION= ',D17.10)
 5040 FORMAT(9X,'VALUE OF BINOMIAL PROBABILITY= ',D17.10)
      END
```

(d) Output results

```
     *** D1DDBP ***
      ** INPUT **
         N  =  10
         M  =   3
         PI =  0.3


      ** OUTPUT **
         IERR =    0
         VALUE OF P.D.F. OF BINOMIAL DISTRIBUTION=  0.2668279320D+00
         IERR =    0
         VALUE OF C.D.F. OF BINOMIAL DISTRIBUTION=  0.6172172136D+00
         IERR =    0
         VALUE OF P.D.F. OF NEGATIVE BINOMIAL DISTRIBUTION=  0.4455837540D-03
         IERR =    0
         VALUE OF C.D.F. OF NEGATIVE BINOMIAL DISTRIBUTION=  0.6519600090D-03
         IERR =    0
         VALUE OF BINOMIAL PROBABILITY=  0.1059207840D-01
```

### 3.3.2 D1DDGO, R1DDGO Geometric Distribution

(1) **Function**

Given the probability of success in $m$ trial $p$, the D1DDBP or R1DDBP obtains the values of the geometric distribution probability $P_{NB}(X = m; p)$ and cumulative distribution function (c.d.f.) $P_{NB}(X \leq m; p)$ , which are defined by the following equations.

$$P_{NB}(X = m; p) = q^{m-1}p \qquad (q = 1 - p)$$

$$P_{NB}(X \leq m; p) = \sum_{i=0}^{m} q^{m-1}p$$

(2) **Usage**

Double precision:
CALL D1DDGO (M, PI, PO, ISW, IERR)
Single precision:
CALL R1DDGO (M, PI, PO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: {INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer}

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | M | I | 1 | Input | Variable $m$ |
| 2 | PI | {D / R} | 1 | Input | Probability $p$ |
| 3 | PO | {D / R} | 1 | Output | Value of geometric distribution probability $P_{NB}(X = m; p)$, geometric distribution cumulative distribution function $P_{NB}(X \leq m; p)$ |
| 4 | ISW | I | 1 | Input | Processing switch<br>ISW=1: Obtain the value of the geometric distribution probability $P_{BIN}(X = m; p)$ for PO<br>ISW=2: Obtain the value of the geometric distribution cumulative distribution function $P_{BIN}(X \leq m; p)$ for PO |
| 5 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) 0.0<PI<1.0

(b) M≥0

(c) ISW∈{1, 2}

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3200 | Restriction (b) was not satisfied. | |
| 3300 | Restriction (c) was not satisfied. | |

(6) **Notes**
None

(7) **Example**

(a) Problem
Let the probability that an event will occur $p$=0.3 and the number of trials $m$=3 and obtain the values of the binomial distribution probability $P_{BIN}(X = m; p)$ and cumulative distribution function $P_{BIN}(X \leq m; p)$.

(b) Input data
PI=0.3 and M=3.

(c) Main program

```
      PROGRAM B1DDGO
! *** MAIN - GEOMETRIC DISTRIBUTION ***
      REAL(8) PI,PO
      INTEGER          M,IERR,ISW
!
!CCCC INPUT DATA
!
      M  = 3
      PI = 0.3
!
!CCCC ISW = 1
!
      ISW = 1
      WRITE(6,1000)
      WRITE(6,1100)
      WRITE(6,1200) 'M  ',M
      WRITE(6,1200) 'ISW',ISW
      WRITE(6,1300) 'PI ',PI
      CALL D1DDGO( M, PI, PO, ISW, IERR )
      WRITE(6,1700)
      WRITE(6,1200) 'IERR',IERR
      WRITE(6,1400) PO
!
!CCCC ISW = 2
!
      ISW = 2
      WRITE(6,1100)
      WRITE(6,1200) 'M  ',M
      WRITE(6,1200) 'ISW',ISW
      WRITE(6,1300) 'PI ',PI
      CALL D1DDGO( M, PI, PO, ISW, IERR )
      WRITE(6,1700)
      WRITE(6,1200) 'IERR',IERR
      WRITE(6,1500) PO
!
 1000 FORMAT( 1X, 2X, '*** D1DDGO ***',/)
 1100 FORMAT( 1X, 3X, '** INPUT **',/)
 1200 FORMAT( 6X, A, 2X, '=', 2X, I5)
 1300 FORMAT( 6X, A, 2X, '=', 2X, F5.2)
 1400 FORMAT( 6X,'VALUE OF P.D.F. OF GEOMETRIC DISTRIBUTION= ',D17.10,/)
 1500 FORMAT( 6X,'VALUE OF C.D.F. OF GEOMETRIC DISTRIBUTION= ',D17.10,/)
 1700 FORMAT( 1X, 3X, '** OUTPUT **',/)
!
      STOP
      END
```

(d) Output results

```
*** D1DDGO ***

 ** INPUT **

   M    =      3
   ISW  =      1
   PI   =   0.30
 ** OUTPUT **

   IERR  =      0
   VALUE OF P.D.F. OF GEOMETRIC DISTRIBUTION=  0.1028999988D+00

 ** INPUT **

   M    =      3
   ISW  =      2
   PI   =   0.30
 ** OUTPUT **

   IERR  =      0
   VALUE OF C.D.F. OF GEOMETRIC DISTRIBUTION=  0.7599000164D+00
```

### 3.3.3 D1DDPO, R1DDPO Poisson Distribution

(1) **Function**

Given the mean $\lambda$ and random variable $k$, the D1DDPO or R1DDPO obtains the value of the probability $Pr.\{X = k\}$ or cumulative distribution function $F(k)$ of a Poisson distribution, which are defined by the following expressions.

$$Pr.\{X = k\} = e^{-\lambda}\frac{\lambda^k}{k!} \qquad (k = 0, 1, 2, \cdots; \lambda > 0)$$

$$F(k) = \sum_{i=0}^{k} Pr.\{X = i\} = e^{-\lambda}\sum_{i=0}^{k}\frac{\lambda^k}{k!}$$

(2) **Usage**

Double precision:
CALL D1DDPO (XE, K, XO, ISW, IERR)
Single precision:
CALL R1DDPO (XE, K, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | XE | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Mean $\lambda$. |
| 2 | K | I | 1 | Input | Value of random variable $k$. |
| 3 | XO | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Output | Value of the probability $Pr.\{X = k\}$ or cumulative distribution function $F(k)$ of the Poisson distribution. |
| 4 | ISW | I | 1 | Input | Processing switch<br>ISW=1:Obtain the value of the probability $Pr.\{X = k\}$ of the Poisson distribution for XO<br>ISW=2:Obtain the value of the cumulative distribution function $F(k)$ of the Poisson distribution for XO |
| 5 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW$\in\{1, 2\}$

(b) 0.0<XE

(c) 0$\leq$K

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) or (c) was not satisfied. | |

(6) **Notes**

(a) When K > 1000, the cumulative distribution function $F(k)$ is obtained according to the Peizer-Pratt approximation.

(7) **Example**

(a) Problem
Let the mean $\lambda = 3.0$ and the random variable $k = 2$ and obtain the values of the probability $Pr.\{X = k\}$ and cumulative distribution function $F(k)$ of the Poisson distribution.

(b) Input data
XE=3.0 and K=2.

(c) Main program

```
      PROGRAM B1DDPO
! *** EXAMPLE OF D1DDPO ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER ISW,K
      REAL(8) XE
!
      XE=3.0D0
      K=2
      WRITE(6,1000)
      WRITE(6,2000) XE
      WRITE(6,2010) K
      WRITE(6,3000)
      ISW=1
      CALL D1DDPO(XE,K,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) XO
!
      ISW=2
      CALL D1DDPO(XE,K,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1DDPO ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'XE = ',F4.1)
 2010 FORMAT(9X,' K = ',I4)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'VALUE OF P.D.F = ',D17.10)
 5010 FORMAT(9X,'VALUE OF C.D.F = ',D17.10)
      END
```

(d) Output results

```
     *** D1DDPO ***
      ** INPUT **
         XE =  3.0
          K =    2


      ** OUTPUT **
         IERR =    0
         VALUE OF P.D.F =  0.2240418077D+00
         IERR =    0
         VALUE OF C.D.F =  0.4231900811D+00
```

### 3.3.4 D1DDHG, R1DDHG Hypergeometric Distribution

(1) **Function**

Assume that there is a lot of size $N$ in which $M$ of the $N$ articles are inferior goods and $N-M$ articles are of good quality. The D1DDHG or R1DDHG obtains the values of the hypergeometric distribution probability $Pr.\{X = k\}$ and cumulative distribution function $F(k)$ corresponding to the probability distribution in which $k$ inferior goods appear when an arbitrary sample of size $n$ is extracted from this lot. The hypergeometric distribution probability $Pr.\{X = k\}$ and cumulative distribution function $F(k)$ are defined by the following equations.

$$Pr.\{X=k\} = \begin{cases} \dfrac{\begin{pmatrix} M \\ k \end{pmatrix}\begin{pmatrix} N-M \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} & k=0,1,2,\cdots,\min\{M,n\} \\ 0 & \text{otherwise} \end{cases}$$

$$F(k) = \sum_{i=0}^{k} Pr.\{X=i\} = \frac{\sum_{i=0}^{k} \begin{pmatrix} M \\ i \end{pmatrix}\begin{pmatrix} N-M \\ n-i \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

(2) **Usage**

Double precision:
CALL D1DDHG (NN, M, N, K, XO, ISW, IERR)
Single precision:
CALL R1DDHG (NN, M, N, K, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | NN | I | 1 | Input | Size of population $N$ |
| 2 | M | I | 1 | Input | Number of inferior goods in population $M$ |
| 3 | N | I | 1 | Input | Size of sample $n$ |
| 4 | K | I | 1 | Input | Number of inferior goods in sample $k$ |
| 5 | XO | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Output | Value of hypergeometric distribution probability $Pr.\{X = k\}$ or cumulative distribution function $F(k)$ |
| 6 | ISW | I | 1 | Input | ISW=0: Obtain $Pr.\{X = k\}$<br>ISW=1: Obtain $F(k)$ |
| 7 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{0, 1\}$

(b) $0 \leq \mathrm{M} \leq \mathrm{NN}$

(c) $1 \leq \mathrm{N} \leq \mathrm{NN}$

(d) $0 \leq \mathrm{K} \leq \min(\mathrm{M}, \mathrm{N})$

(e) $\mathrm{M} + \mathrm{N} - \mathrm{NN} \leq \mathrm{K}$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (d) was not satisfied. | 0.0 is set for XO. (When ISW = 0 or K < 0.)<br>1.0 is set for XO. (When min(M, N) < K.) |
| 1010 | Restriction (e) was not satisfied. | 0.0 is set for XO. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |

(6) **Notes**

(a) The mathematical expectation and variance of the hypergeometric distribution are given by the following equations.

$$E(X) = np,\ \sigma^2(X) = \frac{N-k}{N-1}np(1-p) \quad (p = \frac{\mathrm{M}}{\mathrm{N}})$$

(b) Since the hypergeometric distribution probability $Pr.\{X = k\}$ satisfies the following relationship

$$Pr.\{X = k\} = \frac{\begin{pmatrix} M \\ k \end{pmatrix}\begin{pmatrix} N-M \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} \simeq \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^k,\ (p = \frac{M}{N}, \frac{n}{N} \ll 1)$$

it can be approximated by a binomial distribution $B(n, p)$ when the population is sufficiently large.

(7) **Example**

(a) Problem
Let $N = 1000$, $M = 50$, $n = 20$ and $k = 1$ and obtain the values of the hypergeometric distribution probability $Pr.\{X = k\}$ and cumulative distribution function $F(k)$.

(b) Input data
NN = 1000, M = 20, N = 20 and K = 1.

(c) Main program

```
      PROGRAM B1DDHG
! *** EXAMPLE OF D1DDHG ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER NN,M,N,K,ISW
      REAL(8) XO
!
      NN=1000
      M=50
```

```
      N=20
      K=1
      WRITE(6,1000)
      WRITE(6,2000) NN
      WRITE(6,2010) M
      WRITE(6,2020) N
      WRITE(6,2030) K
      WRITE(6,3000)
      ISW=0
      CALL D1DDHG(NN,M,N,K,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) XO
!
      ISW=1
      CALL D1DDHG(NN,M,N,K,XO,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5010) XO
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D1DDHG ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'NN = ',I5)
 2010 FORMAT(9X,'M  = ',I5)
 2020 FORMAT(9X,'N  = ',I5)
 2030 FORMAT(9X,'K  = ',I5)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'VALUE OF P.D.F. = ',D17.10)
 5010 FORMAT(9X,'VALUE OF C.D.F. = ',D17.10)
      END
```

(d) Output results

```
     *** D1DDHG ***
      ** INPUT **
         NN =  1000
         M  =    50
         N  =    20
         K  =     1


      ** OUTPUT **
         IERR =    0
         VALUE OF P.D.F. =  0.3811717030D+00
         IERR =    0
         VALUE OF C.D.F. =  0.7360425584D+00
```

### 3.3.5 D1DDHN, R1DDHN Negative Hypergeometric Distribution

(1) **Function**

Assume that there is a lot of size $NN$ in which $M$ of the $NN$ articles are inferior goods and $NN-M$ articles are of good quality. Sampling from this lot is continued until $n$ inferior goods are extracted. The negative hypergeometric distribution probability $Pr.\{X = k\}$ is defined as the probability of such occurrences that exactly $k$ goods has been extracted at this time. The D1DDHN or R1DDHN obtains the values of the negative hypergeometric distribution probability $Pr.\{X = k\}$ and cumulative distribution function $F(k)$ corresponding to the probability distribution in which exactly $k$ goods have been extracted from this lot when the $n$-th inferior good appears. The negative hypergeometric distribution probability $Pr.\{X = k\}$ and cumulative distribution function $F(k)$ are defined by the following equations.

$$Pr.\{X = k\} = \begin{cases} \dfrac{\begin{pmatrix} M \\ NN-1 \end{pmatrix}\begin{pmatrix} NN-M \\ k-n \end{pmatrix}}{\begin{pmatrix} NN \\ k-1 \end{pmatrix}} \times \dfrac{M-n+1}{NN-k+1} \\ = \dfrac{\begin{pmatrix} k-1 \\ n-1 \end{pmatrix}\begin{pmatrix} NN-k \\ M-n \end{pmatrix}}{\begin{pmatrix} NN \\ M \end{pmatrix}} & \text{when } k = n, n+1, n+2, \cdots, NN-M+n \\ 0 & \text{Otherwise.} \end{cases}$$

$$F(k) = \sum_{i=n}^{k} Pr.\{X = i\} = \frac{\sum_{i=0}^{k} \begin{pmatrix} i-1 \\ n-1 \end{pmatrix}\begin{pmatrix} NN-i \\ M-n \end{pmatrix}}{\begin{pmatrix} NN \\ M \end{pmatrix}}$$

(2) **Usage**

Double precision:

CALL D1DDHN (NN, M, N, K, XO, ISW, IERR)

Single precision:

CALL R1DDHN (NN, M, N, K, XO, ISW, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: { INTEGER(4) as for 32bit Integer; INTEGER(8) as for 64bit Integer }
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | NN | I | 1 | Input | Size $NN$ of population |
| 2 | M | I | 1 | Input | Number of inferior goods $M$ in population |
| 3 | N | I | 1 | Input | Number $n$ of inferior goods extracted from the lot |
| 4 | K | I | 1 | Input | Total number $k$ of goods extracted from the lot |
| 5 | XO | {D, R} | 1 | Output | Value of negative hypergeometric distribution probability $Pr.\{X = k\}$ or cumulative distribution function $F(k)$ |
| 6 | ISW | I | 1 | Input | ISW=0: Obtain $Pr.\{X = k\}$<br>ISW=1: Obtain $F(k)$ |
| 7 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{0, 1\}$

(b) $1 \leq$ N $\leq$ M $\leq$ NN

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | K < N | 0.0 is set for XO. |
| 1010 | K > NN − M + N | If ISW=0 then 0.0 is set for XO.<br>If ISW=1 then 1.0 is set for XO. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |

(6) **Notes**

(a) If $M = 1$ and if $n = 1$:
$Pr.\{X = k\} = \frac{1}{NN} \quad k = 1, 2, \cdots, NN$
When $n \neq 1$ : $Pr.\{X = k\} = 0$.

(b) If $M > 1$, define $N1$ as the greatest integer no more than

$$\frac{NN \times (n-1)}{M-1} + 1.$$

Then the following stands for the distribution probability $Pr.\{X = k\}$:

$$Pr.\{X = n\} \leq Pr.\{X = n+1\} \leq Pr.\{X = n+2\} \leq \cdots \leq Pr.\{X = N1\}.$$

$$Pr.\{X = N1\} \geq Pr.\{X = N1+1\} \geq Pr.\{X = N1+2\} \geq \cdots \geq Pr.\{X = NN - M + n\}.$$

(7) **Example**

(a) Problem

Let $N = 30$, $M = 25$ and $n = 10$ and obtain the values of the hypergeometric distribution probability $Pr.\{X = k\}$ and cumulative distribution function $F(k)$ for $k = 10, 11, \cdots, 15$.

(b) Input data

NN = 30, M = 25 and N = 10.

(c) Main program

```
      PROGRAM B1DDHN
! *** EXAMPLE OF D1DDHN ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER NN,M,N,K,ISW
      INTEGER KINC,K1,K2,KMIN,KMAX
      REAL(8) XO
!
      NN=30
      M=25
      N=10
      KMIN=N
      KMAX=NN-M+N
      K1=KMIN
      K2=KMAX
      KINC=1
      WRITE(6,1000)
      WRITE(6,2000) NN
      WRITE(6,2010) M
      WRITE(6,2020) N
      WRITE(6,3000)
!
      DO 100 K=K1,K2,KINC
        WRITE(6,2030) K
        ISW=0
        CALL D1DDHN(NN,M,N,K,XO,ISW,IERR)
        WRITE(6,4000) IERR
        WRITE(6,5000) XO
!
        ISW=1
        CALL D1DDHN(NN,M,N,K,XO,ISW,IERR)
        WRITE(6,4000) IERR
        WRITE(6,5010) XO
  100 CONTINUE
!
      STOP
!
 1000 FORMAT(/,1X,'     *** D1DDHN ***',/,/,&
      1X,'     ** INPUT **')
 2000 FORMAT(1X,'        NN = ',I5)
 2010 FORMAT(1X,'        M  = ',I5)
 2020 FORMAT(1X,'        N  = ',I5)
 2030 FORMAT(/,1X,'        *** K  = ',I5,' ****')
 3000 FORMAT(/,/,1X,'     ** OUTPUT **')
 4000 FORMAT(1X,'        IERR = ',I4)
 5000 FORMAT(1X,'        VALUE OF P.D.F. = ',F17.10)
 5010 FORMAT(1X,'        VALUE OF C.D.F. = ',F17.10)
      END
```

(d) Output results

```
*** D1DDHN ***

** INPUT **
   NN =    30
   M  =    25
   N  =    10


** OUTPUT **

   *** K  =    10 ****
   IERR =    0
   VALUE OF P.D.F. =      0.1087954191
   IERR =    0
   VALUE OF C.D.F. =      0.1087954191

   *** K  =    11 ****
   IERR =    0
   VALUE OF P.D.F. =      0.2719885479
   IERR =    0
   VALUE OF C.D.F. =      0.3807839670

   *** K  =    12 ****
   IERR =    0
   VALUE OF P.D.F. =      0.3149341080
   IERR =    0
   VALUE OF C.D.F. =      0.6957180750
```

```
*** K  =    13 ****
IERR =    0
VALUE OF P.D.F. =      0.2099560720
IERR =    0
VALUE OF C.D.F. =      0.9056741471

*** K  =    14 ****
IERR =    0
VALUE OF P.D.F. =      0.0802773217
IERR =    0
VALUE OF C.D.F. =      0.9859514687

*** K  =    15 ****
IERR =    0
VALUE OF P.D.F. =      0.0140485313
IERR =    0
VALUE OF C.D.F. =      1.0000000000
```

# Chapter 4

# SAMPLE STATISTICS

## 4.1 INTRODUCTION

This library provides the following functions for calculating sample statistics.

- One Sample Basic Statistics
- Two Samples Basic Statistics
- Geometric Mean
- Moment
- $m$ Samples Basic Statistics
- Harmonic Mean
- Root Mean Square
- Variance-Covariance Matrices
- Variance-Covariance Matrices (Grouped Data)
- Correlation Matrices
- Multiple Correlation Coefficients
- Partial Correlation Coefficients

### 4.1.1 Explanation

(1) **One Sample Basic Statistics**

The sum, mean, standard deviation, median of $n$ observation data $\{x_i\}$ $(i = 1, \cdots, n)$ are defined by the following equations.

Sum :

$$t = \sum_{i=1}^{n} x_i$$

Mean :

$$\bar{x} = \frac{t}{n}$$

Standard deviation :

$$d = \sqrt{\frac{1}{\alpha}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Here, $\alpha$ is $n-1$ when an unbiased estimate is used or $\alpha$ is $n$ when a sample variance is used.

Assume that we are given $m$ classes of equal width $C_{i+1} = [c_i, c_i + h)$ $(i = 1, 2, \cdots, m)$ and the two classes $C_1 = (-\infty, c_1)$ and $C_{m+2} = [c_m + h, \infty)$, and that the following relationships are satisfied.

$$\begin{aligned} c_1 &= c_{min} \\ c_{i+1} &= c_i + h \qquad i = 1, 2, \cdots, m \\ c_{max} &= c_m + h \end{aligned}$$

Here, $h$ is the class width, which is defined as follows.

$$h = \frac{c_{max} - c_{min}}{m}$$

At this time, if the frequency of observation data in class $C_i$ $(i = 1, 2, \cdots, m+2)$ is $e_i$, the frequency percentage $f_i$ is defined as follows.

$$f_i = \frac{e_i}{n} \times 100 \qquad i = 1, \cdots, m+2$$

Median :

$$p = a_{i-1} + \frac{h \times (N/2 - s_{i-1})}{g_i}$$

Here, $a_{i-1}$ is lower boundary limit of class of median, $g_i$ is frequency of class of median, $s_{i-1}$ is frequency to class of median.

(2) **Two Samples Basic Statistics**

The sum, mean, and standard deviation of $n$ two-sample observation data $\{x_i\}$ $(i = 1, \cdots, n)$ and $\{y_i\}$ $(i = 1, \cdots, n)$ are defined by the following equations.

Sum :

$$S_x = \sum_{i=1}^{n} x_i$$

$$S_y = \sum_{i=1}^{n} y_i$$

Mean :

$$\bar{x} = \frac{S_x}{n}$$

$$\bar{y} = \frac{S_y}{n}$$

Standard deviation :

$$SD_x = \sqrt{\frac{1}{\alpha}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Here, $\alpha$ is $n-1$ when an unbiased estimate is used or $\alpha$ is $n$ when a sample variance is used.
Assume that for $x_i$, we are given $m_x$ classes of equal width ${C_{i+1}}^{(x)} = [c_i, c_i + h_x)$ $(i = 1, 2, \cdots, m_x)$ and the two classes ${C_1}^{(x)} = (-\infty, c_1)$ and ${C_{m+2}}^{(x)} = [c_m + h_x, \infty)$. Assume that for $y_i$, we are given $m_y$ classes of equal width ${C_{i+1}}^{(y)} = [d_i, d_i + h_y)$ $(i = 1, 2, \cdots, m_y)$ and the two classes ${C_1}^{(y)} = (-\infty, d_1)$ and ${C_{m+2}}^{(y)} = [d_m + h_y, \infty)$. Furthermore, assume that $(m_x+2)\cdot(m_y+2)$ classes on the $x-y$ plane defined as the direct product of $x_i$ and $y_i$ are defined by $C_{i,j} = (C_i^{(x)}, C_j^{(y)})$ $(i = 1, 2, \cdots, m_x + 2; j = 1, 2, \cdots, m_y + 2)$ and that the following relationships are satisfied.

$$\begin{aligned} c_1 &= c_{min} \\ c_{i+1} &= c_i + h_x \qquad i = 1, 2, \cdots, m_x \\ c_{max} &= c_{m_x} + h_x \end{aligned}$$

$$\begin{aligned} d_1 &= d_{min} \\ d_{i+1} &= d_i + h_y \qquad i = 1, 2, \cdots, m_y \\ d_{max} &= d_{m_y} + h_y \end{aligned}$$

Here, $h_x$ and $h_y$ are the class widths, which are defined as follows.

$$h_x = \frac{c_{max} - c_{min}}{m_x}$$

$$h_y = \frac{d_{may} - d_{min}}{m_y}$$

At this time, if the frequency of observation data in class $C_{i,j}$ is $e_{i,j}$, the frequency percentage $\{f_{ij}\}$ is defined as follows.

$$f_{ij} = \frac{e_{ij}}{n} \times 100 \qquad i = 1, \cdots, m_y + 2, \quad j = 1, \cdots, m_x + 2$$

(3) **Geometric Mean**

Given a sample consisting of $n$ observed values $\{x_i\}(i = 1, \cdots, n)$, the geometric mean and its standard deviation are defined by the following equations.

Geometric mean :

$$\mathrm{GM} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$$

Standard deviation :

$$\mathrm{GSD} = \exp\left(\sqrt{\frac{\sum_{i=1}^{n}(\log x_i)^2 - n(\log \mathrm{GM})^2}{\alpha}}\right)$$

Here, $\alpha$ is $n-1$ when an unbiased estimate is used or $\alpha$ is $n$ when a sample variance is used.

(4) **Moment**

Assume that we are given $m$ classes of equal width $C_i = [x_i, x_i + h)$ $(i = 1, 2, \cdots, m)$, and that the following relationships are satisfied.

$$\begin{aligned} x_1 &= x_{min} \\ x_{i+1} &= x_i + h \qquad i = 1, 2, \cdots, m \\ x_{max} &= x_m + h \end{aligned}$$

Here, $h$ is the class width, which is defined as follows.

$$h = \frac{x_{max} - x_{min}}{m}$$

At this time, if the frequency of observation data in class $C_i$ is $f_i$, the linear moment about the origin and the moment of order $r$ about the mean value are defined as follows. Linear moment about the origin :

$$\mu'_1 = \frac{\sum_{i=1}^{m} f_i \hat{x}_i}{\sum_{i=1}^{m} f_i}$$

Moment of order $r$ about the mean value :

$$\mu_r = \frac{\sum_{i=1}^{m}[f_i(\hat{x}_i - \mu'_1)^r]}{\sum_{i=1}^{m} f_i} \qquad (r = 2, 3, \cdots)$$

$\hat{x}_i$, which represents the class value of each class, is defined as follows.

$$\hat{x}_i = x_{min} + (i - 0.5)h$$

(5) $m$ **Samples Basic Statistics**

Given $m$ samples consisting of $n$ observed values $\{x_{ij}\}$, $(i = 1, \cdots, n;\ j = 1, \cdots, m)$, the basic statistics (sum, mean, sum of squares of deviation, variance, and standard deviation) for each sample are defined by the following equations.

Sum :

$$t_j = \sum_{i=1}^{n} x_{ij}, \qquad (j = 1, \cdots, m)$$

Mean :

$$\bar{x}_j = \frac{t_j}{n}, \quad (j = 1, \cdots, m)$$

Sum of squares of deviation :

$$s_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2, \quad (j = 1, \cdots, m)$$

Variance :

$$v_j = \frac{s_j}{\alpha}, \quad (j = 1, \cdots, m)$$

Standard deviation :

$$d_j = \sqrt{v_j}, \quad (j = 1, \cdots, m)$$

Here, $\alpha$ is $n-1$ when an unbiased estimate is used or $\alpha$ is $n$ when a sample variance is used.

(6) **Harmonic Mean**

Given a sample consisting of $n$ observed values $\{x_i\}(i = 1, \cdots, n)$, the harmonic mean is defined by the following equations.

Harmonic mean :

$$\mathrm{HM} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}}$$

(7) **Root Mean Square**

Given a sample consisting of $n$ observed values $\{x_i\}(i = 1, \cdots, n)$, the root mean square is defined by the following equations.

Root mean square :

$$\mathrm{SM} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$$

(8) **Variance-Covariance Matrices**

Given $m$ samples consisting of $n$ observed values $\{x_{ki}\}(k = 1, \cdots, n; i = 1, \cdots, m)$, the variance-covariance among the samples $d_{i,j}$ is defined as follows.

$$d_{ij} = \frac{s_{ij}}{\alpha} \qquad i, j = 1, \cdots, m$$

Here, $\alpha$ is $n$ when a sample covariance is used or $\alpha$ is $n-1$ when an unbiased covariance is used. $s_{ij}$ is defined as follows.

$$s_{ij} = \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j), \qquad i = 1, \cdots, m; j = 1, \cdots, m$$

$\bar{x}_j$ is defined as follows.

$$\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}, \qquad j = 1, \cdots, m$$

(9) **Variance-Covariance Matrices (Grouped Data)**

Assume there are $g$ groups, and for each group, we are given $m$ samples consisting of $n_r$ observed values $\{x_{ij}^{(r)}\}(i = 1, \cdots, n_r; j = 1, \cdots, m; r = 1, \cdots, g)$. The variance-covariance over all groups $d_{i,j}$ is defined as follows.

$$d_{ij} = \frac{\sum_{r=1}^{g} s_{ij}^{(r)}}{\sum_{r=1}^{g} \alpha_r} \qquad i, j = 1, \cdots, m$$

For each group, $s_{ij}^{(r)}$ (deviation sum of product matrix) is defined as follows.

$$s_{ij}^{(r)} = \sum_{k=1}^{n_r} (x_{ki}^{(r)} - \bar{x}_i^{(r)})(x_{kj}^{(r)} - \bar{x}_j^{(r)}) \qquad i, j = 1, \cdots, m$$

Here, $\alpha_r$ is $n_r$ when a sample covariance is used or $\alpha_r$ is $n_r - 1$ when an unbiased covariance is used. Also, $\bar{x}_i^{(r)}$ is the mean of each group.

(10) **Correlation Matrices**

Given $m$ samples consisting of $n$ observed values $\boldsymbol{x}_i = \{x_{ki}\}(k = 1, \cdots, n; i = 1, \cdots, m)$, the correlation coefficient $r_{ij}$ between samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is defined as follows.

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii} \cdot s_{jj}}}, \qquad i = 1, \cdots, m; j = 1, \cdots, m$$

$s_{ij}$ is defined as follows.

$$s_{ij} = \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j), \qquad i = 1, \cdots, m; j = 1, \cdots, m$$

$\bar{x}_j$ is defined as follows.

$$\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}, \qquad j = 1, \cdots, m$$

Also, the matrix $R = (r_{ij})$ is called the correlation coefficient matrix. The correlation coefficients have the following properties.

- $|r_{ij}| \leq 1$
- $r_{ij} = r_{ji}$

(11) **Multiple Correlation Coefficients** and **Partial Correlation Coefficients**

For $m$ variates $X_i$ $(i = 1, \cdots, m)$, let $\bar{x}_i$ be the mean of each variate, ${\sigma_i}^2$ be the variance, and $r_{ij}$ be the correlation coefficient of variates $X_i$ and $X_j$. The multiple correlation coefficient $r_{p \cdot 1, \cdots, n}$, which is defined as the maximum value of the correlation coefficients of $X_p$ $(l \geq p \geq m)$ and the linear expression $U = b + \sum_{i=1}^{l} c_i X_i$ of $X_1, \cdots, X_l$, can be calculated as follows.

$$r_{p \cdot 1, \cdots, l} = \sqrt{1 - \frac{\Delta}{\Delta_{pp}}}$$

Here, $\Delta$ and $\Delta_{ij}$ are the determinant and cofactor matrix of the matrix having the correlation coefficients $r_{ij}$ $(i, j = 1, \cdots, l, p)$ as elements. Representing the value of $U$ corresponding to this maximum value by $\hat{X}_p$, this can be expressed as follows.

$$\hat{X}_p = \bar{x}_p - \sum_{i=1}^{l} \sigma_p \frac{\Delta_{ip}}{\Delta_{pp}} \frac{X_i - \bar{x}_i}{\sigma_i}$$

$X_p - \hat{X}_p$, which is considered to be the variate when the influence of $X_1, \cdots, X_l$ has been eliminated from $X_p$, is called the remainder. The partial correlation coefficient $r_{p,q\cdot 1,\cdots,l}$ $(l \geq p, q \geq m)$, which is defined as the correlation coefficient of $X_p - \hat{X}_p$ and $X_q - \hat{X}_q$, can be calculated as follows.

$$r_{p,q\cdot 1,\cdots,l} = -\frac{\Delta_{pq}}{\sqrt{\Delta_{pp}\Delta_{qq}}}$$

### 4.1.2 Reference Bibliography

(1) Spiegel, M. R. , “Theory and Problems of Probability and Statistics”, Schaum’s outline series, McGraw-Hill, Inc. (1975).

## 4.2 BASIC STATISTICS

### 4.2.1 D2BA1T, R2BA1T
### One Sample Basic Statistics

(1) **Function**

The D2BA1T or R2BA1T obtains the minimum, maximum, sum, mean, standard deviation, median, frequency, and frequency percentage of one-sample observation data. It also obtains the minimum, maximum, sum, mean, standard deviation, median, frequency, and frequency percentage when observation data is added.

The sum, mean, standard deviation and median of $n$ observation data $\{x_i\}$ $(i = 1, \cdots, n)$ are defined by the following equations.

Sum :

$$t = \sum_{i=1}^{n} x_i$$

Mean :

$$\bar{x} = \frac{t}{n}$$

Standard deviation :

$$d = \sqrt{\frac{1}{\alpha} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Here, $\alpha$ is $n-1$ when an unbiased estimate is used or $\alpha$ is $n$ when a sample variance is used.

Also, consider $m$ classes of equal width $C_{i+1} = [c_i, c_i+h)$ $(i = 1, 2, \cdots, m)$ and the two classes $C_1 = (-\infty, c_1)$ and $C_{m+2} = [c_m + h, \infty)$, and assume that the following relationships are satisfied.

$$\begin{aligned} c_1 &= c_{min} \\ c_{i+1} &= c_i + h \quad\quad i = 1, 2, \cdots, m \\ c_{max} &= c_m + h \end{aligned}$$

The D2BA1T or R2BA1T obtains the frequency $e_i$ of the observation data in each class and the frequency percentage $f_i$, which is defined as follows.

$$f_i = \frac{e_i}{n} \times 100 \quad\quad i = 1, \cdots, m+2$$

$h$ is the class width, which is defined as follows.

$$h = \frac{c_{max} - c_{min}}{m}$$

Median :

$$p = a_{i-1} + \frac{h \times (N/2 - s_{i-1})}{g_i}$$

Here, $a_{i-1}$ is the lower bound of the class of the median, $g_i$ is the frequency of the class of the median, and $s_{i-1}$ is the frequency to the class of the median.

(2) **Usage**

Double precision:

CALL D2BA1T (A, N, NC, BL, BU, NS, STAT, IFRQ1, FREQ2, ISW, IWK, WK, IERR)

Single precision:

CALL R2BA1T (A, N, NC, BL, BU, NS, STAT, IFRQ1, FREQ2, ISW, IWK, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: {INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer}
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | {D, R} | N | Input | Sequence of observed values $\{x_i\}$ |
| 2 | N | I | 1 | Input | Number of observed values $n$ |
| 3 | NC | I | 1 | Input | Number of classes $m+2$ |
| 4 | BL | {D, R} | 1 | Input | Value of $c_{min}$ |
| 5 | BU | {D, R} | 1 | Input | Value of $c_{max}$ |
| 6 | NS | I | 1 | Input | Number of observed values before adding observed values (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Number of observed values $n$ |
| 7 | STAT | {D, R} | 6 | Output | Basic statistics of observed values (See Note (b)) |
| 8 | IFRQ1 | I | NC | Output | Frequency for each class $\{e_i\}$ |
| 9 | FREQ2 | {D, R} | NC | Output | Frequency percentage for each class $\{f_i\}$ |
| 10 | ISW | I | 1 | Input | Processing switch<br>0: Perform calculations using an unbiased estimate (no previously established basic statistics)<br>1: Perform calculations using an unbiased estimate (added observed values)<br>2: Perform calculations using a sample variance (no previously established basic statistics)<br>3: Perform calculations using a sample variance (added observed values) |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 11 | IWK | I | NC | Work | Work area |
| 12 | WK | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | NC + N | Work | Work area |
| 13 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW = 0, 1, 2, 3

(b) N $\geq$ 1

(c) NC $\geq$ 3

(d) BU > BL

(e) NS $\geq$ 1 (When ISW=1 or 3)

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (a) was not satisfied. | Processing continues with ISW=0. |
| 1010 | An attempt was made to obtain the standard deviation when ISW=0 and N=1. | The absolute value maximum that can be represented is set for the standard deviation. |
| 1020 | BU<BL occurred. | BU and BL are switched, and processing continues. |
| 3000 | Restriction (b), (c) was not satisfied. | Processing is aborted. |
| 3010 | BU=BL occurred. | |
| 3020 | Restriction (e) was not satisfied. | |

(6) **Notes**

(a) When an observed value < BL, it is included in the frequency for IFRQ1(1).
When an observed value $\geq$ BU, it is included in the frequency for IFRQ1(NC).

(b) The basic statistics are stored as follows in the array STAT.
STAT(1): Minimum of observed values
STAT(2): Maximum of observed values
STAT(3): Sum of observed values $t$
STAT(4): Mean of observed values $\bar{x}$
STAT(5): Standard deviation of observed values $d$
STAT(6): Median of observed values $p$

(c) To obtain the basic statistics when observed values are added, use the contents of NC, BL, BU, NS, STAT, IFRQ1, FREQ2 and WK which were calculated before adding the observed values, set the added observed values for A and the number of added observed values for N, set ISW to 1 or 3, and perform the calculation. However, when obtaining the standard deviation, you must set the ISW value so that the calculation is performed using a sample variance following a calculation that used a sample variance the previous time or so that the calculation is performed using an unbiased estimate following a calculation that used an unbiased estimate the previous time.

(d) When there are an extremely large number of data values that are widely dispersed, better results are obtained by grouping them into data having absolute values of the same relative size and adding them to the samples in increasing order of size.

(e) Statistics obtained when calculations are performed using an unbiased estimate can be applied to a population for which sampling with replacement is performed from an infinite or finite population. Statistics obtained when calculations are performed using a sample variance can be applied to a population for which the population and sample match.

(7) **Example**

(a) Problem

Obtain the minimum, maximum, sum, mean, standard deviation, frequencies, and frequency percentages of the following one-sample observation data.

$$\{x_i\} = \{-10, -9, -8, -6, -5, -4, -3, -2, -1, 0, 1\}$$

Also, obtain the minimum, maximum, sum, mean, standard deviation, frequencies, and frequency percentages when the following one-sample observation data is added.

$$\{y_i\} = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$$

(b) Input data

First processing :

One-sample observation data $\{x_i\}$, N = 11, NC = 7, BL = −6.5, BU = 5.8 and ISW = 0.

Second processing :

One-sample observation data $\{y_i\}$, N = 10, NC = 7, BL = −6.5, BU = 5.8 and ISW = 1.

(c) Main program

```
      PROGRAM B2BA1T
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 100, NC = 7 )
      DIMENSION IFRQ1(NC),IWK(NC)
      DIMENSION A(NA),STAT(6),FREQ2(NC),WK(NA+NC)
!
      WRITE(6,6000)
      IERR =    0
      ISW  =    0
      READ(5,*) N
      READ(5,*) BL
      READ(5,*) BU
      DO 100 I=1,N
         READ(5,*) A(I)
  100 CONTINUE
      WRITE(6,6010) ISW,N,NC,BL,BU
      WRITE(6,6020) (A(I),I=1,N)
      CALL D2BA1T(A,N,NC,BL,BU,NS,STAT,IFRQ1,FREQ2,ISW,IWK,WK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS,(STAT(I),I=1,6)
      WRITE(6,6050)
      DO 110 I=1,NC
         WRITE(6,6060) I,IFRQ1(I),FREQ2(I)
  110 CONTINUE
!
      WRITE(6,6070)
      IERR =  0
      ISW  =  1
      READ(5,*) N
      DO 120 I=1,N
         READ(5,*) A(I)
  120 CONTINUE
      WRITE(6,6010) ISW,N,NC,BL,BU
      WRITE(6,6020) (A(I),I=1,N)
      CALL D2BA1T(A,N,NC,BL,BU,NS,STAT,IFRQ1,FREQ2,ISW,IWK,WK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS,(STAT(I),I=1,6)
      WRITE(6,6050)
      DO 130 I=1,NC
         WRITE(6,6060) I,IFRQ1(I),FREQ2(I)
  130 CONTINUE
```

```
!
      STOP
 6000 FORMAT(  ' *** D2BA1T ***',/,&
             /,3X,'*** FIRST PROCESSING ***',/,&
             /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'ISW = ',I6,/,&
             /,7X,'N = ',I6,5X,'NC = ',I6,/,&
             /,7X,'BL = ',F11.2,5X,'BU = ',F11.2)
 6020 FORMAT( /,7X,'OBSERVATIONS',/,/,&
                  5(6X,5(2X,F11.2),/))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
             /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'TOTAL SAMPLE SIZE  =      ',I6,/,&
             /,7X,'VMIN               = ',F11.2,/,&
             /,7X,'VMAX               = ',F11.2,/,&
             /,7X,'SUM                = ',F11.2,/,&
             /,7X,'VMEAN              = ',D15.8,/,&
             /,7X,'STANDARD DEVIATION = ',D15.8,/,&
             /,7X,'VMEDIAN            = ',D15.8)
 6050 FORMAT( /,7X,'VALUE OF EACH CLASSES',/,&
             /,10X,'CLASS',5X,'MEMBER NUMBER',5X,'PERCENTAGE',/,&
              7X,42('-'))
 6060 FORMAT(  7X,I6,8X,I6,8X,F11.2)
 6070 FORMAT(/,/,3X,'*** CONTINUATION PROCESSING ***',/,&
             /,3X,'** INPUT **')
      END
```

(d) Output results

```
*** D2BA1T ***

  *** FIRST PROCESSING ***

  ** INPUT **

      ISW =      0

      N =     11     NC =      7

      BL =        -6.50     BU =         5.80

      OBSERVATIONS

            -10.00        -9.00        -8.00        -6.00        -5.00
             -4.00        -3.00        -2.00        -1.00         0.00
              1.00

  ** OUTPUT **

      IERR =      0

      TOTAL SAMPLE SIZE  =          11

      VMIN               =      -10.00

      VMAX               =        1.00

      SUM                =      -47.00

      VMEAN              = -0.42727273D+01

      STANDARD DEVIATION =  0.36902821D+01

      VMEDIAN            = -0.36300000D+01

      VALUE OF EACH CLASSES

         CLASS     MEMBER NUMBER     PERCENTAGE
      ------------------------------------------
           1              3              27.27
           2              2              18.18
           3              3              27.27
           4              2              18.18
           5              1               9.09
           6              0               0.00
           7              0               0.00


  *** CONTINUATION PROCESSING ***

  ** INPUT **

      ISW =      1

      N =     10     NC =      7

      BL =        -6.50     BU =         5.80

      OBSERVATIONS

              2.00         3.00         4.00         5.00         6.00
              7.00         8.00         9.00        10.00        11.00
```

```
** OUTPUT **

    IERR =      0

    TOTAL SAMPLE SIZE  =          21

    VMIN               =      -10.00

    VMAX               =       11.00

    SUM                =       18.00

    VMEAN              =  0.85714286D+00

    STANDARD DEVIATION =  0.64287302D+01

    VMEDIAN            =  0.12900000D+01

    VALUE OF EACH CLASSES

       CLASS     MEMBER NUMBER     PERCENTAGE
    ------------------------------------------
         1             3              14.29
         2             2               9.52
         3             3              14.29
         4             2               9.52
         5             3              14.29
         6             2               9.52
         7             6              28.57
```

### 4.2.2 D2BA2S, R2BA2S Two Samples Basic Statistics

(1) **Function**

The D2BA2S or R2BA2S obtains the minimum, maximum, sum, mean, standard deviation, frequency, and frequency percentage of two-sample observation data. It also obtains the minimum, maximum, sum, mean, standard deviation, frequency, and frequency percentage when two-sample observation data is added.

The sum, mean, and standard deviation of $n$ two-sample observation data $\{x_i\}$ $(i = 1, \cdots, n)$ and $\{y_i\}$ $(i = 1, \cdots, n)$ are defined by the following equations.

Sum :

$$S_x = \sum_{i=1}^{n} x_i$$

$$S_y = \sum_{i=1}^{n} y_i$$

Mean :

$$\bar{x} = \frac{S_x}{n}$$

$$\bar{y} = \frac{S_y}{n}$$

Standard deviation :

$$SD_x = \sqrt{\frac{1}{\alpha}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$SD_y = \sqrt{\frac{1}{\alpha}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Here, $\alpha$ is $n-1$ when an unbiased estimate is used or $\alpha$ is $n$ when a sample variance is used.

Also, for $x_i$, consider $m_x$ classes of equal width ${C_{i+1}}^{(x)} = [c_i, c_i + h_x)$ $(i = 1, 2, \cdots, m_x)$ and the two classes ${C_1}^{(x)} = (-\infty, c_1)$ and ${C_{m+2}}^{(x)} = [c_m + h_x, \infty)$, and for $y_i$, consider $m_y$ classes of equal width ${C_{i+1}}^{(y)} = [d_i, d_i + h_y)$ $(i = 1, 2, \cdots, m_y)$ and the two classes ${C_1}^{(y)} = (-\infty, d_1)$ and ${C_{m+2}}^{(y)} = [d_m + h_y, \infty)$, and assume that the following relationships are satisfied.

$$\begin{aligned} c_1 &= c_{min} \\ c_{i+1} &= c_i + h_x \qquad i = 1, 2, \cdots, m_x \\ c_{max} &= c_{m_x} + h_x \end{aligned}$$

$$\begin{aligned} d_1 &= d_{min} \\ d_{i+1} &= d_i + h_y \qquad i = 1, 2, \cdots, m_y \\ d_{max} &= d_{m_y} + h_y \end{aligned}$$

On the $x - y$ plane defined as the direct product of $x_i$ and $y_i$, define $(m_x + 2) \cdot (m_y + 2)$ classes as $C_{i,j} = (C_i^{(x)}, C_j^{(y)})$ $(i = 1, 2, \cdots, m_x + 2; j = 1, 2, \cdots, m_y + 2)$. The D2BA2S or R2BA2S obtains the frequency $\{e_{ij}\}$ of observation data in each class and the frequency percentage $\{f_{ij}\}$, which is defined as follows.

$$f_{ij} = \frac{e_{ij}}{n} \times 100 \quad i = 1, \cdots, m_x + 2, \quad j = 1, \cdots, m_y + 2$$

$h_x$ and $h_y$ are the class widths, which are defined as follows.

$$h_x = \frac{c_{max} - c_{min}}{m_x}$$

$$h_y = \frac{d_{may} - d_{min}}{m_y}$$

(2) **Usage**

Double precision:

CALL D2BA2S (X, N, Y, NCX, NCY, BLX, BUX, BLY, BUY, NS, STAT, IFRQ1, FREQ2, ISW, WK, IERR)

Single precision:

CALL R2BA2S (X, N, Y, NCX, NCY, BLX, BUX, BLY, BUY, NS, STAT, IFRQ1, FREQ2, ISW, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | X | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | N | Input | Sequence of observed values for sample X $\{x_i\}$ |
| 2 | N | I | 1 | Input | Number of pairs of observed values $n$ |
| 3 | Y | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | N | Input | Sequence of observed values for sample Y $\{y_i\}$ |
| 4 | NCX | I | 1 | Input | Number of classes for sample X $m_x + 2$ |
| 5 | NCY | I | 1 | Input | Number of classes for sample Y $m_y + 2$ |
| 6 | BLX | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Value of $c_{min}$ |
| 7 | BUX | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Value of $c_{max}$ |
| 8 | BLY | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Value of $d_{min}$ |
| 9 | BUY | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Value of $d_{max}$ |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 10 | NS | I | 1 | Input | Number of pairs of observed values before adding observed values (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Number of pairs of observed values $n$ |
| 11 | STAT | $\left\{\begin{matrix}\mathrm{D}\\\mathrm{R}\end{matrix}\right\}$ | 5,2 | Output | Basic statistics of observed values (See Note (b)) |
| 12 | IFRQ1 | I | NCY, NCX | Output | Frequency distribution table $\{e_{ij}\}$ |
| 13 | FREQ2 | $\left\{\begin{matrix}\mathrm{D}\\\mathrm{R}\end{matrix}\right\}$ | NCY, NCX | Output | Frequency distribution table of percentages $\{f_{ij}\}$ |
| 14 | ISW | I | 1 | Input | Processing switch<br>0: Perform calculations using an unbiased estimate (no previously established basic statistics)<br>1: Perform calculations using an unbiased estimate (added observed values)<br>2: Perform calculations using a sample variance (no previously established basic statistics)<br>3: Perform calculations using a sample variance (added observed values) |
| 15 | WK | $\left\{\begin{matrix}\mathrm{D}\\\mathrm{R}\end{matrix}\right\}$ | See Contents | Work | Work area<br>**Size**: NCX + NCY + 2 × N |
| 16 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW = 0, 1, 2, 3

(b) N $\geq$ 1

(c) NCX $\geq$ 3

(d) NCY $\geq$ 3

(e) BUX > BLX

(f) BUY > BLY

(g) NS $\geq$ 1 (When ISW=1 or 3)

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (a) was not satisfied. | Processing continues with ISW=0. |
| 1010 | An attempt was made to obtain the standard deviation when ISW=0 and N=1. | The absolute value maximum that can be represented is set for the standard deviation. |
| 1020 | BUX<BLX occurred. | BUX and BLX are switched, and processing continues. |
| 1030 | BUY<BLY occurred. | BUY and BLY are switched, and processing continues. |
| 3000 | Restriction (b), (c), (d) was not satisfied. | Processing is aborted. |
| 3010 | BUX = BLX or BUY = BLY occurred. | |
| 3020 | Restriction (g) was not satisfied. | |

(6) **Notes**

(a) The frequencies of observed values $\{e_{ij}\}$ and frequency percentages $\{f_{ij}\}$ are stored in array IFRQ1 and FREQ2, respectively as real matrix (two-dimensional array type) data (See Appendix A).

(b) The basic statistics are stored as follows in the array STAT.

STAT(1, 1): Minimum for sample X
STAT(2, 1): Maximum for sample X
STAT(3, 1): Sum for sample X $S_x$
STAT(4, 1): Mean for sample X $\bar{x}$
STAT(5, 1): Standard deviation for sample X $SD_x$

STAT(1, 2): Minimum for sample Y
STAT(2, 2): Maximum for sample Y
STAT(3, 2): Sum for sample Y $S_y$
STAT(4, 2): Mean for sample Y $\bar{y}$
STAT(5, 2): Standard deviation for sample Y $SD_y$

(c) To obtain the basic statistics, frequencies, and frequency percentages when observed values are added, use the contents of NCX, NCY, BLX, BUX, BLY, BUY, NS, STAT, IFRQ1, FREQ2 and WK, which were calculated before adding the observed values, set the added pairs of observed values for X and Y and the number of added pairs of observed values for N, set ISW to 1 or 3, and perform the calculation. However, when obtaining the standard deviation, you must set the ISW value so that the calculation is performed using a sample variance following a calculation that used a sample variance the previous time or so that the calculation is performed using an unbiased estimate following a calculation that used an unbiased estimate the previous time.

(d) When there are an extremely large number of data values that are widely dispersed, better results are obtained by grouping them into data having absolute values of the same relative size and adding them to the samples in increasing order of size.

(e) Statistics obtained when calculations are performed using an unbiased estimate can be applied to a population for which sampling with replacement is performed from an infinite or finite population. Statistics obtained when calculations are performed using a sample variance can be applied to a population for which the population and sample match.

(7) **Example**

(a) Problem

Obtain the minimum, maximum, sum, mean, standard deviation, frequencies, and frequency percentages of the following two-sample observation data.

$$\{x_i\} = \{1, 3, 5, 7, 9, 11, 13\}$$
$$\{y_i\} = \{2, 4, 6, 8, 10, 12, 14\}$$

Also, obtain the minimum, maximum, sum, mean, standard deviation, frequencies, and frequency percentages when the following two-sample observation data is added.

$$\{x'_i\} = \{15, 17, 19, 21\}$$
$$\{y'_i\} = \{16, 18, 20, 22\}$$

(b) Input data

First processing :

Two-sample observation data $\{x_i\}$, $\{y_i\}$,

N = 7, NCX = 8, NCY = 7, BLX = 5.5, BLY = 2.5,

BUX = 11.5, BUY = 17.5 and ISW = 0.

Second processing :

Two-sample observation data $\{x'_i\}$, $\{y'_i\}$,

N = 4, NCX = 8, NCY = 7, BLX = 5.5, BLY = 2.5,

BUX = 11.5, BUY = 17.5 and ISW = 1.

(c) Main program

```
      PROGRAM B2BA2S
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 100, NCX = 8, NCY = 7 )
      DIMENSION IFRQ1(NCY,NCX)
      DIMENSION X(NA),Y(NA),STAT(5,2),FREQ2(NCY,NCX),WK(NCX+NCY+2*NA)
!
      WRITE(6,6000)
      IERR =     0
      ISW  =     0
      READ(5,*) N
      READ(5,*) BLX
      READ(5,*) BUX
      READ(5,*) BLY
      READ(5,*) BUY
      DO 100 I=1,N
         READ(5,*) X(I),Y(I)
  100 CONTINUE
      WRITE(6,6010) ISW,N,NCX,NCY,BLX,BLY,BUX,BUY
      WRITE(6,6020) 'X',(X(I),I=1,N)
      WRITE(6,6020) 'Y',(Y(I),I=1,N)
      CALL D2BA2S(X,N,Y,NCX,NCY,BLX,BUX,BLY,BUY,NS,&
      STAT,IFRQ1,FREQ2,ISW,WK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS
      WRITE(6,6050) 'X',(STAT(I,1),I=1,5)
      WRITE(6,6050) 'Y',(STAT(I,2),I=1,5)
      WRITE(6,6060)
      DO 110 I=1,NCY
         WRITE(6,6070) (IFRQ1(I,J),J=1,NCX)
  110 CONTINUE
      WRITE(6,6080)
      DO 120 I=1,NCY
         WRITE(6,6090) (FREQ2(I,J),J=1,NCX)
  120 CONTINUE
!
      WRITE(6,6100)
      IERR =     0
      ISW  =     1
```

```
      READ(5,*) N
      DO 130 I=1,N
         READ(5,*) X(I),Y(I)
  130 CONTINUE
      WRITE(6,6010) ISW,N,NCX,NCY,BLX,BLY,BUX,BUY
      WRITE(6,6020) 'X',(X(I),I=1,N)
      WRITE(6,6020) 'Y',(Y(I),I=1,N)
      CALL D2BA2S(X,N,Y,NCX,NCY,BLX,BUX,BLY,BUY,NS,&
      STAT,IFRQ1,FREQ2,ISW,WK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS
      WRITE(6,6050) 'X',(STAT(I,1),I=1,5)
      WRITE(6,6050) 'Y',(STAT(I,2),I=1,5)
      WRITE(6,6060)
      DO 140 I=1,NCY
         WRITE(6,6070) (IFRQ1(I,J),J=1,NCX)
  140 CONTINUE
      WRITE(6,6080)
      DO 150 I=1,NCY
         WRITE(6,6090) (FREQ2(I,J),J=1,NCX)
  150 CONTINUE
!
      STOP
 6000 FORMAT(  ' *** D2BA2S ***',/,&
              /,3X,'*** FIRST PROCESSING ***',/,&
              /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'ISW = ',I6,/,&
              /,7X,'N = ',I6,/,&
              /,7X,'NCX = ',5X,I6,5X,'NCY = ',5X,I6,/,&
              /,7X,'BLX = ',F11.2,5X,'BLY = ',F11.2,/,&
              /,7X,'BUX = ',F11.2,5X,'BUY = ',F11.2)
 6020 FORMAT( /,7X,'OBSERVATIONS ',A,/,/,&
                   5(6X,5(2X,F11.2),/))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
              /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'TOTAL SAMPLE SIZE = ',I6,/,&
              /,66X,'STANDARD',/,&
               8X,'OBSERVATION',3X,'VMIN',8X,'VMAX',8X,'SUM',8X,&
                  'VMEAN',4X,'DEVIATION',/,&
               7X,69('-'))
 6050 FORMAT(  13X,A,4(1X,F11.2),2X,D11.4)
 6060 FORMAT( /,7X,'MATRIX OF FREQUENCIES',/)
 6070 FORMAT(  9X,10I6)
 6080 FORMAT( /,7X,'MATRIX OF PERCENT FREQUENCIES',/)
 6090 FORMAT(  9X,10F6.2)
 6100 FORMAT(/,/,3X,'*** CONTINUATION PROCESSING ***',/,&
              /,3X,'** INPUT **')
      END
```

(d) Output results

```
 *** D2BA2S ***

   *** FIRST PROCESSING ***

   ** INPUT **

       ISW =      0

       N =      7

       NCX =            8     NCY =            7

       BLX =         5.50     BLY =         2.50

       BUX =        11.50     BUY =        17.50

       OBSERVATIONS X

               1.00          3.00          5.00          7.00          9.00
              11.00         13.00

       OBSERVATIONS Y

               2.00          4.00          6.00          8.00         10.00
              12.00         14.00

   ** OUTPUT **

       IERR =      0

       TOTAL SAMPLE SIZE =      7

                                                                  STANDARD
        OBSERVATION   VMIN        VMAX        SUM        VMEAN    DEVIATION
       ---------------------------------------------------------------------
             X        1.00       13.00       49.00        7.00  0.4320D+01
             Y        2.00       14.00       56.00        8.00  0.4320D+01

       MATRIX OF FREQUENCIES

              1     0     0     0     0     0     0     0
              1     0     0     0     0     0     0     0
              1     0     1     0     0     0     0     0
```

```
          0     0     0     0     1     0     0     0
          0     0     0     0     0     0     1     1
          0     0     0     0     0     0     0     0
          0     0     0     0     0     0     0     0

   MATRIX OF PERCENT FREQUENCIES

      14.29  0.00  0.00  0.00  0.00  0.00  0.00  0.00
      14.29  0.00  0.00  0.00  0.00  0.00  0.00  0.00
      14.29  0.00 14.29  0.00  0.00  0.00  0.00  0.00
       0.00  0.00  0.00  0.00 14.29  0.00  0.00  0.00
       0.00  0.00  0.00  0.00  0.00  0.00 14.29 14.29
       0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
       0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00


*** CONTINUATION PROCESSING ***

** INPUT **

   ISW =      1

   N =      4

   NCX =           8     NCY =           7

   BLX =        5.50     BLY =        2.50

   BUX =       11.50     BUY =       17.50

   OBSERVATIONS X

          15.00        17.00        19.00        21.00

   OBSERVATIONS Y

          16.00        18.00        20.00        22.00

** OUTPUT **

   IERR =      0

   TOTAL SAMPLE SIZE =     11

                                                                STANDARD
    OBSERVATION   VMIN        VMAX        SUM         VMEAN    DEVIATION
   -------------------------------------------------------------------------
         X        1.00       21.00      121.00        11.00  0.6633D+01
         Y        2.00       22.00      132.00        12.00  0.6633D+01

   MATRIX OF FREQUENCIES

          1     0     0     0     0     0     0     0
          1     0     0     0     0     0     0     0
          1     0     1     0     0     0     0     0
          0     0     0     0     1     0     0     0
          0     0     0     0     0     0     1     1
          0     0     0     0     0     0     0     1
          0     0     0     0     0     0     0     3

   MATRIX OF PERCENT FREQUENCIES

       9.09  0.00  0.00  0.00  0.00  0.00  0.00  0.00
       9.09  0.00  0.00  0.00  0.00  0.00  0.00  0.00
       9.09  0.00  9.09  0.00  0.00  0.00  0.00  0.00
       0.00  0.00  0.00  0.00  9.09  0.00  0.00  0.00
       0.00  0.00  0.00  0.00  0.00  0.00  9.09  9.09
       0.00  0.00  0.00  0.00  0.00  0.00  0.00  9.09
       0.00  0.00  0.00  0.00  0.00  0.00  0.00 27.27
```

### 4.2.3 D2BAMS, R2BAMS $m$ Samples Basic Statistics

(1) **Function**

Given $m$ samples consisting of n observed values $\{x_{ij}\}$, $(i = 1, \cdots, n;\ j = 1, \cdots, m)$, the D2BAMS or R2BAMS obtains the basic statistics (sum, mean, sum of squares of deviation, variance, and standard deviation) for each sample. It also obtains the basic statistics when $n$ observed values $\{y_{ij}\}$, $(i = 1, \cdots, n;\ j = 1, \cdots, m)$ are added to each of the $m$ samples for which the basic statistics are known.

The basic statistics for $m$ samples consisting of n observed values $\{x_{ij}\}$, $(i = 1, \cdots, n;\ j = 1, \cdots, m)$ are defined by the following equations.

Sum :

$$t_j = \sum_{i=1}^{n} x_{ij}, \quad (j = 1, \cdots, m)$$

Mean :

$$\bar{x}_j = \frac{t_j}{n}, \quad (j = 1, \cdots, m)$$

Sum of squares of deviation :

$$s_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2, \quad (j = 1, \cdots, m)$$

Variance :

$$v_j = \frac{s_j}{\alpha}, \quad (j = 1, \cdots, m)$$

Standard deviation :

$$d_j = \sqrt{v_j}, \quad (j = 1, \cdots, m)$$

Here, $\alpha$ is $n - 1$ when an unbiased estimate is used or $\alpha$ is $n$ when a sample variance is used.

(2) **Usage**

Double precision:

CALL D2BAMS (A, NA, N, M, NS, STAT, ISW, IERR)

Single precision:

CALL R2BAMS (A, NA, N, M, NS, STAT, ISW, IERR)

## (3) Arguments

D:Double precision real　Z:Double precision complex　I: { INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer }
R:Single precision real　C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | {D / R} | NA,M | Input | Matrix in which observed values are stored ($x_{ij}$) or ($y_{ij}$) |
| 2 | NA | I | 1 | Input | Adjustable dimension of array A |
| 3 | N | I | 1 | Input | Number of observed values per sample stored in array A $n$ |
| 4 | M | I | 1 | Input | Number of samples $m$ |
| 5 | NS | I | 1 | Input | Number of observed values per sample before adding observed values (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Number of observed values per sample $n$ |
| 6 | STAT | {D / R} | M,5 | Input | Basic statistics before adding observed values (See Note (a)) (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Obtained basic statistics (See Note (a)) |
| 7 | ISW | I | 1 | Input | Processing switch<br>0: Perform calculations using an unbiased estimate (no previously established basic statistics)<br>1: Perform calculations using an unbiased estimate (added observed values)<br>2: Perform calculations using a sample variance (no previously established basic statistics)<br>3: Perform calculations using a sample variance (added observed values) |
| 8 | IERR | I | 1 | Output | Error indicator |

## (4) Restrictions

(a) ISW=0, 1, 2, 3

(b) NA≥N≥1

(c) M≥1

(d) NS≥1 (When ISW=1 or 3)

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (a) was not satisfied. | Processing continues with ISW=0. |
| 1010 | An attempt was made to obtain an unbiased estimate when N=1. | The absolute value maximum that can be represented is set for the variance and standard deviation. |
| 3000 | Restriction (b) or (c) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (d) was not satisfied. | |

(6) **Notes**

(a) The basic statistics are stored as follows in the array STAT.

STAT($j$, 1) : Sum $t_j$
STAT($j$, 2) : Mean $\bar{x}_j$
STAT($j$, 3) : Sum of squares of deviation $s_j$ , $(j = 1, \cdots, \mathrm{M})$
STAT($j$, 4) : Variance $v_j$
STAT($j$, 5) : Standard deviation $d_j$

(b) To obtain the basic statistics when the same numbers of observed values are added for each sample, use the contents of STAT and NS, which were calculated before adding the observed values, set the added observed values for A and the number of added observed values for N, set ISW to 1 or 3, and perform the calculation. However, when obtaining the variance or standard deviation, you must set the ISW value so that the calculation is performed using a sample variance following a calculation that used a sample variance the previous time or so that the calculation is performed using an unbiased estimate following a calculation that used an unbiased estimate the previous time.

(c) When there are an extremely large number of data values that are widely dispersed, better results are obtained by grouping them into data having absolute values of the same relative size and adding them to the samples in increasing order of size.

(d) Statistics obtained when calculations are performed using an unbiased estimate can be applied to a population for which sampling with replacement is performed from an infinite or finite population. Statistics obtained when calculations are performed using a sample variance can be applied to a population for which the population and sample match.

(7) **Example**

(a) Problem

Obtain the basic statistics (sum, mean, sum of squares of deviation, variance, and standard deviation) for each sample when the observed values are given by matrix $X$ shown below.

$$X = \begin{bmatrix} 30 & 35 & 44 & 44 & 45 \\ 424 & 365 & 346 & 349 & 297 \\ 246 & 219 & 255 & 252 & 256 \end{bmatrix}$$

Also, obtain the basic statistics (sum, mean, sum of squares of deviation, variance, and standard deviation) for each sample when the observed values given by matrix $Y$ shown below are added.

$$Y = \begin{bmatrix} 18 & 21 & 56 & 21 & 45 \\ 2 & 2 & 3 & 1 & 3 \end{bmatrix}$$

(b) Input data

First processing :

Matrix $X$ in which observed values are stored, NA = 100, N = 3, M = 5 and ISW = 0.

Second processing :

Matrix $Y$ in which observed values are stored, NA = 100, N = 2, M = 5 and ISW = 1.

(c) Main program

```
      PROGRAM B2BAMS
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 100, M = 5 )
      DIMENSION A(NA,M),STAT(M,5)
!
      WRITE(6,6000)
      IERR =  0
      ISW  =  0
      READ(5,*) N
      DO 100 I=1,N
         READ(5,*) (A(I,J),J=1,M)
  100 CONTINUE
      WRITE(6,6010) ISW,N,M
      DO 110 I=1,N
         WRITE(6,6020) (A(I,J),J=1,M)
  110 CONTINUE
      CALL D2BAMS(A,NA,N,M,NS,STAT,ISW,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS
      DO 120 J=1,M
         WRITE(6,6050) J,(STAT(J,I),I=1,5)
  120 CONTINUE
!
      WRITE(6,6060)
      IERR =  0
      ISW  =  1
      READ(5,*) N
      DO 130 I=1,N
         READ(5,*) (A(I,J),J=1,M)
  130 CONTINUE
      WRITE(6,6010) ISW,N,M
      DO 140 I=1,N
         WRITE(6,6020) (A(I,J),J=1,M)
  140 CONTINUE
      CALL D2BAMS(A,NA,N,M,NS,STAT,ISW,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS
      DO 150 J=1,M
         WRITE(6,6050) J,(STAT(J,I),I=1,5)
  150 CONTINUE
!
      STOP
 6000 FORMAT(  ' *** D2BAMS ***',/,&
              /,3X,'*** FIRST PROCESSING ***',/,&
              /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'ISW = ',I6,/,&
              /,7X,'N = ',I6,5X,'M = ',I6,/,&
              /,7X,'OBSERVATION MATRIX',/)
 6020 FORMAT(  7X,5(2X,F11.2))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
              /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'TOTAL SAMPLE SIZE = ',I6,/,&
              /,45X,'SUM OF',16X,'STANDARD',/,&
               8X,'VARIABLE',8X,'SUM',8X,'MEAN',5X,'SQUARES',4X,&
                  'VARIANCE',3X,'DEVIATION',/,&
               7X,70('-'))
 6050 FORMAT(  7X,I6,3X,3(1X,F11.2),2(1X,D11.4))
 6060 FORMAT(/,/,3X,'*** CONTINUATION PROCESSING ***',/,&
              /,3X,'** INPUT **')
      END
```

(d) Output results

```
*** D2BAMS ***

  *** FIRST PROCESSING ***

  ** INPUT **

      ISW =      0

      N =      3     M =      5

      OBSERVATION MATRIX

               30.00        35.00        44.00        44.00        45.00
              424.00       365.00       346.00       349.00       297.00
              246.00       219.00       255.00       252.00       256.00
```

```
** OUTPUT **

    IERR =      0

    TOTAL SAMPLE SIZE =      3

                                          SUM OF                     STANDARD
     VARIABLE        SUM         MEAN     SQUARES     VARIANCE     DEVIATION
    ----------------------------------------------------------------------
          1        700.00       233.33    77858.67  0.3893D+05  0.1973D+03
          2        619.00       206.33    54690.67  0.2735D+05  0.1654D+03
          3        645.00       215.00    48002.00  0.2400D+05  0.1549D+03
          4        645.00       215.00    48566.00  0.2428D+05  0.1558D+03
          5        598.00       199.33    36568.67  0.1828D+05  0.1352D+03

*** CONTINUATION PROCESSING ***

** INPUT **

    ISW =      1

    N =      2     M =      5

    OBSERVATION MATRIX

            18.00         21.00         56.00         21.00         45.00
             2.00          2.00          3.00          1.00          3.00

** OUTPUT **

    IERR =      0

    TOTAL SAMPLE SIZE =      5

                                          SUM OF                     STANDARD
     VARIABLE        SUM         MEAN     SQUARES     VARIANCE     DEVIATION
    ----------------------------------------------------------------------
          1        720.00       144.00   137840.00  0.3446D+05  0.1856D+03
          2        642.00       128.40   100423.20  0.2511D+05  0.1584D+03
          3        704.00       140.80    90698.80  0.2267D+05  0.1506D+03
          4        667.00       133.40    98705.20  0.2468D+05  0.1571D+03
          5        646.00       129.20    74340.80  0.1859D+05  0.1363D+03
```

### 4.2.4 D2BAGM, R2BAGM<br>Geometric Mean

(1) **Function**

Given a sample consisting of $n$ observed values $\{x_i\}(i = 1, \cdots, n)$, The D2BAGM or R2BAGM obtains the geometric mean and its standard deviation. Also, the D2BAGM or R2BAGM obtains the geometric mean and its standard deviation when $n$ observed values $\{y_i\}(i = 1, \cdots, n)$ are added.

For a sample consisting of $n$ observed values $\{x_i\}(i = 1, \cdots, n)$, the geometric mean and its standard deviation are defined by the following equations.

Geometric mean :

$$\mathrm{GM} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$$

Standard deviation :

$$\mathrm{GSD} \quad = \quad \exp\left(\sqrt{\frac{\displaystyle\sum_{i=1}^{n}(\log x_i)^2 - n(\log \mathrm{GM})^2}{\alpha}}\right)$$

Here, $\alpha$ is $n-1$ when an unbiased estimate is used or $\alpha$ is $n$ when a sample variance is used.

(2) **Usage**

Double precision:

CALL D2BAGM (A, N, NS, GM, GSD, ISW, IERR)

Single precision:

CALL R2BAGM (A, N, NS, GM, GSD, ISW, IERR)

(3) **Arguments**

D:Double precision real   Z:Double precision complex   I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real   C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | N | Input | Sequence of observed values $\{x_i\}$ or $\{y_i\}$ |
| 2 | N | I | 1 | Input | Number of observed values $n$ |
| 3 | NS | I | 1 | Input | Number of observed values before adding observed values (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Number of observed values $n$ |
| 4 | GM | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Geometric mean before adding observed values (See Note (a)) (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Geometric mean that was obtained (See Note (a)) |
| 5 | GSD | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Standard deviation before adding observed values (See Note (a)) (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Standard deviation that was obtained (See Note (a)) |
| 6 | ISW | I | 1 | Input | Processing switch<br>0: Perform calculations using an unbiased estimate (no previously established basic statistics)<br>1: Perform calculations using an unbiased estimate (added observed values)<br>2: Perform calculations using a sample variance (no previously established basic statistics)<br>3: Perform calculations using a sample variance (added observed values) |
| 7 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW = 0, 1, 2, 3

(b) N $\geq$ 1

(c) NS $\geq$ 1 (When ISW=1 or 3)

(d) $\mathrm{A(i)} > 0.0(\mathrm{i} = 1, 2, \cdots, \mathrm{N})$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (a) was not satisfied. | Processing continues with ISW=0. |
| 1010 | An attempt was made to obtain an unbiased estimate when N = 1. | The absolute value maximum that can be represented is set for the standard deviation. |
| 3000 | Restriction (b) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (c) was not satisfied. | |
| 3020 | Restriction (d) was not satisfied. | |

(6) **Notes**

(a) To obtain the statistics when observed values are added, use the contents of GM, GSD and NS, which were calculated before adding the observed values, set the added observed values for A and the number of added observed values for N, set ISW to 1 or 3, and perform the calculation.

(b) When there are an extremely large number of data values that are widely dispersed, better results are obtained by grouping them into data having absolute values of the same relative size and adding them to the samples in increasing order of size.

(c) Statistics obtained when calculations are performed using an unbiased estimate can be applied to a population for which sampling with replacement is performed from an infinite or finite population. Statistics obtained when calculations are performed using a sample variance can be applied to a population for which the population and sample match.

(7) **Example**

(a) Problem
Given the following sequence of observed values, obtain the geometric mean and its standard deviation.

$$\begin{aligned}\{x_i\} &= \{1100, 2630, 695, 3630, 1550, 1010, 2110, 736, 1260, 1690,\\ &\quad 2680, 2520, 2030, 1280, 2400\}\end{aligned}$$

Also, obtain the geometric mean and its standard deviation when the following sequence of observed values are added.

$$\{y_i\} = \{938, 1860, 2410, 3370, 1380, 2200, 2290, 1220, 1150\}$$

(b) Input data
First processing :

Sequence of observed values $\{x_i\}$, N = 15 and ISW = 0.

Second processing :

Sequence of observed values $\{y_i\}$, N = 9 and ISW = 1.

(c) Main program

```
      PROGRAM B2BAGM
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 100 )
      DIMENSION A(NA)
!
      WRITE(6,6000)
```

```
      IERR =  0
      ISW  =  0
      READ(5,*) N
      DO 100 I=1,N
         READ(5,*) A(I)
  100 CONTINUE
      WRITE(6,6010) ISW,N
      WRITE(6,6020) (A(I),I=1,N)
      CALL D2BAGM(A,N,NS,GM,GSD,ISW,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS,GM,GSD
!
      WRITE(6,6050)
      IERR =  0
      ISW  =  1
      READ(5,*) N
      DO 110 I=1,N
         READ(5,*) A(I)
  110 CONTINUE
      WRITE(6,6010) ISW,N
      WRITE(6,6020) (A(I),I=1,N)
      CALL D2BAGM(A,N,NS,GM,GSD,ISW,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS,GM,GSD
!
      STOP
 6000 FORMAT(  ' *** D2BAGM ***',/,&
              /,3X,'*** FIRST PROCESSING ***',/,&
              /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'ISW = ',I6,/,&
              /,7X,'N = ',I6)
 6020 FORMAT( /,7X,'OBSERVATIONS',/,/,&
                   3(6X,5(2X,F11.2),/))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
              /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'TOTAL SAMPLE SIZE = ',5X,I6,/,&
              /,7X,'GM                = ',D15.8,/,&
              /,7X,'GSD               = ',D15.8)
 6050 FORMAT(/,/,3X,'*** CONTINUATION PROCESSING ***',/,&
              /,3X,'** INPUT **')
      END
```

(d) Output results

```
 *** D2BAGM ***

   *** FIRST PROCESSING ***

   ** INPUT **

       ISW =      0

       N =     15

       OBSERVATIONS

            1100.00      2630.00       695.00      3630.00      1550.00
            1010.00      2110.00       736.00      1260.00      1690.00
            2680.00      2520.00      2030.00      1280.00      2400.00


   ** OUTPUT **

       IERR =      0

       TOTAL SAMPLE SIZE =          15

       GM                =  0.16347482D+04

       GSD               =  0.16419280D+01


   *** CONTINUATION PROCESSING ***

   ** INPUT **

       ISW =      1

       N =      9

       OBSERVATIONS

             938.00      1860.00      2410.00      3370.00      1380.00
            2200.00      2290.00      1220.00      1150.00

   ** OUTPUT **

       IERR =      0

       TOTAL SAMPLE SIZE =          24

       GM                =  0.16695395D+04

       GSD               =  0.15843787D+01
```

### 4.2.5 D2BAMO, R2BAMO Moment

(1) **Function**

Assume that we are given $m$ classes of equal width $C_i = [x_i, x_i + h)$ $(i = 1, 2, \cdots, m)$, and that the following relationships are satisfied.

$$\begin{aligned} x_1 &= x_{min} \\ x_{i+1} &= x_i + h \qquad i = 1, 2, \cdots, m \\ x_{max} &= x_m + h \end{aligned}$$

Here, $h$ is the class width, which is defined as follows.

$$h = \frac{x_{max} - x_{min}}{m}$$

The D2BAMO or R2BAMO obtains the linear moment about the origin and the moment of order $r$ about the mean value, which are defined as follows, where the frequency of observation data in class $C_i$ is $f_i$.
Linear moment about the origin :

$$\mu'_1 = \frac{\sum_{i=1}^{m} f_i \hat{x}_i}{\sum_{i=1}^{m} f_i}$$

Moment of order $r$ about the mean value :

$$\mu_r = \frac{\sum_{i=1}^{m} [f_i(\hat{x}_i - \mu'_1)^r]}{\sum_{i=1}^{m} f_i} \qquad (r = 2, 3, \cdots)$$

$\hat{x}_i$, which represents the class value of each class, is defined as follows.

$$\hat{x}_i = x_{min} + (i - 0.5)h$$

(2) **Usage**

Double precision:
CALL D2BAMO (F, M, XMAX, XMIN, NP, XM, WK, IERR)
Single precision:
CALL R2BAMO (F, M, XMAX, XMIN, NP, XM, WK, IERR)

## (3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{ \begin{array}{l} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{array} \right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | F | $\left\{\begin{matrix}D\\R\end{matrix}\right\}$ | M | Input | Frequency $f_i$. |
| 2 | M | I | 1 | Input | Number of classes $m$ |
| 3 | XMAX | $\left\{\begin{matrix}D\\R\end{matrix}\right\}$ | 1 | Input | Upper limit of largest class $x_{max}$ |
| 4 | XMIN | $\left\{\begin{matrix}D\\R\end{matrix}\right\}$ | 1 | Input | Lower limit of smallest class $x_{min}$ |
| 5 | NP | I | 1 | Input | Order of moment to be obtained $r$<br>When NP=1, the linear moment about the origin is obtained. |
| 6 | XM | $\left\{\begin{matrix}D\\R\end{matrix}\right\}$ | 1 | Output | Moment ${\mu'}_1$ or $\mu_r$ $(r = 2, \cdots, m)$ |
| 7 | WK | $\left\{\begin{matrix}D\\R\end{matrix}\right\}$ | M | Work | Work area |
| 8 | IERR | I | 1 | Output | Error indicator |

## (4) **Restrictions**

(a) XMAX $>$ XMIN

(b) M $\geq 1$

(c) NP $\geq 1$

(d) F(i) $\geq 0.0$  $(\mathrm{i} = 1, 2, \cdots, \mathrm{M})$

(e) $\sum_{\mathrm{i}=1}^{\mathrm{M}} \mathrm{F(i)} > 0.0$

## (5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | XMAX $<$ XMIN | XMAX and XMIN are switched, and processing continues. |
| 3000 | XMAX $=$ XMIN is specified. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 3040 | Restriction (e) was not satisfied. | |

(6) **Notes**

(a) According to the definition, the linear moment about the origin is the mean, and the moment of order 2 about the mean is the variance.

(b) The **skewness** is an indicator representing the degree of asymmetry in the distribution of observed values. The value $\alpha_3 = \frac{\mu_3}{\sqrt{{\mu_2}^3}}$, which is defined using the moments of order 2 and 3 about the mean $(\mu_2, \mu_3)$ or its square is used as the skewness.

(c) The **kurtosis** is an indicator representing the degree of peakedness in the distribution of observed values. The value $\alpha_4 = \frac{\mu_4}{{\mu_2}^2}$, which is defined using the moments of order 2 and 4 about the mean $(\mu_2, \mu_4)$ or $\alpha_4 - 3$ is used as the kurtosis.

(7) **Example**

(a) Problem

Assume that the interval $[0, 80]$ is divided into 20 equal-width class intervals and that the observation frequencies of each class $F_1, F_2, \cdots, F_{20}$ are given as follows. Obtain the moment of order 4.

| Class | $F_i$ | Class | $F_i$ | Class | $F_i$ | Class | $F_i$ |
|---|---|---|---|---|---|---|---|
| 1 | 527 | 6 | 3942 | 11 | 1496 | 16 | 448 |
| 2 | 501 | 7 | 3737 | 12 | 1044 | 17 | 283 |
| 3 | 1082 | 8 | 3012 | 13 | 874 | 18 | 260 |
| 4 | 2177 | 9 | 2489 | 14 | 607 | 19 | 207 |
| 5 | 2958 | 10 | 1801 | 15 | 450 | 20 | 144 |

(b) Input data

Observation frequencies $\{F_i\}$, M = 20, XMAX = 80.0, XMIN = 0.0 and NP = 20.

(c) Main program

```
      PROGRAM B2BAMO
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NF = 100 )
      DIMENSION F(NF),WK(NF)
!
      WRITE(6,6000)
      IERR =     0
      READ(5,*) M
      READ(5,*) XMAX
      READ(5,*) XMIN
      READ(5,*) NP
      DO 100 I=1,M
         READ(5,*) F(I)
  100 CONTINUE
      WRITE(6,6010) M,XMAX,XMIN,NP
      WRITE(6,6020) (F(I),I=1,M)
      CALL D2BAMO(F,M,XMAX,XMIN,NP,XM,WK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) XM
!
      STOP
 6000 FORMAT(  ' *** D2BAMO ***',/,&
              /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'M  = ',I6,/,&
              /,7X,'UPPER LIMIT = ',F11.2,5X,'LOWER LIMIT = ',F11.2,/,&
              /,7X,'ORDER OF MOMENT = ',I6)
 6020 FORMAT( /,7X,'FREQUENCIES',/,/,&
                   4(6X,5(2X,F11.2),/))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
              /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'MOMENT = ',F11.2)
      END
```

(d) Output results

```
*** D2BAMO ***

  ** INPUT **

      M  =     20
```

```
    UPPER LIMIT =       80.00     LOWER LIMIT =        0.00

    ORDER OF MOMENT =      4

    FREQUENCIES

          527.00       501.00      1082.00      2177.00      2958.00
         3942.00      3737.00      3012.00      2489.00      1801.00
         1496.00      1044.00       874.00       607.00       450.00
          448.00       283.00       260.00       207.00       144.00

** OUTPUT **

    IERR =      0

    MOMENT =   173638.61
```

## 4.2.6 D2BAHM, R2BAHM Harmonic Mean

(1) **Function**

Given a sample consisting of $n$ observed values $\{x_i\}(i = 1, \cdots, n)$, The D2BAHM or R2BAHM obtains the harmonic mean. Also, the D2BAHM or R2BAHM obtains the harmonic mean when $n$ observed values $\{y_i\}(i = 1, \cdots, n)$ are added.

For a sample consisting of $n$ observed values $\{x_i\}(i = 1, \cdots, n)$, the harmonic mean is defined by the following equations.

Harmonic mean :

$$\mathrm{HM} = \frac{1}{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\frac{1}{x_i}}$$

(2) **Usage**

Double precision:

CALL D2BAHM (A, N, NS, HM, ISW, IERR)

Single precision:

CALL R2BAHM (A, N, NS, HM, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{matrix}\text{D}\\ \text{R}\end{matrix}\right\}$ | N | Input | Sequence of observed values $\{x_i\}$ or $\{y_i\}$ |
| 2 | N | I | 1 | Input | Number of observed values $n$ |
| 3 | NS | I | 1 | Input | Number of observed values before adding observed values (for ISW=0 , no initial setting is required) |
| | | | | Output | Number of observed values $n$ |
| 4 | HM | $\left\{\begin{matrix}\text{D}\\ \text{R}\end{matrix}\right\}$ | 1 | Input | Harmonic mean before adding observed values (See Note (a)) (for ISW=0, no initial setting is required) |
| | | | | Output | Harmonic mean that was obtained (See Note (a)) |
| 5 | ISW | I | 1 | Input | Processing switch<br>0: No previously established basic statistics<br>1: Added observed values |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW = 0, 1

(b) $N \geq 1$

(c) $NS \geq 1$ (When ISW=1)

(d) $A(i) > 0.0 (i = 1, 2, \cdots, N)$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (a) was not satisfied. | Processing continues with ISW=0. |
| 3000 | Restriction (b) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (c) was not satisfied. | |
| 3020 | Restriction (d) was not satisfied. | |
| 4000 | Denominator of harmonic mean was zero. | The absolute value maximum that can be represented is set for the standard deviation. |

(6) **Notes**

(a) To obtain the statistics when observed values are added, use the contents of HM and NS, which were calculated before adding the observed values, set the added observed values for A and the number of added observed values for N, set ISW to 1, and perform the calculation.

(b) When there are an extremely large number of data values that are widely dispersed, better results are obtained by grouping them into data having absolute values of the same relative size and adding them to the samples in increasing order of size.

(c) Statistics obtained when calculations are performed using an unbiased estimate can be applied to a population for which sampling with replacement is performed from an infinite or finite population. Statistics obtained when calculations are performed using a sample variance can be applied to a population for which the population and sample match.

(7) **Example**

(a) Problem

Given the following sequence of observed values, obtain the harmonic mean and its standard deviation.

$$\{x_i\} = \{300, 600, 150, 30, 20, 120, 200, 100, 50, 40, 50\}$$

Also, obtain the harmonic mean and its standard deviation when the following sequence of observed values are added.

$$\{y_i\} = \{120, 1200, 300, 150, 600, -120, 240, 200, -100, -50\}$$

(b) Input data

First processing :

Sequence of observed values $\{x_i\}$, N = 11 and ISW = 0.

Second processing :

Sequence of observed values $\{y_i\}$, N = 10 and ISW = 1.

(c) Main program

```
      PROGRAM B2BAHM
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 100 )
      DIMENSION A(NA)
!
      WRITE(6,6000)
      IERR =  0
      ISW  =  0
      READ(5,*) N
      DO 100 I=1,N
         READ(5,*) A(I)
  100 CONTINUE
      WRITE(6,6010) ISW,N
      WRITE(6,6020) (A(I),I=1,N)
      CALL D2BAHM(A,N,NS,HM,ISW,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS,HM
!
      WRITE(6,6050)
      IERR =  0
      ISW  =  1
      READ(5,*) N
      DO 110 I=1,N
         READ(5,*) A(I)
  110 CONTINUE
      WRITE(6,6010) ISW,N
      WRITE(6,6020) (A(I),I=1,N)
      CALL D2BAHM(A,N,NS,HM,ISW,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS,HM
!
      STOP
 6000 FORMAT(  ' *** D2BAHM ***',/,&
             /,3X,'*** FIRST PROCESSING ***',/,&
             /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'ISW = ',I6,/,&
             /,7X,'N = ',I6)
 6020 FORMAT( /,7X,'OBSERVATIONS',/,/,&
                  3(6X,5(2X,F11.2),/))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
             /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'TOTAL SAMPLE SIZE = ',5X,I6,/,&
             /,7X,'HM                = ',D15.8)
 6050 FORMAT(/,/,3X,'*** CONTINUATION PROCESSING ***',/,&
             /,3X,'** INPUT **')
      END
```

(d) Output results

```
*** D2BAHM ***

  *** FIRST PROCESSING ***

  ** INPUT **

      ISW =      0

      N =     11

      OBSERVATIONS

            300.00       600.00       150.00        30.00        20.00
            120.00       200.00       100.00        50.00        40.00
             50.00

  ** OUTPUT **

      IERR =      0

      TOTAL SAMPLE SIZE =          11

      HM                =  0.60000000D+02


  *** CONTINUATION PROCESSING ***

  ** INPUT **

      ISW =      1

      N =     10

      OBSERVATIONS

            120.00      1200.00       300.00       150.00       600.00
           -120.00       240.00       200.00      -100.00       -50.00


  ** OUTPUT **

      IERR =      0
```

```
TOTAL SAMPLE SIZE =          21
HM                =  0.12000000D+03
```

### 4.2.7 D2BASM, R2BASM Root Mean Square

(1) **Function**

Given a sample consisting of $n$ observed values $\{x_i\}(i = 1, \cdots, n)$, The D2BASM or R2BASM obtains the root mean square Also, the D2BAHM or R2BAHM obtains the root mean square when $n$ observed values $\{y_i\}(i = 1, \cdots, n)$ are added.

For a sample consisting of $n$ observed values $\{x_i\}(i = 1, \cdots, n)$, the root mean square is defined by the following equations.

Root mean square :

$$\mathrm{SM} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$$

(2) **Usage**

Double precision:
  CALL D2BASM (A, N, NS, SM, ISW, IERR)
Single precision:
  CALL R2BASM (A, N, NS, SM, ISW, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\\text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | N | Input | Sequence of observed values $\{x_i\}$ or $\{y_i\}$ |
| 2 | N | I | 1 | Input | Number of observed values $n$ |
| 3 | NS | I | 1 | Input | Number of observed values before adding observed values (for ISW=0, no initial setting is required) |
| | | | | Output | Number of observed values $n$ |
| 4 | SM | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 1 | Input | Root mean square before adding observed values (See Note (a)) (for ISW=0, no initial setting is required) |
| | | | | Output | Root mean square that was obtained (See Note (a)) |
| 5 | ISW | I | 1 | Input | Processing switch<br>0: No previously established basic statistics<br>1: Added observed values |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW = 0, 1

(b) N $\geq$ 1

(c) NS $\geq$ 1 (When ISW=1 )

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (a) was not satisfied. | Processing continues with ISW=0. |
| 3000 | Restriction (b) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (c) was not satisfied. | |

(6) **Notes**

(a) To obtain the statistics when observed values are added, use the contents of SM and NS, which were calculated before adding the observed values, set the added observed values for A and the number of added observed values for N, set ISW to 1, and perform the calculation.

(b) When there are an extremely large number of data values that are widely dispersed, better results are obtained by grouping them into data having absolute values of the same relative size and adding them to the samples in increasing order of size.

(c) Statistics obtained when calculations are performed using an unbiased estimate can be applied to a population for which sampling with replacement is performed from an infinite or finite population. Statistics obtained when calculations are performed using a sample variance can be applied to a population for which the population and sample match.

(7) **Example**

(a) Problem
Given the following sequence of observed values, obtain the root mean square.

$$\{x_i\} \quad = \quad \{90, 60, 30, 95, 40, 80, 25, 50, 70, 50\}$$

Also, obtain the root mean square when the following sequence of observed values are added.

$$\{y_i\} = \{60, 60, 60, 70, 60, 60, 50, 60, 60, 60\}$$

(b) Input data
First processing :

Sequence of observed values $\{x_i\}$, N = 11 and ISW = 0.

Second processing :

Sequence of observed values $\{y_i\}$, N = 10 and ISW = 1.

(c) Main program

```
      PROGRAM B2BASM
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 100 )
      DIMENSION A(NA)
!
      WRITE(6,6000)
      IERR =  0
      ISW  =  0
      READ(5,*) N
      DO 100 I=1,N
         READ(5,*) A(I)
  100 CONTINUE
      WRITE(6,6010) ISW,N
      WRITE(6,6020) (A(I),I=1,N)
      CALL D2BASM(A,N,NS,SM,ISW,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS,SM
!
      WRITE(6,6050)
      IERR =  0
      ISW  =  1
      READ(5,*) N
      DO 110 I=1,N
         READ(5,*) A(I)
  110 CONTINUE
      WRITE(6,6010) ISW,N
      WRITE(6,6020) (A(I),I=1,N)
      CALL D2BASM(A,N,NS,SM,ISW,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS,SM
!
      STOP
 6000 FORMAT(  ' *** D2BASM ***',/,&
             /,3X,'*** FIRST PROCESSING ***',/,&
             /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'ISW = ',I6,/,&
             /,7X,'N = ',I6)
 6020 FORMAT( /,7X,'OBSERVATIONS',/,/,&
                  3(6X,5(2X,F11.2),/))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
             /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'TOTAL SAMPLE SIZE = ',5X,I6,/,&
             /,7X,'SM                = ',D15.8)
 6050 FORMAT(/,/,3X,'*** CONTINUATION PROCESSING ***',/,&
             /,3X,'** INPUT **')
      END
```

(d) Output results

```
*** D2BASM ***

  *** FIRST PROCESSING ***

  ** INPUT **

      ISW =      0

      N =     11

      OBSERVATIONS

             90.00         60.00         30.00         95.00         40.00
             80.00         25.00         50.00         70.00         70.00
             50.00

  ** OUTPUT **

      IERR =      0

      TOTAL SAMPLE SIZE =           11

      SM                =  0.63995738D+02


  *** CONTINUATION PROCESSING ***

  ** INPUT **

      ISW =      1

      N =     10

      OBSERVATIONS

             60.00         60.00         60.00         70.00         60.00
             60.00         50.00         60.00         60.00         60.00


  ** OUTPUT **

      IERR =      0
```

```
TOTAL SAMPLE SIZE =          21
SM                =  0.62201669D+02
```

## 4.3 VARIANCE-COVARIANCE

### 4.3.1 D2VCMT, R2VCMT Variance-Covariance Matrices

(1) **Function**

Given $m$ samples consisting of n observed values $\{x_{ki}\}(k = 1, \cdots, n; i = 1, \cdots, m)$, the D2VCMT or R2VCMT obtains the mean of each sample and the variance-covariance among the samples. It also obtains the mean and variance-covariance when $n$ observed values $\{y_{ki}\}$ $(k = 1, \cdots, n; i = 1, \cdots, m)$ are added to each of the $m$ samples for which the mean and variance-covariance are known.

The mean and variance-covariance among the samples for $m$ samples consisting of $n$ observed values $\{x_{ki}\}$ $(k = 1, \cdots, n; i = 1, \cdots, m)$ are defined by the following equations.

Mean :

$$\bar{x}_i = \frac{\sum_{k=1}^{n} x_{ki}}{n} \qquad i = 1, \cdots, m$$

Variance-covariance :

$$d_{ij} = \frac{s_{ij}}{\alpha} \qquad i, j = 1, \cdots, m$$

$s_{ij}$ (sum of squares of deviation matrix) is defined as follows.

$$s_{ij} = \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \qquad i, j = 1, \cdots, m$$

Here, $\alpha$ is $n$ when a sample covariance is used or $\alpha$ is $n-1$ when an unbiased covariance is used. The diagonal elements of the variance-covariance matrix are the variance values.

(2) **Usage**

Double precision:

CALL D2VCMT (A, NA, N, M, NS, X1, D, ND, ISW, WK, IERR)

Single precision:

CALL R2VCMT (A, NA, N, M, NS, X1, D, ND, ISW, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{ \begin{array}{l} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{array} \right\}$

R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{ \begin{array}{c} \text{D} \\ \text{R} \end{array} \right\}$ | NA,M | Input | Matrix in which observed values are stored $(x_{ki})$ or $(y_{ki})$ (See Note (a)) |
| 2 | NA | I | 1 | Input | Adjustable dimension of array A |
| 3 | N | I | 1 | Input | Number of observed values per sample stored in array A $n$ |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 4 | M | I | 1 | Input | Number of samples $m$ |
| 5 | NS | I | 1 | Input | Number of observed values per sample before adding observed values (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Number of observed values per sample $n$ |
| 6 | X1 | $\begin{Bmatrix} \mathrm{D} \\ \mathrm{R} \end{Bmatrix}$ | M | Input | Mean of each sample before adding observed values (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Obtained mean of each sample |
| 7 | D | $\begin{Bmatrix} \mathrm{D} \\ \mathrm{R} \end{Bmatrix}$ | ND,M | Input | Variance-covariance matrix before adding observed values (See Note (b)) (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Obtained variance-covariance matrix (See Note (b)) |
| 8 | ND | I | 1 | Input | Adjustable dimension of array D |
| 9 | ISW | I | 1 | Input | Processing switch<br>0: Perform calculations using an unbiased covariance (when the result of the previous calculation is not used)<br>1: Perform calculations using an unbiased covariance (when the result of the previous calculation is used)<br>2: Perform calculations using a sample covariance (when the result of the previous calculation is not used)<br>3: Perform calculations using a sample covariance (when the result of the previous calculation is used) |
| 10 | WK | $\begin{Bmatrix} \mathrm{D} \\ \mathrm{R} \end{Bmatrix}$ | M | Work | Work area |
| 11 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW = 0, 1, 2, 3

(b) NA $\geq$ N $\geq$ 1

(c) ND $\geq$ M $\geq$ 1

(d) NS $\geq$ 1 (When ISW=1 or 3)

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (a) was not satisfied. | Processing continues with ISW=0. |
| 1010 | An attempt was made to obtain an unbiased covariance when N = 1. | The absolute value maximum that can be represented is set for the covariance. |
| 3000 | Restriction (b) or (c) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (d) was not satisfied. | |

(6) **Notes**

(a) The observed values $x_{ki}$ or $y_{ki}$ $(k = 1, 2, \cdots, n; i = 1, 2, \cdots, m)$ are stored in array A as real matrix (two-dimensional array type) data (See Appendix A).

(b) To obtain the mean and covariance when the same numbers of observed values are added for each sample, use the contents of D and NS, which were calculated before adding the observed values, set the added observed values for A and the number of added observed values for N, set ISW to 1 or 3, and perform the calculation. However, when obtaining the covariance, you must set the ISW value so that the calculation is performed using a sample covariance following a calculation that used a sample covariance the previous time or so that the calculation is performed using an unbiased covariance following a calculation that used an unbiased covariance the previous time.

(c) When there are an extremely large number of data values that are widely dispersed, better results are obtained by grouping them into data having absolute values of the same relative size and adding them to the samples in increasing order of size.

(d) Statistics obtained when calculations are performed using an unbiased covariance can be applied to a population for which sampling with replacement is performed from an infinite or finite population. Statistics obtained when calculations are performed using a sample covariance can be applied to a population for which the population and sample match.

(7) **Example**

(a) Problem
Obtain the mean of each sample and the covariance among the samples when the observed values are given by matrix $X$ shown below.

$$X = \begin{bmatrix} 7 & 15 & 36 & 61 & 24 \\ 18 & 36 & 43 & 63 & 31 \\ 8 & 11 & 46 & 27 & 15 \\ 6 & 16 & 35 & 64 & 25 \\ 22 & 30 & 40 & 66 & 32 \\ 10 & 11 & 40 & 30 & 18 \\ 17 & 27 & 45 & 55 & 30 \end{bmatrix}$$

Also, obtain the mean of each sample and the covariance among the samples when the observed values given by matrix $Y$ shown below are added.

$$Y = \begin{bmatrix} 15 & 19 & 29 & 57 & 26 \\ 9 & 14 & 31 & 67 & 7 \\ 18 & 18 & 37 & 61 & 20 \end{bmatrix}$$

(b) Input data

First processing :

Observed value matrix $X$, NA = 100, N = 7, M = 5 and ISW = 0.

Second processing :

Observed value matrix $Y$, NA = 100, N = 3, M = 5 and ISW = 1.

(c) Main program

```
      PROGRAM B2VCMT
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 100, M = 5, ND = 10 )
      DIMENSION A(NA,M),X1(M),D(ND,M),WK(M)
!
      WRITE(6,6000)
      IERR =   0
      ISW  =   0
      READ(5,*) N
      DO 100 I = 1,N
         READ(5,*) (A(I,J),J=1,M)
  100 CONTINUE
      WRITE(6,6010) ISW,N,M
      DO 110 I=1,N
         WRITE(6,6020) (A(I,J),J=1,M)
  110 CONTINUE
      CALL D2VCMT(A,NA,N,M,NS,X1,D,ND,ISW,WK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS
      WRITE(6,6050) (X1(J),J=1,M)
      WRITE(6,6060)
      DO 120 I=1,M
         WRITE(6,6070) (D(I,J),J=1,M)
  120 CONTINUE
!
      WRITE(6,6080)
      IERR =  0
      ISW  =  1
      READ(5,*) N
      DO 130 I=1,N
         READ(5,*) (A(I,J),J=1,M)
  130 CONTINUE
      WRITE(6,6010) ISW,N,M
      DO 140 I=1,N
         WRITE(6,6020) (A(I,J),J=1,M)
  140 CONTINUE
      CALL D2VCMT(A,NA,N,M,NS,X1,D,ND,ISW,WK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS
      WRITE(6,6050) (X1(J),J=1,M)
      WRITE(6,6060)
      DO 150 I=1,M
         WRITE(6,6070) (D(I,J),J=1,M)
  150 CONTINUE
!
      STOP
 6000 FORMAT(  ' *** D2VCMT ***',/,&
              /,3X,'*** FIRST PROCESSING ***',/,&
              /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'ISW = ',I6,/,&
              /,7X,'N = ',I6,5X,'M = ',I6,/,&
              /,7X,'OBSERVATION MATRIX',/)
 6020 FORMAT(  7X,5(2X,F11.2))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
              /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'TOTAL SAMPLE SIZE = ',I6,/,&
              /,7X,'MEAN',/)
 6050 FORMAT(      5(2X,F11.2))
 6060 FORMAT( /,7X,'COVARIANCE MATRIX',/)
 6070 FORMAT(  9X,5(1X,D11.4))
 6080 FORMAT(/,/,3X,'*** CONTINUATION PROCESSING ***',/,&
              /,3X,'** INPUT **')
      END
```

(d) Output results

```
*** D2VCMT ***

  *** FIRST PROCESSING ***

  ** INPUT **

      ISW =      0

      N =      7     M =      5

      OBSERVATION MATRIX
```

```
            7.00          15.00          36.00          61.00          24.00
           18.00          36.00          43.00          63.00          31.00
            8.00          11.00          46.00          27.00          15.00
            6.00          16.00          35.00          64.00          25.00
           22.00          30.00          40.00          66.00          32.00
           10.00          11.00          40.00          30.00          18.00
           17.00          27.00          45.00          55.00          30.00

** OUTPUT **

    IERR =      0

    TOTAL SAMPLE SIZE =      7

    MEAN

     12.57         20.86         40.71         52.29         25.00

    COVARIANCE MATRIX

        0.3995D+02  0.5510D+02  0.1102D+02  0.4114D+02  0.3167D+02
        0.5510D+02  0.1005D+03  0.1079D+02  0.1109D+03  0.5983D+02
        0.1102D+02  0.1079D+02  0.1790D+02 -0.3324D+02 -0.2167D+01
        0.4114D+02  0.1109D+03 -0.3324D+02  0.2766D+03  0.9567D+02
        0.3167D+02  0.5983D+02 -0.2167D+01  0.9567D+02  0.4333D+02

*** CONTINUATION PROCESSING ***

** INPUT **

    ISW =      1

    N =      3     M =      5

    OBSERVATION MATRIX

           15.00          19.00          29.00          57.00          26.00
            9.00          14.00          31.00          67.00           7.00
           18.00          18.00          37.00          61.00          20.00

** OUTPUT **

    IERR =      0

    TOTAL SAMPLE SIZE =     10

    MEAN

     13.00         19.70         38.20         55.10         22.80

    COVARIANCE MATRIX

        0.3178D+02  0.3778D+02  0.7000D+01  0.2678D+02  0.2656D+02
        0.3778D+02  0.7201D+02  0.1496D+02  0.6259D+02  0.5216D+02
        0.7000D+01  0.1496D+02  0.3218D+02 -0.3991D+02  0.1260D+02
        0.2678D+02  0.6259D+02 -0.3991D+02  0.2105D+03  0.3691D+02
        0.2656D+02  0.5216D+02  0.1260D+02  0.3691D+02  0.6240D+02
```

### 4.3.2 D2VCGR, R2VCGR Variance-Covariance Matrices (Grouped Data)

(1) **Function**

Given $g$ groups and $m$ samples consisting of $n_r$ observed values for each group $\{x_{ki}^{(r)}\}(k = 1, \cdots, n_r; i = 1, \cdots, m; r = 1, \cdots, g)$, the D2VCGR or R2VCGR obtains the mean of each sample in each group, the mean of each sample over all groups, and the covariance among the samples over all groups. It also obtains the mean in each group and the mean and covariance over all groups when $n_r$ observed values $\{y_{ki}^{(r)}\}$ $(k = 1, \cdots, n_r; i = 1, \cdots, m; r = 1, \cdots, g)$ are added in each group to each of the $m$ samples for which the mean of each group and the mean and covariance over all groups are known.

The mean of each group and the mean and covariance over all groups for $\{x_{ki}^{(r)}\}$ $(k = 1, \cdots, n_r; i = 1, \cdots, m; r = 1, \cdots, g)$ are defined by the following equations.

Mean of each group :

$$\bar{x}_i^{(r)} = \frac{\sum_{k=1}^{n_r} x_{ki}^{(r)}}{n_r} \qquad i = 1, \cdots, m; r = 1, \cdots, g$$

Mean over all groups :

$$\bar{x}_i = \frac{\sum_{r=1}^{g} n_r \bar{x}^{(r)}}{\sum_{r=1}^{g} n_r} \qquad i = 1, \cdots, m$$

Covariance over all groups :

$$d_{ij} = \frac{\sum_{r=1}^{g} s_{ij}^{(r)}}{\sum_{r=1}^{g} \alpha_r} \qquad i, j = 1, \cdots, m$$

$s_{ij}^{(r)}$ (deviation sum of product matrix) of each group is defined as follows.

$$s_{ij}^{(r)} = \sum_{k=1}^{n_r} (x_{ki}^{(r)} - \bar{x}_i^{(r)})(x_{kj}^{(r)} - \bar{x}_j^{(r)}) \qquad i, j = 1, \cdots, m$$

Here, $\alpha_r$ is $n_r$ when a sample covariance is used or $\alpha_r$ is $n_r - 1$ when an unbiased covariance is used. The diagonal elements of the covariance matrix are the variance values.

(2) **Usage**

Double precision:

CALL D2VCGR (A, NA, M, N, K, NS, X1, Y, NY, D, ND, ISW, WK, IERR)

Single precision:

CALL R2VCGR (A, NA, M, N, K, NS, X1, Y, NY, D, ND, ISW, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\begin{cases} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{cases}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | NA,M | Input | Matrix in which observed values are stored ($x_{ki}^{(r)}$) or ($y_{ki}^{(r)}$) (See Note (a)) |
| 2 | NA | I | 1 | Input | Adjustable dimension of array A |
| 3 | M | I | 1 | Input | Number of samples $m$ |
| 4 | N | I | K | Input | Number of observed values per sample in each group stored in array A $n_r$ |
| 5 | K | I | 1 | Input | Number of groups $g$ |
| 6 | NS | I | K | Input | Number of observed values per sample in each group before adding observed values (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Number of observed values per sample in each group $n_r$ |
| 7 | X1 | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | M | Input | Mean of each sample over all groups before adding observed values (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Obtained mean of each sample over all groups |
| 8 | Y | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | NY,K | Input | Mean of each sample in each group before adding observed values (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Obtained mean of each sample in each group |
| 9 | NY | I | 1 | Input | Adjustable dimension of array Y |
| 10 | D | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | ND,M | Input | Covariance matrix over all groups before adding observed values (See Note (a)) (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Obtained covariance matrix over all groups (See Note (a)) |
| 11 | ND | I | 1 | Input | Adjustable dimension of array D |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 12 | ISW | I | 1 | Input | Processing switch<br>0: Perform calculations using an unbiased covariance (when the result of the previous calculation is not used)<br>1: Perform calculations using an unbiased covariance (when the result of the previous calculation is used)<br>2: Perform calculations using a sample covariance (when the result of the previous calculation is not used)<br>3: Perform calculations using a sample covariance (when the result of the previous calculation is used) |
| 13 | WK | $\left\{\begin{matrix} \mathrm{D} \\ \mathrm{R} \end{matrix}\right\}$ | See Contents | Work | Work area<br>**Size**: $\mathrm{M} \times (\mathrm{M} \times \mathrm{K} + 1)$ |
| 14 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW = 0, 1, 2, 3

(b) $\mathrm{M} \geq 1$

(c) $\mathrm{K} \geq 1$

(d) $\mathrm{ND} \geq \mathrm{M}$

(e) $\mathrm{NY} \geq \mathrm{M}$

(f) $\mathrm{N(i)} \geq \begin{cases} 1 & \text{(When ISW = 0 or 2)} \\ 0 & \text{(When ISW = 1 or 3)} \end{cases} \quad (\mathrm{i} = 1, \cdots, \mathrm{K})$

(g) $\mathrm{NS(i)} \geq 1 \quad (\mathrm{i} = 1, \cdots, \mathrm{K})$ (When ISW = 1 or 3)

(h) $\mathrm{NA} \geq \sum_{\mathrm{i}=1}^{\mathrm{K}} \mathrm{N(i)}$

(i) $\sum_{\mathrm{i}=1}^{\mathrm{K}} \mathrm{N(i)} \geq 1$ (When ISW = 1 or 3)

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (a) was not satisfied. | Processing continues with ISW=0. |
| 1010 | An attempt was made to obtain an unbiased covariance when N(1) $= \cdots =$ N(K) $= 1$. | The absolute value maximum that can be represented is set for the covariance. |
| 3000 | Any of restriction (b) to (f) were not satisfied. | Processing is aborted. |
| 3010 | Restriction (g) was not satisfied. | |
| 3020 | Restriction (h) was not satisfied. | |
| 3030 | Restriction (i) was not satisfied. | |
| 4000 | An error occurred in a lower level subroutine. | |

(6) **Notes**

(a) The observed values are stored as follows in array A as real matrix (two-dimensional array type) data (See Appendix A).

$$\begin{bmatrix}
x_{11}^{(1)} & x_{12}^{(1)} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & x_{1m}^{(1)} \\
x_{21}^{(1)} & x_{22}^{(1)} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & x_{2m}^{(1)} \\
\vdots & \vdots & & & & & & & & & \vdots \\
x_{n_1 1}^{(1)} & x_{n_1 2}^{(1)} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & x_{n_1 m}^{(1)} \\
 & & & & & & & & & & \\
x_{11}^{(2)} & x_{12}^{(2)} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & x_{1m}^{(2)} \\
x_{21}^{(2)} & x_{22}^{(2)} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & x_{2m}^{(2)} \\
\vdots & \vdots & & & & & & & & & \vdots \\
\vdots & \vdots & & & & & & & & & \vdots \\
x_{n_2 1}^{(2)} & x_{n_2 2}^{(2)} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & x_{n_2 m}^{(2)} \\
 & & & & & & & & & & \\
\vdots & \vdots & & & & & & & & & \vdots \\
x_{11}^{(g)} & x_{12}^{(g)} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & x_{1m}^{(g)} \\
x_{21}^{(g)} & x_{22}^{(g)} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & x_{2m}^{(1)} \\
\vdots & \vdots & & & & & & & & & \vdots \\
x_{n_g 1}^{(g)} & x_{n_g 2}^{(g)} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & x_{n_g m}^{(g)}
\end{bmatrix}$$

(b) To obtain the mean and covariance when the same numbers of observed values are added for each sample in each group, use the contents of NS, Y and WK, which were calculated before adding the observed values, set the added observed values for A and the number of added observed values for N, set ISW to 1 or 3, and perform the calculation. However, when obtaining the covariance, you must set the ISW value so that the calculation is performed using a sample covariance following a calculation that used a sample covariance the previous time or so that the calculation is performed using an unbiased covariance following a calculation that used an unbiased covariance the previous time.

(c) When there are an extremely large number of data values that are widely dispersed, better results are obtained by grouping them into data having absolute values of the same relative size and adding them to the samples in increasing order of size.

(d) Statistics obtained when calculations are performed using an unbiased covariance can be applied to a population for which sampling with replacement is performed from an infinite or finite population. Statistics obtained when calculations are performed using a sample covariance can be applied to a population for which the population and sample match.

(7) **Example**

(a) Problem
When three groups of observed values are given, and the observed values in each group are given by the matrices $X_1$, $X_2$ and $X_3$ shown below, obtain the mean of each sample in each group, the mean of each sample over all groups, and the covariance among the samples over all groups.

$$X_1 = \begin{bmatrix} 10 & 3 & 7 \\ 11 & 5 & 8 \\ 12 & 7 & 6 \end{bmatrix}$$

$$X_2 = \begin{bmatrix} 17 & 12 & 8 \\ 18 & 11 & 6 \\ 18 & 13 & 7 \\ 17 & 11 & 6 \end{bmatrix}$$

$$X_3 = \begin{bmatrix} 11 & 4 & 11 \\ 12 & 6 & 12 \\ 13 & 8 & 10 \\ 15 & 5 & 6 \end{bmatrix}$$

Also, obtain the mean of each sample in each group, the mean of each sample over all groups, and the covariance among the samples over all groups when the observed values given by matrices $Y_1$ and $Y_3$ shown below are added to groups 1 and 3.

$$Y_1 = \begin{bmatrix} 14 & 4 & 9 \\ 15 & 6 & 8 \end{bmatrix}$$

$$Y_3 = \begin{bmatrix} 18 & 10 & 13 \\ 14 & 5 & 7 \end{bmatrix}$$

(b) Input data
First processing :

Observed value matrices of each group $X_1$, $X_2$ and $X_3$,
NA=100, M=3, K=3,
N(1)=3, N(2)=4, N(3)=4,
NY=10, ND=10 and ISW=0.

Second processing :

Observed value matrices of each group $Y_1$ and $Y_3$,
NA=100, M=3, K=3,
N(1)=2, N(2)=0, N(3)=2,

NY=10, ND=10 and ISW=1.

(c) Main program

```
      PROGRAM B2VCGR
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 100, M = 3, NY = 10, K = 3, ND = 10 )
      DIMENSION N(K),NS(K)
      DIMENSION A(NA,M),X1(M),Y(NY,K),D(ND,M),WK(M*M*K+M)
!
      WRITE(6,6000)
      IERR =  0
      ISW  =  0
      DO 100 I=1,K
         READ(5,*) N(I)
  100 CONTINUE
      IA = 0
      DO 110 II=1,K
         DO 120 I=1,N(II)
            READ(5,*) (A(IA+I,J),J=1,M)
  120    CONTINUE
         IA = IA + N(II)
  110 CONTINUE
      WRITE(6,6010) ISW,M,K
      WRITE(6,6020) (N(II),II=1,K)
      WRITE(6,6030)
      IA = 0
      DO 130 II=1,K
         WRITE(6,6040) II
         DO 140 I=1,N(II)
            WRITE(6,6050) (A(IA+I,J),J=1,M)
  140    CONTINUE
         IA = IA + N(II)
  130 CONTINUE
      CALL D2VCGR(A,NA,M,N,K,NS,X1,Y,NY,D,ND,ISW,WK,IERR)
      WRITE(6,6060) IERR
      WRITE(6,6070)
      WRITE(6,6020) (NS(II),II=1,K)
      WRITE(6,6080)
      DO 150 II=1,K
         WRITE(6,6090) II,(Y(J,II),J=1,M)
  150 CONTINUE
      WRITE(6,6100)
      WRITE(6,6110) (X1(J),J=1,M)
      WRITE(6,6120)
      DO 160 J=1,M
         WRITE(6,6130) (D(J,L),L=1,M)
  160 CONTINUE
!
      WRITE(6,6140)
      IERR =  0
      ISW  =  1
      DO 170 I=1,K
         READ(5,*) N(I)
  170 CONTINUE
      IA = 0
      DO 180 II=1,K
         DO 190 I=1,N(II)
            READ(5,*) (A(IA+I,J),J=1,M)
  190    CONTINUE
         IA = IA + N(II)
  180 CONTINUE
      WRITE(6,6010) ISW,M,K
      WRITE(6,6020) (N(II),II=1,K)
      WRITE(6,6030)
      IA = 0
      DO 200 II=1,K
         WRITE(6,6040) II
         DO 210 I=1,N(II)
            WRITE(6,6050) (A(IA+I,J),J=1,M)
  210    CONTINUE
         IA = IA + N(II)
  200 CONTINUE
      CALL D2VCGR(A,NA,M,N,K,NS,X1,Y,NY,D,ND,ISW,WK,IERR)
      WRITE(6,6060) IERR
      WRITE(6,6070)
      WRITE(6,6020) (NS(II),II=1,K)
      WRITE(6,6080)
      DO 220 II=1,K
         WRITE(6,6090) II,(Y(J,II),J=1,M)
  220 CONTINUE
      WRITE(6,6100)
      WRITE(6,6110) (X1(J),J=1,M)
      WRITE(6,6120)
      DO 230 J=1,M
         WRITE(6,6130) (D(J,L),L=1,M)
  230 CONTINUE
!
      STOP
 6000 FORMAT(  ' *** D2VCGR ***',/,&
              /,3X,'*** FIRST PROCESSING ***',/,&
              /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'ISW = ',I6,/,&
```

```
             /,7X,'M = ',I6,5X,'K = ',I6,/,&
             /,7X,'NUMBER OF OBSERVATIONS IN EACH GROUP',/)
6020 FORMAT(  9X,5(2X,I6))
6030 FORMAT( /,7X,'OBSERVATION MATRIX')
6040 FORMAT( /,9X,'GROUP ',I2)
6050 FORMAT(  9X,5(2X,F11.2))
6060 FORMAT( /,3X,'** OUTPUT **',/,&
             /,7X,'IERR = ',I6)
6070 FORMAT( /,7X,'TOTAL NUMBER OF OBSERVATIONS IN EACH GROUP',/)
6080 FORMAT( /,7X,'MEAN OF EACH GROUP',/,&
             /,10X,'GROUP',/,&
              9X,43('-'))
6090 FORMAT(  7X,I6,2X,5(1X,F11.2))
6100 FORMAT( /,7X,'MEAN OVER ALL GROUPS',/)
6110 FORMAT(  9X,5(1X,F11.2))
6120 FORMAT( /,7X,'COVARIANCE MATRIX',/)
6130 FORMAT(  9X,5(1X,D11.4))
6140 FORMAT(/,/,3X,'*** CONTINUATION PROCESSING ***',/,&
             /,3X,'** INPUT **')
     END
```

(d) Output results

```
*** D2VCGR ***

  *** FIRST PROCESSING ***

  ** INPUT **

      ISW =      0

      M =      3     K =      3

      NUMBER OF OBSERVATIONS IN EACH GROUP

              3       4       4

      OBSERVATION MATRIX

        GROUP  1
                10.00         3.00         7.00
                11.00         5.00         8.00
                12.00         7.00         6.00

        GROUP  2
                17.00        12.00         8.00
                18.00        11.00         6.00
                18.00        13.00         7.00
                17.00        11.00         6.00

        GROUP  3
                11.00         4.00        11.00
                12.00         6.00        12.00
                13.00         8.00        10.00
                15.00         5.00         6.00

  ** OUTPUT **

      IERR =      0

      TOTAL NUMBER OF OBSERVATIONS IN EACH GROUP

              3       4       4

      MEAN OF EACH GROUP

         GROUP
        -------------------------------------------
           1         11.00        5.00        7.00
           2         17.50       11.75        6.75
           3         12.75        5.75        9.75

      MEAN OVER ALL GROUPS

              14.00        7.73        7.91

      COVARIANCE MATRIX

          0.1469D+01  0.7813D+00 -0.1719D+01
          0.7813D+00  0.2438D+01  0.1875D+00
         -0.1719D+01  0.1875D+00  0.3188D+01


  *** CONTINUATION PROCESSING ***

  ** INPUT **

      ISW =      1

      M =      3     K =      3

      NUMBER OF OBSERVATIONS IN EACH GROUP

              2       0       2
```

```
    OBSERVATION MATRIX

      GROUP  1
              14.00          4.00          9.00
              15.00          6.00          8.00

      GROUP  2

      GROUP  3
              18.00         10.00         13.00
              14.00          5.00          7.00

** OUTPUT **

    IERR =      0

    TOTAL NUMBER OF OBSERVATIONS IN EACH GROUP

             5       4       6

    MEAN OF EACH GROUP

       GROUP
      -------------------------------------------
         1        12.40         5.00         7.60
         2        17.50        11.75         6.75
         3        13.83         6.33         9.83

    MEAN OVER ALL GROUPS

             14.33        7.33        8.27

    COVARIANCE MATRIX

        0.4086D+01  0.2069D+01  0.4278D+00
        0.2069D+01  0.3174D+01  0.1340D+01
        0.4278D+00  0.1340D+01  0.3899D+01
```

## 4.4 CORRELATION COEFFICIENTS

### 4.4.1 D2CCMT, R2CCMT
### Correlation Matrices

(1) **Function**

Given $m$ samples consisting of $n$ observed values $\boldsymbol{x}_i = \{x_{ki}\}(k = 1, \cdots, n; i = 1, \cdots, m)$, the D2CCMT or R2CCMT obtains the mean of each sample and the correlation coefficients between the samples. It also obtains the mean of each sample and the correlation coefficients between the samples when $n$ observed values $\boldsymbol{y}_i = \{y_{ki}\}(k = 1, \cdots, n; i = 1, \cdots, m)$ are added to each of the $m$ samples.

The mean $\bar{x}_i$ of sample $\boldsymbol{x}_i$ $(i = 1, \cdots, m)$ is defined as follows.

$$\bar{x}_i = \frac{t_i}{n}, \qquad i = 1, \cdots, m$$

Here, $t_i$ is the sum, which is defined as follows.

$$t_i = \sum_{k=1}^{n} x_{ki}, \qquad i = 1, \cdots, m$$

The correlation coefficient $r_{ij}$ between samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is defined as follows.

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii} \cdot s_{jj}}}, \qquad i = 1, \cdots, m; j = 1, \cdots, m$$

$s_{ij}$ is defined as follows.

$$s_{ij} = \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j), \qquad i = 1, \cdots, m; j = 1, \cdots, m$$

The matrix $R = (r_{ij})$ is called the correlation coefficient matrix.

(2) **Usage**

Double precision:

CALL D2CCMT (A, NA, N, M, NS, X1, R, NR, ISW, WK, IERR)

Single precision:

CALL R2CCMT (A, NA, N, M, NS, X1, R, NR, ISW, WK, IERR)

## (3) Arguments

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | NA,M | Input | Matrix in which observed values are stored ($x_{ki}$) or ($y_{ki}$) $(k = 1, \cdots, n;\ i = 1, \cdots, m)$ |
| 2 | NA | I | 1 | Input | Adjustable dimension of array A |
| 3 | N | I | 1 | Input | Number of observed values per sample stored in array A $n$ |
| 4 | M | I | 1 | Input | Number of samples $m$ |
| 5 | NS | I | 1 | Input | Number of observed values per sample before adding observed values (for ISW=0, no initial setting is required) |
| | | | | Output | Number of observed values per sample $n$ |
| 6 | X1 | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | M | Input | Mean of each sample before adding observed values (for ISW=0, no initial setting is required) |
| | | | | Output | Obtained mean of each sample |
| 7 | R | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | NR,M | Input | Correlation coefficients between samples before adding observed values (for ISW=0, no initial setting is required) |
| | | | | Output | Obtained correlation coefficients between samples |
| 8 | NR | I | 1 | Input | Adjustable dimension of array R |
| 9 | ISW | I | 1 | Input | Processing switch<br>0: First calculation (when the result of the previous calculation is not used)<br>1: Add observed values (when the result of the previous calculation is used) |
| 10 | WK | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | M | Work | Work area |
| 11 | IERR | I | 1 | Output | Error indicator |

## (4) Restrictions

(a) ISW = 0, 1

(b) NA $\geq$ N $\geq$ 1

(c) NR $\geq$ M $\geq$ 1

(d) NS $\geq$ 1 (When ISW=1)

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (a) was not satisfied. | Processing continues with ISW=0. |
| 1010 | An attempt was made to obtain correlation coefficients when ISW=0 and N=1. | The absolute value maximum that can be represented is set for the correlation coefficients. |
| 1020 | The variance was zero because all values of a certain sample were equal. | |
| 3000 | Restriction (b), (c) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (d) was not satisfied. | |

(6) **Notes**

(a) To obtain the statistics when the same numbers of observed values are added for each sample, use the contents of X1, R, NS and WK, which were calculated before adding the observed values, set the added observed values for A and the number of added observed values for N, set ISW to 1, and perform the calculation.

(b) If a job that had been executed with ISW=0 terminated with IERR=1020, the result will not be guaranteed if a job is executed with ISW=1 following it.

(c) When there are an extremely large number of data values that are widely dispersed, better results are obtained by grouping them into data having absolute values of the same relative size and adding them to the samples in increasing order of size.

(7) **Example**

(a) Problem

Obtain the mean of each sample and the correlation coefficients between the samples when the observed values are given by matrix $X$ shown below.

$$X = \begin{bmatrix} 7 & 9 & 15 & 60 & 24 \\ 13 & 25 & 13 & 61 & 30 \\ 9 & 24 & 12 & 62 & 31 \\ 7 & 25 & 11 & 63 & 32 \\ 6 & 20 & 15 & 18 & 15 \\ 10 & 30 & 10 & 27 & 17 \\ 7 & 11 & 15 & 60 & 25 \end{bmatrix}$$

Also, obtain the mean of each sample and the correlation coefficients between the samples when the observed values given by matrix $Y$ shown below are added.

$$Y = \begin{bmatrix} 16 & 25 & 13 & 64 & 30 \\ 9 & 26 & 13 & 66 & 32 \\ 8 & 26 & 13 & 66 & 34 \end{bmatrix}$$

(b) Input data

First processing :

Matrix $X$ in which observed values are stored, NA = 100, N = 7, M = 5 and ISW = 0.

Second processing :

Matrix $Y$ in which observed values are stored, NA = 100, N = 3, M = 5 and ISW = 1.

(c) Main program

```
      PROGRAM B2CCMT
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 20, NR = 5, M = 5 )
      DIMENSION A(NA,M),R(NR,M),X1(M),WK(M)
!
      WRITE(6,6000)
      IERR =  0
      ISW  =  0
      READ(5,*) N
      DO 100 I=1,N
         READ(5,*) (A(I,J),J=1,M)
  100 CONTINUE
      WRITE(6,6010) ISW,N,M
      DO 110 I=1,N
         WRITE(6,6020) (A(I,J),J=1,M)
  110 CONTINUE
      CALL D2CCMT(A,NA,N,M,NS,X1,R,NR,ISW,WK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS
      WRITE(6,6050) (X1(I),I=1,M)
      WRITE(6,6060)
      DO 120 I=1,M
         WRITE(6,6070) (R(I,J),J=1,M)
  120 CONTINUE
!
      WRITE(6,6080)
      IERR = 0
      ISW  = 1
      READ(5,*) N
      DO 130 I=1,N
         READ(5,*) (A(I,J),J=1,M)
  130 CONTINUE
      WRITE(6,6010) ISW,N,M
      DO 140 I=1,N
         WRITE(6,6020) (A(I,J),J=1,M)
  140 CONTINUE
      CALL D2CCMT(A,NA,N,M,NS,X1,R,NR,ISW,WK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS
      WRITE(6,6050) (X1(I),I=1,M)
      WRITE(6,6060)
      DO 150 I=1,M
         WRITE(6,6070) (R(I,J),J=1,M)
  150 CONTINUE
!
      STOP
 6000 FORMAT(  ' *** D2CCMT ***',/,&
              /,3X,'*** FIRST PROCESSING ***',/,&
              /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'ISW = ',I6,/,&
              /,7X,'N = ',I6,5X,'M = ',I6,/,&
              /,7X,'OBSERVATION MATRIX',/)
 6020 FORMAT(  7X,5(2X,F11.2))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
              /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'TOTAL SAMPLE SIZE = ',I6,/,&
              /,7X,'MEAN',/)
 6050 FORMAT(      5(2X,F11.2))
 6060 FORMAT( /,7X,'CORRELATION COEFFICIENT MATRIX',/)
 6070 FORMAT(  7X,5(2X,D11.4))
 6080 FORMAT(/,/,3X,'*** CONTINUATION PROCESSING ***',/,&
              /,3X,'** INPUT **')
      END
```

(d) Output results

```
*** D2CCMT ***

  *** FIRST PROCESSING ***

  ** INPUT **

      ISW =      0

      N =      7     M =      5

      OBSERVATION MATRIX

              7.00         9.00        15.00        60.00        24.00
             13.00        25.00        13.00        61.00        30.00
              9.00        24.00        12.00        62.00        31.00
              7.00        25.00        11.00        63.00        32.00
              6.00        20.00        15.00        18.00        15.00
             10.00        30.00        10.00        27.00        17.00
              7.00        11.00        15.00        60.00        25.00
```

```
** OUTPUT **

    IERR =      0

    TOTAL SAMPLE SIZE =      7

    MEAN

      8.43        20.57        13.00        50.14        24.86

    CORRELATION COEFFICIENT MATRIX

       0.1000D+01   0.5450D+00  -0.4266D+00   0.1845D+00   0.2970D+00
       0.5450D+00   0.1000D+01  -0.8614D+00  -0.2936D+00   0.4911D-01
      -0.4266D+00  -0.8614D+00   0.1000D+01   0.2936D-01  -0.2129D+00
       0.1845D+00  -0.2936D+00   0.2936D-01   0.1000D+01   0.9186D+00
       0.2970D+00   0.4911D-01  -0.2129D+00   0.9186D+00   0.1000D+01

*** CONTINUATION PROCESSING ***

** INPUT **

    ISW =      1

    N =      3     M =      5

    OBSERVATION MATRIX

            16.00        25.00        13.00        64.00        30.00
             9.00        26.00        13.00        66.00        32.00
             8.00        26.00        13.00        66.00        34.00

** OUTPUT **

    IERR =      0

    TOTAL SAMPLE SIZE =     10

    MEAN

      9.20        22.10        13.00        54.70        27.00

    CORRELATION COEFFICIENT MATRIX

       0.1000D+01   0.4416D+00  -0.2724D+00   0.2554D+00   0.2813D+00
       0.4416D+00   0.1000D+01  -0.8030D+00  -0.9303D-01   0.2320D+00
      -0.2724D+00  -0.8030D+00   0.1000D+01   0.2655D-01  -0.1788D+00
       0.2554D+00  -0.9303D-01   0.2655D-01   0.1000D+01   0.9244D+00
       0.2813D+00   0.2320D+00  -0.1788D+00   0.9244D+00   0.1000D+01
```

### 4.4.2 D2CCMA, R2CCMA Multiple Correlation Coefficients

(1) **Function**

Given $m$ samples consisting of $n$ observed values $\{x_{ki}\}(k = 1, \cdots, n; i = 1, \cdots, m)$, the D2CCMA or R2CCMA obtains the basic statistics (sum, mean, sum of squares of deviation, variance, and standard deviation) and multiple correlation coefficients of each sample. It also obtains the basic statistics and multiple correlation coefficients when $n$ observed values $\{y_{ki}\}$ $(k = 1, \cdots, n; i = 1, \cdots, m)$ are added to each of the $m$ samples for which the basic statistics are known.

The basic statistics and multiple correlation coefficients for $m$ samples consisting of $n$ observed values $\{x_{ki}\}$ $(k = 1, \cdots, n; i = 1, \cdots, m)$ are defined by the following equations.

Sum :

$$t_i = \sum_{k=1}^{n} x_{ki}, \quad i = 1, \cdots, m$$

Mean :

$$\bar{x}_i = \frac{t_i}{n}, \quad i = 1, \cdots, m$$

Sum of squares of deviation :

$$s_i = \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)^2, \quad i = 1, \cdots, m$$

Variance :

$$v_i = \frac{s_i}{\alpha}, \quad i = 1, \cdots, m$$

Standard deviation :

$$d_i = \sqrt{v_i}, \quad i = 1, \cdots, m$$

Here, $\alpha$ is $n - 1$ when an unbiased estimate is used or $\alpha$ is $n$ when a sample variance is used.

Multiple correlation coefficients :

$$r_{i \cdot 1, \cdots, i-1, i+1, \cdots, m} = \sqrt{1 - \frac{\Delta}{\Delta_{ii}}}, \quad i = 1, \cdots, m$$

Here, $\Delta$ and $\Delta_{ij}$ are the determinant and cofactor matrix of the matrix having the correlation coefficients $r_{ij}$ $(i, j = 1, \cdots, m)$ as elements.

(2) **Usage**

Double precision:

CALL D2CCMA (A, NA, N, M, NS, STAT, R, ISW, WK, IERR)

Single precision:

CALL R2CCMA (A, NA, N, M, NS, STAT, R, ISW, WK, IERR)

## (3) **Arguments**

D:Double precision real  Z:Double precision complex  I: { INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer }
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | {D / R} | NA,M | Input | Matrix in which observed values are stored ($x_{ki}$) or ($y_{ki}$) (See Note (a)) |
| 2 | NA | I | 1 | Input | Adjustable dimension of array A |
| 3 | N | I | 1 | Input | Number of observed values per sample stored in array A $n$ |
| 4 | M | I | 1 | Input | Number of samples $m$ |
| 5 | NS | I | 1 | Input | Number of observed values per sample before adding observed values (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Number of observed values per sample $n$ |
| 6 | STAT | {D / R} | M,5 | Input | Basic statistics before adding observed values (See Note (b)) (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Obtained basic statistics (See Note (b)) |
| 7 | R | {D / R} | M | Output | Multiple correlation coefficients of each sample |
| 8 | ISW | I | 1 | Input | Processing switch<br>0: Perform calculations using an unbiased estimate (no previously established basic statistics)<br>1: Perform calculations using an unbiased estimate (added observed values)<br>2: Perform calculations using a sample variance (no previously established basic statistics)<br>3: Perform calculations using a sample variance (added observed values) |
| 9 | WK | {D / R} | See Contents | Work | Work area<br>**Size:**2×M×M+3×M |
| 10 | IERR | I | 1 | Output | Error indicator |

## (4) **Restrictions**

(a) ISW = 0, 1, 2, 3

(b) NA $\geq$ N $\geq$ 1

(c) M $\geq$ 1

(d) NS $\geq$ 1 (When ISW=1 or 3)

(e) N $\geq$ M (When ISW=0 or 2)

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (a) was not satisfied. | Processing continues with ISW=0. |
| 1010 | N = 1 was specified. | When ISW = 0, the absolute value maximum that can be represented is set for the variance, standard deviation, and multiple correlation coefficients.<br>When ISW = 2, the absolute value maximum that can be represented is set for the multiple correlation coefficients. |
| 1020 | The variance was zero because all values of a certain sample were equal. | The absolute value maximum that can be represented is set for the multiple correlation coefficients. |
| 3000 | Restriction (b), (c) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (d) was not satisfied. | |
| 3020 | Restriction (e) was not satisfied. | |
| 4000 | The inverse matrix of the simple correlation matrix was not obtained. | |

(6) **Notes**

(a) The observed values $x_{ki}$ or $y_{ki}$ $(k = 1, 2, \cdots, n; i = 1, 2, \cdots, m)$ are stored in array A as real matrix (two-dimensional array type) data (See Appendix A).

(b) The basic statistics are stored as follows in the array STAT.

STAT($i$, 1) : Sum $t_i$
STAT($i$, 2) : Mean $\bar{x}_i$
STAT($i$, 3) : Sum of squares of deviation $s_i$ , i = $1, \cdots,$ M
STAT($i$, 4) : Variance $v_i$
STAT($i$, 5) : Standard deviation $d_i$

(c) To obtain the basic statistics and multiple correlation coefficients when the same numbers of observed values are added for each sample, use the contents of STAT, NS and WK, which were calculated before adding the observed values, set the added observed values for A and the number of added observed values for N, set ISW to 1 or 3, and perform the calculation. However, when obtaining the variance or standard deviation, you must set the ISW value so that the calculation is performed using a sample variance following a calculation that used a sample variance the previous time or so that the calculation is performed using an unbiased estimate following a calculation that used an unbiased estimate the previous time.

(d) If a job that had been executed with ISW=0 or ISW=2 terminated with IERR=1020, the result will not be guaranteed if a job is executed with ISW=1 or ISW=3 following it.

(e) When there are an extremely large number of data values that are widely dispersed, better results are obtained by grouping them into data having absolute values of the same relative size and adding them to the samples in increasing order of size.

(f) Statistics obtained when calculations are performed using an unbiased estimate can be applied to a population for which sampling with replacement is performed from an infinite or finite population. Statistics obtained when calculations are performed using a sample variance can be applied to a population for which the population and sample match.

(7) **Example**

(a) Problem

Obtain the basic statistics (sum, mean, sum of squares of deviation, variance, and standard deviation) and multiple correlation coefficients for each sample when the observed values are given by matrix $X$ shown below.

$$X = \begin{bmatrix} 23.9 & 64.6 & 2.41 \\ 21.4 & 65.2 & 2.14 \\ 23.6 & 57.7 & 2.61 \\ 23.2 & 61.0 & 2.24 \\ 25.0 & 86.5 & 2.78 \\ 25.7 & 88.8 & 2.95 \\ 23.2 & 91.0 & 2.91 \end{bmatrix}$$

Also, obtain the basic statistics (sum, mean, sum of squares of deviation, variance, and standard deviation) and multiple correlation coefficients for each sample when the observed values given by matrix $Y$ shown below are added.

$$Y = \begin{bmatrix} 24.3 & 84.6 & 3.12 \\ 25.3 & 89.2 & 3.01 \\ 26.5 & 90.8 & 2.73 \end{bmatrix}$$

(b) Input data

First processing :

Matrix $X$ in which observed values are stored, NA = 100, N = 7, M = 3 and ISW = 0.

Second processing :

Matrix $Y$ in which observed values are stored, NA = 100, N = 3, M = 3 and ISW = 1.

(c) Main program

```
      PROGRAM B2CCMA
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 100, M = 3 )
      DIMENSION A(NA,M),STAT(M,5),R(M),WK(2*M*M+3*M)
!
      WRITE(6,6000)
      IERR =  0
      ISW  =  0
      READ(5,*) N
      DO 100 I=1,N
         READ(5,*) (A(I,J),J=1,M)
  100 CONTINUE
      WRITE(6,6010) ISW,N,M
      DO 110 I=1,N
         WRITE(6,6020) (A(I,J),J=1,M)
  110 CONTINUE
      CALL D2CCMA(A,NA,N,M,NS,STAT,R,ISW,WK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS
      DO 120 J=1,M
         WRITE(6,6050) J,(STAT(J,I),I=1,5),R(J)
  120 CONTINUE
!
      WRITE(6,6060)
      IERR = 0
      ISW  = 1
      READ(5,*) N
      DO 130 I=1,N
```

```
         READ(5,*) (A(I,J),J=1,M)
  130 CONTINUE
      WRITE(6,6010) ISW,N,M
      DO 140 I=1,N
         WRITE(6,6020) (A(I,J),J=1,M)
  140 CONTINUE
      CALL D2CCMA(A,NA,N,M,NS,STAT,R,ISW,WK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS
      DO 150 J=1,M
         WRITE(6,6050) J,(STAT(J,I),I=1,5),R(J)
  150 CONTINUE
!
      STOP
 6000 FORMAT(  ' *** D2CCMA ***',/,&
              /,3X,'*** FIRST PROCESSING ***',/,&
              /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'ISW = ',I6,/,&
              /,7X,'N = ',I6,5X,'M = ',I6,/,&
              /,7X,'OBSERVATION MATRIX',/)
 6020 FORMAT(  7X,5(2X,F11.2))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
              /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'TOTAL SAMPLE SIZE = ',I6,/,&
              /,68X,'MULTIPLE',/,&
               34X,'SUM OF',15X,'STANDARD',3X,'CORRELATION',/,&
               3X,'VARIABLE',4X,'SUM',7X,'MEAN',4X,'SQUARES',3X,&
                   'VARIANCE',3X,'DEVIATION',3X,'COEFFICIENT',/,&
               3X,74('-'))
 6050 FORMAT(  ' ',I6,3F11.2,3(1X,D11.4))
 6060 FORMAT(/,/,3X,'*** CONTINUATION PROCESSING ***',/,&
              /,3X,'** INPUT **')
      END
```

(d) Output results

```
 *** D2CCMA ***

   *** FIRST PROCESSING ***

   ** INPUT **

       ISW =      0

       N =      7     M =      3

       OBSERVATION MATRIX

               23.90        64.60         2.41
               21.40        65.20         2.14
               23.60        57.70         2.61
               23.20        61.00         2.24
               25.00        86.50         2.78
               25.70        88.80         2.95
               23.20        91.00         2.91

   ** OUTPUT **

       IERR =      0

       TOTAL SAMPLE SIZE =      7

                                                                    MULTIPLE
                                  SUM OF               STANDARD   CORRELATION
   VARIABLE    SUM       MEAN    SQUARES   VARIANCE   DEVIATION   COEFFICIENT
   --------------------------------------------------------------------------
      1     166.00      23.71      11.53 0.1921D+01 0.1386D+01 0.7491D+00
      2     514.80      73.54    1263.32 0.2106D+03 0.1451D+02 0.8253D+00
      3      18.04       2.58       0.62 0.1041D+00 0.3227D+00 0.8934D+00


   *** CONTINUATION PROCESSING ***

   ** INPUT **

       ISW =      1

       N =      3     M =      3

       OBSERVATION MATRIX

               24.30        84.60         3.12
               25.30        89.20         3.01
               26.50        90.80         2.73

   ** OUTPUT **

       IERR =      0

       TOTAL SAMPLE SIZE =     10

                                                                    MULTIPLE
                                  SUM OF               STANDARD   CORRELATION
   VARIABLE    SUM       MEAN    SQUARES   VARIANCE   DEVIATION   COEFFICIENT
```

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 242.10 | 24.21 | 19.69 | 0.2188D+01 | 0.1479D+01 | 0.6847D+00 |
| 2 | 779.40 | 77.94 | 1735.18 | 0.1928D+03 | 0.1389D+02 | 0.8229D+00 |
| 3 | 26.90 | 2.69 | 1.00 | 0.1114D+00 | 0.3338D+00 | 0.8159D+00 |

### 4.4.3 D2CCPR, R2CCPR Partial Correlation Coefficients

(1) **Function**

Given $m$ samples consisting of n observed values $\{x_{ki}\}(k = 1, \cdots, n; i = 1, \cdots, m)$, the D2CCPR or R2CCPR obtains the basic statistics (sum, mean, sum of squares of deviation, variance, and standard deviation) and partial correlation coefficients of each sample. It also obtains the basic statistics and partial correlation coefficients when $n$ observed values $\{y_{ki}\}$ $(k = 1, \cdots, n; i = 1, \cdots, m)$ are added to each of the $m$ samples for which the basic statistics are known.

The basic statistics and partial correlation coefficients for $m$ samples consisting of $n$ observed values $\{x_{ki}\}$ $(k = 1, \cdots, n; i = 1, \cdots, m)$ are defined by the following equations.

Sum :

$$t_i = \sum_{k=1}^{n} x_{ki}, \quad i = 1, \cdots, m$$

Mean :

$$\bar{x}_i = \frac{t_i}{n}, \quad i = 1, \cdots, m$$

Sum of squares of deviation :

$$s_i = \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)^2, \quad i = 1, \cdots, m$$

Variance :

$$v_i = \frac{s_i}{\alpha}, \quad i = 1, \cdots, m$$

Standard deviation :

$$d_i = \sqrt{v_i}, \quad i = 1, \cdots, m$$

Here, $\alpha$ is $n - 1$ when an unbiased estimate is used or $\alpha$ is $n$ when a sample variance is used.

Partial correlation coefficients :

$$r_{i,j \cdot 1,\cdots,i-1,i+1,\cdots j-1,j+1,\cdots m} = -\frac{\Delta_{ij}}{\sqrt{\Delta_{ii}\Delta_{jj}}} \quad i, j = 1, \cdots, m$$

Here, $\Delta_{ij}$ is the cofactor matrix of the matrix having the correlation coefficients $r_{ij}$ $(i, j = 1, \cdots, m)$ as elements.

(2) **Usage**

Double precision:

CALL D2CCPR (A, NA, N, M, NS, STAT, R, NR, ISW, WK, IERR)

Single precision:

CALL R2CCPR (A, NA, N, M, NS, STAT, R, NR, ISW, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | NA,M | Input | Matrix in which observed values are stored ($x_{ki}$) or ($y_{ki}$)(See Note (a)) |
| 2 | NA | I | 1 | Input | Adjustable dimension of array A |
| 3 | N | I | 1 | Input | Number of observed values per sample stored in array A $n$ |
| 4 | M | I | 1 | Input | Number of samples $m$ |
| 5 | NS | I | 1 | Input | Number of observed values per sample before adding observed values (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Number of observed values per sample $n$ |
| 6 | STAT | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | M,5 | Input | Basic statistics before adding observed values (See Note (b)) (for ISW=0 or 2, no initial setting is required) |
| | | | | Output | Obtained basic statistics (See Note (b)) |
| 7 | R | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | NR,M | Output | Partial correlation coefficients between samples $r_{i,j\cdot 1,\cdots,i-1,i+1,\cdots j-1,j+1,\cdots m}$ (See Note (a)) |
| 8 | NR | I | 1 | Input | Adjustable dimension of array R |
| 9 | ISW | I | 1 | Input | Processing switch<br>0: Perform calculations using an unbiased estimate (no previously established basic statistics)<br>1: Perform calculations using an unbiased estimate (added observed values)<br>2: Perform calculations using a sample variance (no previously established basic statistics)<br>3: Perform calculations using a sample variance (added observed values) |
| 10 | WK | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | See Contents | Work | Work area<br>2×M×M+3×M |
| 11 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW = 0, 1, 2, 3

(b) NA $\geq$ N $\geq$ 1

(c) NR $\geq$ M $\geq$ 1

(d) NS $\geq$ 1 (When ISW=1 or 3)

(e) N $\geq$ M (When ISW=0 or 2)

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (a) was not satisfied. | Processing continues with ISW=0. |
| 1010 | N = 1 was specified. | When ISW = 0, the absolute value maximum that can be represented is set for the variance, standard deviation, and partial correlation coefficients. When ISW = 2, the absolute value maximum that can be represented is set for the partial correlation coefficients. |
| 1020 | The variance was zero because all values of a certain sample were equal. | The absolute value maximum that can be represented is set for the partial correlation coefficients. |
| 3000 | Restriction (b), (c) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (d) was not satisfied. | |
| 3020 | Restriction (e) was not satisfied. | |
| 4000 | The inverse matrix of the simple correlation matrix was not obtained. | |

(6) **Notes**

(a) The observed values $x_{ki}$ or $y_{ki}$ $(k = 1, 2, \cdots, n; i = 1, 2, \cdots, m)$ are stored in array A as real matrix (two-dimensional array type) data.
Also, the partial correlation coefficients $\hat{r}_{ij} = r_{i,j\cdot 1,\cdots,i-1,i+1,\cdots j-1,j+1,\cdots m}$ $(i, j = 1, 2, \cdots, m)$ are stored in array R as real matrix (two-dimensional array type) data.
For the method of array data storage, see Appendix A.

(b) The basic statistics are stored as follows in the array STAT.

STAT($i$, 1) : Sum $t_i$
STAT($i$, 2) : Mean $\bar{x}_i$
STAT($i$, 3) : Sum of squares of deviation $s_i$ , i = $1, \cdots,$ M
STAT($i$, 4) : Variance $v_i$
STAT($i$, 5) : Standard deviation $d_i$

(c) To obtain the basic statistics and partial correlation coefficients when the same numbers of observed values are added for each sample, use the contents of STAT, NS and WK, which were calculated before adding the observed values, set the added observed values for A and the number of added observed values for N, set ISW to 1 or 3, and perform the calculation. However, when obtaining the variance or standard deviation, you must set the ISW value so that the calculation is performed using a sample variance following a calculation that used a sample variance the previous time or so that the calculation is performed using an unbiased estimate following a calculation that used an unbiased estimate the previous time.

(d) If a job that had been executed with ISW=0 or ISW=2 terminated with IERR=1020, the result will not be guaranteed if a job is executed with ISW=1 or ISW=3 following it.

(e) When there are an extremely large number of data values that are widely dispersed, better results are obtained by grouping them into data having absolute values of the same relative size and adding them to the samples in increasing order of size.

(f) Statistics obtained when calculations are performed using an unbiased estimate can be applied to a population for which sampling with replacement is performed from an infinite or finite population. Statistics obtained when calculations are performed using a sample variance can be applied to a population for which the population and sample match.

(7) **Example**

(a) Problem

Obtain the basic statistics (sum, mean, sum of squares of deviation, variance, and standard deviation) and partial correlation coefficients for each sample when the observed values are given by matrix $X$ shown below.

$$X = \begin{bmatrix} 84 & 58 & 42 \\ 92 & 88 & 86 \\ 88 & 80 & 98 \\ 66 & 72 & 64 \\ 64 & 50 & 40 \\ 80 & 94 & 74 \\ 90 & 92 & 94 \end{bmatrix}$$

Also, obtain the basic statistics (sum, mean, sum of squares of deviation, variance, and standard deviation) and partial correlation coefficients for each sample when the observed values given by matrix $Y$ shown below are added.

$$Y = \begin{bmatrix} 40 & 36 & 8 \\ 78 & 50 & 86 \\ 82 & 62 & 66 \end{bmatrix}$$

(b) Input data

First processing :

Matrix $X$ in which observed values are stored, NA $= 100$, N $= 7$, M $= 3$ and ISW $= 0$.

Second processing :

Matrix $Y$ in which observed values are stored, NA $= 100$, N $= 3$, M $= 3$ and ISW $= 1$.

(c) Main program

```
      PROGRAM B2CCPR
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 100, M = 3, NR = 3 )
      DIMENSION A(NA,M),STAT(M,5),R(NR,M),WK(2*M*M+3*M)
!
      WRITE(6,6000)
      IERR = 0
      ISW  = 0
      READ(5,*) N
      DO 100 I=1,N
         READ(5,*) (A(I,J),J=1,M)
  100 CONTINUE
      WRITE(6,6010) ISW,N,M
      DO 110 I=1,N
         WRITE(6,6020) (A(I,J),J=1,M)
  110 CONTINUE
      CALL D2CCPR(A,NA,N,M,NS,STAT,R,NR,ISW,WK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS
      DO 120 J=1,M
         WRITE(6,6050) J,(STAT(J,I),I=1,5)
  120 CONTINUE
      WRITE(6,6060)
```

```
      DO 130 I=1,M
         WRITE(6,6070) (R(I,J),J=1,M)
  130 CONTINUE
!
      WRITE(6,6080)
      IERR = 0
      ISW  = 1
      READ(5,*) N
      DO 140 I=1,N
         READ(5,*) (A(I,J),J=1,M)
  140 CONTINUE
      WRITE(6,6010) ISW,N,M
      DO 150 I=1,N
         WRITE(6,6020) (A(I,J),J=1,M)
  150 CONTINUE
      CALL D2CCPR(A,NA,N,M,NS,STAT,R,NR,ISW,WK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) NS
      DO 160 J=1,M
         WRITE(6,6050) J,(STAT(J,I),I=1,5)
  160 CONTINUE
      WRITE(6,6060)
      DO 170 I=1,M
         WRITE(6,6070) (R(I,J),J=1,M)
  170 CONTINUE
!
      STOP
 6000 FORMAT(  ' *** D2CCPR ***',/,&
              /,3X,'*** FIRST PROCESSING ***',/,&
              /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'ISW = ',I6,/,&
              /,7X,'N = ',I6,5X,'M = ',I6,/,&
              /,7X,'OBSERVATION MATRIX',/)
 6020 FORMAT(  7X,5(2X,F11.2))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
              /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'TOTAL SAMPLE SIZE = ',I6,/,&
              /,45X,'SUM OF',16X,'STANDARD',/,&
               8X,'VARIABLE',8X,'SUM',8X,'MEAN',5X,'SQUARES',4X,&
                   'VARIANCE',3X,'DEVIATION',/,&
               7X,70('-'))
 6050 FORMAT(  7X,I6,3X,3(1X,F11.2),2(1X,D11.4))
 6060 FORMAT( /,7X,'PARTIAL CORRELATION COEFFICIENT MATRIX',/)
 6070 FORMAT(  9X,5(1X,D15.8))
 6080 FORMAT(/,/,3X,'*** CONTINUATION PROCESSING ***',/,&
              /,3X,'** INPUT **')
      END
```

(d) Output results

```
*** D2CCPR ***

  *** FIRST PROCESSING ***

  ** INPUT **

      ISW =      0

      N =      7     M =      3

      OBSERVATION MATRIX

              84.00        58.00        42.00
              92.00        88.00        86.00
              88.00        80.00        98.00
              66.00        72.00        64.00
              64.00        50.00        40.00
              80.00        94.00        74.00
              90.00        92.00        94.00

  ** OUTPUT **

      IERR =      0

      TOTAL SAMPLE SIZE =      7

                                             SUM OF                STANDARD
       VARIABLE        SUM        MEAN     SQUARES    VARIANCE   DEVIATION
      ----------------------------------------------------------------------
            1        564.00       80.57      773.71  0.1290D+03  0.1136D+02
            2        534.00       76.29     1755.43  0.2926D+03  0.1710D+02
            3        498.00       71.14     3342.86  0.5571D+03  0.2360D+02

      PARTIAL CORRELATION COEFFICIENT MATRIX

         -0.10000000D+01  0.11966857D+00  0.36608424D+00
          0.11966857D+00 -0.10000000D+01  0.74403866D+00
          0.36608424D+00  0.74403866D+00 -0.10000000D+01


  *** CONTINUATION PROCESSING ***

  ** INPUT **
```

```
    ISW =      1

    N =      3     M =      3

    OBSERVATION MATRIX

            40.00        36.00         8.00
            78.00        50.00        86.00
            82.00        62.00        66.00

** OUTPUT **

    IERR =      0

    TOTAL SAMPLE SIZE =     10

                                          SUM OF                    STANDARD
     VARIABLE        SUM        MEAN     SQUARES    VARIANCE   DEVIATION
    ----------------------------------------------------------------------
          1         764.00       76.40     2254.40  0.2505D+03  0.1583D+02
          2         682.00       68.20     3619.60  0.4022D+03  0.2005D+02
          3         658.00       65.80     7291.60  0.8102D+03  0.2846D+02

    PARTIAL CORRELATION COEFFICIENT MATRIX

       -0.10000000D+01  0.27290340D+00  0.64287069D+00
        0.27290340D+00 -0.10000000D+01  0.37828938D+00
        0.64287069D+00  0.37828938D+00 -0.10000000D+01
```

# Chapter 5

# TIME SERIES ANALYSIS

## 5.1 INTRODUCTION

This library provides the following functions for performing time series analysis.

- Autocovariance and Cross Covariance
- Autocorrelation and Cross Correlation
- Smoothing and Demand Forecasting

### 5.1.1 Explanation

(1) **Basic Statistics of Time Series Data**

Let time series data be represented by $x_1, x_2, \cdots, x_n$. Let the mean of the first and last $n-l$ data values among this time series data be represented by $\mu^{(l)}$ and $\nu^{(l)}$ as follows, where $(l = 0, 1, \cdots, m-1; m \leq n)$.

$$\mu^{(l)} = \frac{\sum_{i=1}^{n-l} x_i}{(n-l)} \quad (l = 0, 1, \cdots, m-1)$$

$$\nu^{(l)} = \frac{\sum_{i=1}^{n-l} x_{i+l}}{(n-l)} \quad (l = 0, 1, \cdots, m-1)$$

The **autocovariance** $c^{(l)}$ at this time is defined as follows.

$$c^{(l)} = \frac{\sum_{i=1}^{n-l}(x_i - \mu^{(l)})(x_{i+l} - \nu^{(l)})}{(n-l)} \quad (l = 0, 1, \cdots, m-1)$$

$c^{(l)}$ can also be expressed as follows.

$$c^{(l)} = \frac{\sum_{j=l+1}^{n} (x_j - \nu^{(l)})(x_{j-l} - \mu^{(l)})}{(n-l)} \quad (l = 0, 1, \cdots, m-1)$$

If the sample variance (not an unbiased estimate) of the first and last $n-l$ data values among this time series data are represented by $u^{(l)}$ and $v^{(l)}$ as follows:

$$u^{(l)} = \frac{\sum_{i=1}^{n-l}(x_i - \mu^{(l)})^2}{(n-l)} \quad (l = 0, 1, \cdots, m-1)$$

$$v^{(l)} = \frac{\sum_{i=1}^{n-l}(x_{i+l} - \nu^{(l)})^2}{(n-l)} \quad (l = 0, 1, \cdots, m-1)$$

the **autocorrelation coefficient** $r^{(l)}$ is defined as follows.

$$\begin{aligned} r^{(l)} &= \frac{c^{(l)}}{\sqrt{u^{(l)} v^{(l)}}} \\ &= \frac{\sum_{i=1}^{n-l}(x_i - \mu^{(l)})(x_{i+l} - \nu^{(l)})}{\sqrt{\sum_{i=1}^{n-l}(x_i - \mu^{(l)})^2}\sqrt{\sum_{i=1}^{n-l}(x_{i+l} - \nu^{(l)})^2}} \end{aligned}$$

The variable $l$ is called the **lag**.

Now, let two sets of time series data be represented by $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_n$. Let the mean of the first and last $n-l$ data values among each of these sets of time series data be represented by ${\mu_x}^{(l)}$ and ${\nu_x}^{(l)}$ for $x_i$ and ${\mu_y}^{(l)}$ and ${\nu_y}^{(l)}$ for $y_i$ $(i = 1, 2, \cdots, n)$ as follows, where $(l = 0, 1, \cdots, m-1; m \leq n)$.

$${\mu_x}^{(l)} = \frac{\sum_{i=1}^{n-l} x_i}{(n-l)} \ (l = 0, 1, \cdots, m-1)$$

$${\nu_x}^{(l)} = \frac{\sum_{i=1}^{n-l} x_{i+l}}{(n-l)} \ (l = 0, 1, \cdots, m-1)$$

$${\mu_y}^{(l)} = \frac{\sum_{i=1}^{n-l} y_i}{(n-l)} \ (l = 0, 1, \cdots, m-1)$$

$${\nu_y}^{(l)} = \frac{\sum_{i=1}^{n-l} y_{i+l}}{(n-l)} \ (l = 0, 1, \cdots, m-1)$$

The **cross covariances** ${c_{xy}}^{(l)}$ and ${c_{yx}}^{(l)}$ at this time are defined as follows.

$${c_{xy}}^{(l)} = \frac{\sum_{i=1}^{n-l} (x_i - {\mu_x}^{(l)})(y_{i+l} - {\nu_y}^{(l)})}{(n-l)} \ (l = 0, 1, \cdots, m-1)$$

$${c_{yx}}^{(l)} = \frac{\sum_{i=1}^{n-l} (y_i - {\mu_y}^{(l)})(x_{i+l} - {\nu_x}^{(l)})}{(n-l)} \ (l = 0, 1, \cdots, m-1)$$

If the sample variances (not unbiased estimates) of the first and last $n-l$ data values among each of these sets of time series data are represented by ${u_x}^{(l)}$ and ${v_x}^{(l)}$ for $x_i$ and ${u_y}^{(l)}$ and ${v_y}^{(l)}$ for $y_i$ $(i = 1, 2, \cdots, n)$ as follows:

$${u_x}^{(l)} = \frac{\sum_{i=1}^{n-l} (x_i - {\mu_x}^{(l)})^2}{(n-l)} \ (l = 0, 1, \cdots, m-1)$$

$${v_x}^{(l)} = \frac{\sum_{i=1}^{n-l} (x_{i+l} - {\nu_x}^{(l)})^2}{(n-l)} \ (l = 0, 1, \cdots, m-1)$$

$${u_y}^{(l)} = \frac{\sum_{i=1}^{n-l} (y_i - {\mu_y}^{(l)})^2}{(n-l)} \ (l = 0, 1, \cdots, m-1)$$

$${v_y}^{(l)} = \frac{\sum_{i=1}^{n-l} (y_{i+l} - {\nu_y}^{(l)})^2}{(n-l)} \ (l = 0, 1, \cdots, m-1)$$

the **cross correlation coefficients** ${r_{xy}}^{(l)}$ and ${r_{yx}}^{(l)}$ are defined as follows.

$$\begin{aligned}
{r_{xy}}^{(l)} &= \frac{{c_{xy}}^{(l)}}{\sqrt{{u_x}^{(l)}{v_y}^{(l)}}} \\
&= \frac{\sum_{i=1}^{n-l}(x_i - {\mu_x}^{(l)})(y_{i+l} - {\nu_y}^{(l)})}{\sqrt{\sum_{i=1}^{n-l}(x_i - {\mu_x}^{(l)})^2}\sqrt{\sum_{i=1}^{n-l}(y_{i+l} - {\nu_y}^{(l)})^2}}
\end{aligned}$$

$$\begin{aligned}
{r_{yx}}^{(l)} &= \frac{{c_{yx}}^{(l)}}{\sqrt{{u_y}^{(l)}{v_x}^{(l)}}} \\
&= \frac{\sum_{i=1}^{n-l}(y_i - {\mu_y}^{(l)})(x_{i+l} - {\nu_x}^{(l)})}{\sqrt{\sum_{i=1}^{n-l}(y_i - {\mu_y}^{(l)})^2}\sqrt{\sum_{i=1}^{n-l}(x_{i+l} - {\nu_x}^{(l)})^2}}
\end{aligned}$$

(2) **Smoothing and Demand Forecasting**

(a) Moving Averages

Let $n$ given time series data values be represented by:

$$x_1, x_2, \cdots, x_n$$

and $m$ specified weights be represented by:

$$w_1, w_2, \cdots, w_m$$

The weighted moving average $M_k^w$ is defined as follows.

$$M_k^w = \sum_{j=1}^{m} \frac{(x_{k+j-1} \cdot w_j)}{\sum_{j=1}^{m} w_j} \ (k = 1, 2, \cdots, n-m+1)$$

$m$ can be called the smoothing bandwidth. Usually, the weight coefficients $w_j$ are determined so that the following relationship is satisfied.

$$\sum_{j=1}^{m} w_j = 1$$

(b) Single Exponential Smoothing

The single exponential smoothing equation for a given time series $\cdots, x_{n-1}, x_n$ ($x_n$ is the most recent data) is defined as follows.

$$S_t = \alpha x_t + (1-\alpha)S_{t-1} \ (t = M, M+1, \cdots, n; M \to -\infty)$$

Here, $S_t$ is the smoothing value of $x_t$ and $\alpha$ is the smoothing constant. Now, assume the forecasting model has the following kind of structure.

$$x_t = a + \varepsilon_t$$

Here, $a$ is a constant, and $\varepsilon_t$ is an error term that independently obeys $N(0, \sigma^2)$. At this time, the mathematical expectation value $E_t$ at time $t$ and the forecast value $E_{t+L}$ at $L$ periods after $t$ are as follows.

$$E_{t+L} = E_t = S_t$$

(c) Double Exponential Smoothing

The double exponential smoothing equation for a given time series $\cdots, x_{n-1}, x_n$ ($x_n$ is the most recent data) is defined as follows.

$$S_t = \alpha x_t + (1 - \alpha)S_{t-1}$$

$$D_t = \alpha S_t + (1 - \alpha)D_{t-1} \ (t = M, M+1, \cdots, n; M \to -\infty)$$

Here, $S_t$ is the single exponential smoothing value of $x_t$, $D_t$ is the double exponential smoothing value of $x_t$, and $\alpha$ is the smoothing constant. Now, assume the forecasting model has the following kind of structure.

$$x_t = a + bt + \varepsilon_t$$

Here, $a$ and $b$ are constants, and $\varepsilon_t$ is an error term that independently obeys $N(0, \sigma^2)$. At this time, the mathematical expectation value $E_t$ at time $t$ and the forecast value $E_{t+L}$ at $L$ periods after $t$ are as follows.

$$E_t = a + bt$$

$$S_t = a + bt - \frac{1-\alpha}{\alpha}b = E_t - \frac{1-\alpha}{\alpha} \cdot b$$

$$D_t = a + bt - \frac{2(1-\alpha)}{\alpha}b = E_t - \frac{2(1-\alpha)}{\alpha} \cdot b$$

Therefore, the following relationships hold:

$$E_t = 2S_t - D_t$$

$$b = B_t = \frac{\alpha}{1-\alpha}(S_t - D_t)$$

Also:

$$E_{t+L} = (2S_t - D_t) + B_t L$$

Here, $B_t$ is called the **linear trend estimate**.

(d) Triple Exponential Smoothing

The triple exponential smoothing equation for a given time series $\cdots, x_{n-1}, x_n$ ($x_n$ is the most recent data) is defined as follows.

$$S_t = \alpha x_t + (1 - \alpha)S_{t-1}$$

$$D_t = \alpha S_t + (1 - \alpha)D_{t-1}$$

$$T_t = \alpha D_t + (1 - \alpha)T_{t-1} \ (t = M, M+1, \cdots, n; M \to -\infty)$$

Here, $S_t$ is the single exponential smoothing value of $x_t$, $D_t$ is the double exponential smoothing value of $x_t$, $T_t$ is the triple exponential smoothing value of $x_t$, and $\alpha$ is the smoothing constant. Now, assume the forecasting model has the following kind of structure.

$$x_t = a + bt + \frac{c}{2}t^2 + \varepsilon_t$$

Here, $a$, $b$, and $c$ are constants, and $\varepsilon_t$ is an error term that independently obeys $N(0, \sigma^2)$. At this time, the mathematical expectation value $E_t$ at time $t$ and the forecast value $E_{t+L}$ at $L$ periods after $t$ are as follows.

$$E_t = a + bt + \frac{c}{2}t^2$$

$$S_t = a + bt + \frac{c}{2}t^2 - (b+ct)\frac{1-\alpha}{\alpha} + \frac{c}{2}\cdot\frac{(1-\alpha)(1+(1-\alpha))}{(\alpha)^2}$$

$$D_t = a + bt + \frac{c}{2}t^2 - (b+ct)\frac{2(1-\alpha)}{\alpha} + \frac{c}{2}\cdot\frac{(1-\alpha)(2+2(1-\alpha))}{(\alpha)^2}$$

$$T_t = a + bt + \frac{c}{2}t^2 - (b+ct)\frac{3(1-\alpha)}{\alpha} + \frac{c}{2}\cdot\frac{(1-\alpha)(3+3(1-\alpha))}{(\alpha)^2}$$

Therefore, the following relationships hold:

$$E_t = 3S_t - 3D_t + T_t$$

$$b = B_t = \frac{\alpha}{2(1-\alpha)^2}\{(6-5\alpha)S_t - 2(5-4\alpha)D_t + (4-3\alpha)T_t\}$$

$$c = C_t = \frac{(\alpha)^2}{(1-\alpha)^2}(S_t - 2D_t + T_t)$$

Also:

$$E_{t+L} = E_t + B_tL + \frac{C_t}{2}L^2$$

Here, $B_t$ is called the **linear trend estimate** and $C_t$ is called the **quadratic trend estimate**.

### 5.1.2 Reference Bibliography

(1) Brigham, E. Oran, “The Fast Fourier Transform”, Prentice-Hall Inc. , (1974).

## 5.2 AUTOCOVARIANCE AND CROSS COVARIANCE

### 5.2.1 DFCVSC, RFCVSC<br>Autocovariances

(1) **Function**

Let time series data be represented by $x_1, x_2, \cdots, x_n$. The DFCVSC or RFCVSC obtains the sum, mean, sum of squares of deviation, and standard deviation (unbiased estimate), which are defined by the following equations, for the first $n-l$ $(l = 1, 2, \cdots, m; m \leq n)$ data values among this time series data.

Sum:

$$s^{(l)} = \sum_{i=1}^{n-l} x_i \ (l = 1, 2, \cdots, m)$$

Mean:

$$\mu^{(l)} = \frac{\sum_{i=1}^{n-l} x_i}{(n-l)} \ (l = 1, 2, \cdots, m)$$

Sum of squares of deviation:

$$v^{(l)} = \sum_{i=1}^{n-l} (x_i - \mu^{(l)})^2 \ (l = 1, 2, \cdots, m)$$

Standard deviation (unbiased estimate):

$$\sigma^{(l)} = \sqrt{\frac{v^{(l)}}{(n-l-1)}}$$

The subroutine also obtains the autocovariance, which is defined by the following equation.

$$c^{(l)} = \frac{\sum_{i=1}^{n-l} (x_i - \mu^{(l)})(x_{i+l} - \nu^{(l)})}{(n-l)} \ (l = 0, 1, \cdots, m-1)$$

Here, $\nu^{(l)}$ is the mean of the last $n-l$ data values, which is defined as follows.

$$\nu^{(l)} = \frac{\sum_{i=1}^{n-l} x_{i+l}}{(n-l)} \ (l = 1, 2, \cdots, m)$$

(2) **Usage**

Double precision:

CALL DFCVSC (A, N, M, V, STAT, IERR)

Single precision:

CALL RFCVSC (A, N, M, V, STAT, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | N | Input | Time series data $x_i$ |
| 2 | N | I | 1 | Input | Number of time series data $n$ |
| 3 | M | I | 1 | Input | Number of autocovariances to be obtained $m$ |
| 4 | V | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | M | Output | Autocovariance $c^{(l)}$ |
| 5 | STAT | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | M,4 | Output | Basic statistics (See Note (a)) |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $\text{N} \geq 2$

(b) $0 < \text{M} \leq \text{N}$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) or (b) was not satisfied. | Processing is aborted. |

(6) **Notes**

(a) The sum $s^{(l)}$, mean $\mu^{(l)}$, sum of squares of deviation $v^{(l)}$, and standard deviation $\sigma^{(l)}$ $(l = 1, 2, \cdots, m)$ are output as follows in array STAT as a real matrix (two-dimensional array type) (See Appendix A).

$$\begin{bmatrix} s^{(1)} & \mu^{(1)} & v^{(1)} & \sigma^{(1)} \\ s^{(2)} & \mu^{(2)} & v^{(2)} & \sigma^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ s^{(m)} & \mu^{(m)} & v^{(m)} & \sigma^{(m)} \end{bmatrix}$$

(7) **Example**

(a) Problem
Obtain autocovariances when the following time series data A is given.
A(1) = 48.1D0, A(13) = 26.1D0
A(2) = 54.7D0, A(14) = 22.3D0
A(3) = 58.0D0, A(15) = 24.1D0
A(4) = 64.2D0, A(16) = 33.1D0
A(5) = 56.1D0, A(17) = 42.3D0
A(6) = 44.9D0, A(18) = 52.5D0

A(7) = 36.1D0, A(19) = 36.0D0
A(8) = 34.2D0, A(20) = 23.5D0
A(9) = 50.1D0, A(21) = 14.9D0
A(10) = 54.9D0, A(22) = 19.8D0
A(11) = 55.1D0, A(23) = 25.0D0
A(12) = 48.4D0, A(24) = 35.4D0

(b) Input data

Time series data A, N=24 and number of autocovariances to be obtained M=14.

(c) Main program

```
      PROGRAM NFCVSC
      PARAMETER (NMAX=24,MMAX=14)
      REAL(8) A(NMAX),STAT(MMAX,4),V(MMAX)
!
      N=24
      M=14
!
      READ(*,*) (A(I),I=1,N)
!
      CALL DFCVSC(A,N,M,V,STAT,IERR)
!
      WRITE(6,110) N,M
      WRITE(6,120) (A(I),I=1,N)
      WRITE(6,130) IERR
      WRITE(6,140) (I,STAT(I,1),STAT(I,2)&
                  ,STAT(I,3),STAT(I,4),V(I),I=1,M)
!
  110 FORMAT(' ',/,6X,'*** AUTOCOVARIANCE (DFCVSC))',&
           ' ***',/,/,6X,'INPUT DATA',6X,'N =',I4,5X,'M =',I4,/)
  120 FORMAT(6X,6F10.3)
  130 FORMAT(6X,'OUTPUT',/,/,6X,'IERR=',I4,/,/,6X,'NO.',1X,'STAT(I,1)',&
         1X,'STAT(I,2)',1X,'STAT(I,3)',1X,'STAT(I,4)',9X,'R',/)
  140 FORMAT(6X,I3,5F10.3)
!
      END
```

(d) Output results

```
     *** AUTOCOVARIANCE (DFCVSC)) ***

     INPUT DATA      N =  24     M =  14

         48.100    54.700    58.000    64.200    56.100    44.900
         36.100    34.200    50.100    54.900    55.100    48.400
         26.100    22.300    24.100    33.100    42.300    52.500
         36.000    23.500    14.900    19.800    25.000    35.400
     OUTPUT

     IERR=   0

     NO. STAT(I,1) STAT(I,2) STAT(I,3) STAT(I,4)         R

       1   959.800    39.992  4686.918    14.275  4686.918
       2   924.400    40.191  4664.918    14.562  3579.078
       3   899.400    40.882  4423.653    14.514  1540.272
       4   879.600    41.886  3958.046    14.068  -232.746
       5   864.700    43.235  3193.405    12.964  -621.458
       6   841.200    44.274  2783.437    12.435   448.051
       7   805.200    44.733  2711.180    12.629  1945.133
       8   752.700    44.276  2647.311    12.863  2522.894
       9   710.400    44.400  2643.160    13.274  1891.180
      10   677.300    45.153  2506.957    13.382   480.529
      11   653.200    46.657  2032.054    12.502  -673.340
      12   630.900    48.531  1393.148    10.775  -878.072
      13   604.800    50.400   848.080     8.781  -136.010
      14   556.400    50.582   843.716     9.185   624.600
```

### 5.2.2 DFCVCS, RFCVCS Cross Covariances

(1) **Function**

Let two sets of time series data be represented as follows.

$$x_1, x_2, \cdots, x_n$$

$$y_1, y_2, \cdots, y_n$$

The DFCVCS or RFCVCS obtains the cross covariances ${c_{xy}}^{(l)}$ and ${c_{yx}}^{(l)}$ $(l = 1, 2, \cdots, m; m \leq n)$, which are defined by the following equations.

$${c_{xy}}^{(l)} = \frac{\sum_{i=1}^{n-l}(x_i - {\mu_x}^{(l)})(y_{i+l} - {\nu_y}^{(l)})}{(n-l)} \quad (l = 1, 2, \cdots, m)$$

$${c_{yx}}^{(l)} = \frac{\sum_{i=1}^{n-l}(y_i - {\mu_y}^{(l)})(x_{i+l} - {\nu_x}^{(l)})}{(n-l)} \quad (l = 1, 2, \cdots, m)$$

Here, ${\mu_x}^{(l)}$, ${\nu_x}^{(l)}$, ${\mu_y}^{(l)}$ and ${\nu_y}^{(l)}$, which represent the means for $x_i$ and $y_i$ $(i = 1, 2, \cdots, n)$ of the first and last $n - l$ data values, respectively, among these sets of time series data where $(l = 1, 2, \cdots, m; m \leq n)$, are defined as follows.

$${\mu_x}^{(l)} = \frac{\sum_{i=1}^{n-l} x_i}{(n-l)} \quad (l = 1, 2, \cdots, m)$$

$${\nu_x}^{(l)} = \frac{\sum_{i=1}^{n-l} x_{i+l}}{(n-l)} \quad (l = 1, 2, \cdots, m)$$

$${\mu_y}^{(l)} = \frac{\sum_{i=1}^{n-l} y_i}{(n-l)} \quad (l = 1, 2, \cdots, m)$$

$${\nu_y}^{(l)} = \frac{\sum_{i=1}^{n-l} y_{i+l}}{(n-l)} \quad (l = 1, 2, \cdots, m)$$

(2) **Usage**

Double precision:

CALL DFCVCS (X, N, Y, M, VX, VY, IERR)

Single precision:

CALL RFCVCS (X, N, Y, M, VX, VY, IERR)

## (3) Arguments

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | X | $\begin{Bmatrix}\text{D}\\\text{R}\end{Bmatrix}$ | N | Input | Time series data $x_i$ |
| 2 | N | I | 1 | Input | Number of time series data $n$ |
| 3 | Y | $\begin{Bmatrix}\text{D}\\\text{R}\end{Bmatrix}$ | N | Input | Time series data $y_i$ |
| 4 | M | I | 1 | Input | Number of cross covariances to be obtained $m$ |
| 5 | VX | $\begin{Bmatrix}\text{D}\\\text{R}\end{Bmatrix}$ | M | Output | Cross covariance ${c_{xy}}^{(l)}$ |
| 6 | VY | $\begin{Bmatrix}\text{D}\\\text{R}\end{Bmatrix}$ | M | Output | Cross covariance ${c_{yx}}^{(l)}$ |
| 7 | IERR | I | 1 | Output | Error indicator |

## (4) Restrictions

(a) $\text{N} \geq 2$

(b) $0 < \text{M} \leq \text{N}$

## (5) Error indicator

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) or (b) was not satisfied. | Processing is aborted. |

## (6) Notes

None

## (7) Example

(a) Problem

Obtain cross covariances when the following two sets of time series data X and Y are given.

X(1) = 2.05D0, Y(1) = 22.23D0
X(2) = 2.10D0, Y(2) = 22.34D0
X(3) = 3.40D0, Y(3) = 22.30D0
X(4) = 5.01D0, Y(4) = 21.90D0
X(5) = 7.35D0, Y(5) = 21.31D0
X(6) = 9.43D0, Y(6) = 20.41D0
X(7) = 10.82D0, Y(7) = 19.43D0
X(8) = 11.09D0, Y(8) = 18.41D0
X(9) = 10.41D0, Y(9) = 17.72D0
X(10) = 7.85D0, Y(10) = 17.81D0

X(11) = 3.97D0, Y(11) = 18.01D0
X(12) = 2.90D0, Y(12) = 18.72D0
X(13) = 2.85D0, Y(13) = 19.51D0
X(14) = 3.50D0, Y(14) = 20.39D0
X(15) = 4.97D0, Y(15) = 21.54D0
X(16) = 6.77D0, Y(16) = 22.33D0
X(17) = 9.61D0, Y(17) = 22.35D0
X(18) = 11.91D0, Y(18) = 21.69D0
X(19) = 12.81D0, Y(19) = 19.97D0
X(20) = 10.92D0, Y(20) = 19.02D0
X(21) = 9.30D0, Y(21) = 18.73D0
X(22) = 6.13D0, Y(22) = 18.91D0
X(23) = 4.54D0, Y(23) = 19.34D0
X(24) = 4.72D0, Y(24) = 19.94D0

(b) Input data
Time series data X and Y, N=24 and number of cross covariances to be obtained M=14.

(c) Main program

```
      PROGRAM NFCVCS
!
      REAL(8) X(24),Y(24),VX(14),VY(14)
!
      N=24
      M=14
!
      READ(*,*) (X(I),I=1,N)
      READ(*,*) (Y(I),I=1,N)
!
      CALL DFCVCS(X,N,Y,M,VX,VY,IERR)
!
      WRITE(6,110) N,M
      WRITE(6,120) (X(I),Y(I),I=1,N)
      WRITE(6,130) IERR
      WRITE(6,140) (I,VX(I),VY(I),I=1,M)
!
  110 FORMAT(' ',/,/,/,6X,'*** CROSS COVARIANCE (DFCVCS) ***',/,/,&
           6X,'INPUT DATA',/,/,6X,'N =',I4,5X,'M =',I4,/,/,&
           15X,'DATA-X',9X,'DATA-Y',/)
  120 FORMAT(6X,2F15.3)
  130 FORMAT(6X,'OUTPUT',/,/,6X,'IERR=',I3,/,/,&
           6X,'NO.',15X,'VX',13X,'VY',/)
  140 FORMAT(6X,I3,2X,2F15.3)
!
      END
```

(d) Output results

```
      *** CROSS COVARIANCE (DFCVCS) ***

      INPUT DATA

      N =  24     M =  14

               DATA-X         DATA-Y

                2.050         22.230
                2.100         22.340
                3.400         22.300
                5.010         21.900
                7.350         21.310
                9.430         20.410
               10.820         19.430
               11.090         18.410
               10.410         17.720
                7.850         17.810
                3.970         18.010
                2.900         18.720
                2.850         19.510
                3.500         20.390
                4.970         21.540
                6.770         22.330
                9.610         22.350
               11.910         21.690
               12.810         19.970
```

```
          10.920          19.020
           9.300          18.730
           6.130          18.910
           4.540          19.340
           4.720          19.940
OUTPUT

IERR=  0

NO.               VX              VY

  1          -32.295         -32.295
  2          -72.567          22.571
  3          -93.644          66.354
  4          -89.968          86.105
  5          -65.973          78.128
  6          -27.864          48.649
  7           13.207           9.811
  8           47.778         -26.477
  9           67.510         -54.681
 10           67.428         -70.349
 11           50.022         -69.464
 12           21.173         -50.006
 13           -9.509         -22.386
 14          -33.709           5.315
```

## 5.3 AUTOCORRELATION AND CROSS CORRELATION

### 5.3.1 DFCRSC, RFCRSC
### Autocorrelation Coefficients

(1) **Function**

Let time series data be represented by $x_1, x_2, \cdots, x_n$. The DFCRSC or RFCRSC obtains the sum, mean, sum of squares of deviation, and standard deviation (unbiased estimate), which are defined by the following equations, for the first $n-l$ $(l = 1, 2, \cdots, m; m \leq n)$ data values among this time series data.

Sum:

$$s^{(l)} = \sum_{i=1}^{n-l} x_i \ (l = 1, 2, \cdots, m)$$

Mean:

$$\mu^{(l)} = \frac{\sum_{i=1}^{n-l} x_i}{(n-l)} \ (l = 1, 2, \cdots, m)$$

Sum of squares of deviation:

$$v^{(l)} = \sum_{i=1}^{n-l} (x_i - \mu^{(l)})^2 \ (l = 1, 2, \cdots, m)$$

Standard deviation (unbiased estimate):

$$\sigma^{(l)} = \sqrt{\frac{v^{(l)}}{(n-l-1)}}$$

The subroutine also obtains the autocorrelation coefficient, which is defined by the following equation.

$$r^{(l)} = \frac{\sum_{i=1}^{n-l} (x_i - \mu^{(l)})(x_{i+l} - \nu^{(l)})}{\sqrt{\sum_{i=1}^{n-l} (x_i - \mu^{(l)})^2} \sqrt{\sum_{i=1}^{n-l} (x_{i+l} - \nu^{(l)})^2}}$$

Here, $\nu^{(l)}$ is the mean of the last $n-l$ data values, which is defined as follows.

$$\nu^{(l)} = \frac{\sum_{i=1}^{n-l} x_{i+l}}{(n-l)} \ (l = 1, 2, \cdots, m)$$

(2) **Usage**

Double precision:

CALL DFCRSC (A, N, M, R, STAT, IERR)

Single precision:

CALL RFCRSC (A, N, M, R, STAT, IERR)

### (3) Arguments

D:Double precision real    Z:Double precision complex    I: $\left\{ \begin{array}{l} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{array} \right\}$
R:Single precision real    C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | N | Input | Time series data $x_i$ |
| 2 | N | I | 1 | Input | Number of time series data $n$ |
| 3 | M | I | 1 | Input | Number of autocorrelation coefficients to be obtained $m$ |
| 4 | R | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | M | Output | Autocorrelation coefficient $r^{(l)}$ |
| 5 | STAT | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | M,4 | Output | Basic statistics (See Note (a)) |
| 6 | IERR | I | 1 | Output | Error indicator |

### (4) Restrictions

(a) N $\geq 2$

(b) $1 \leq$ M $\leq$N

### (5) Error indicator

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | For some $l$, the standard deviation $\sigma^{(l)}$ or correlation coefficient $r^{(l)}$ was not obtained. | 0.0 is set for the portion corresponding to $\sigma^{(l)}$ or $r^{(l)}$. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3100 | Restriction (b) was not satisfied. | |

### (6) Notes

(a) The sum $s^{(l)}$, mean $\mu^{(l)}$, sum of squares of deviation $v^{(l)}$, and standard deviation $\sigma^{(l)}$ $(l = 1, 2, \cdots, m)$ are output as follows in array STAT as a real matrix (two-dimensional array type) (See Appendix A).

$$\begin{bmatrix} s^{(1)} & \mu^{(1)} & v^{(1)} & \sigma^{(1)} \\ s^{(2)} & \mu^{(2)} & v^{(2)} & \sigma^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ s^{(m)} & \mu^{(m)} & v^{(m)} & \sigma^{(m)} \end{bmatrix}$$

(b) When IERR=1000, the sum $s^{(l)}$, mean $\mu^{(l)}$, and sum of squares of deviation $v^{(l)}$ are obtained, and 0.0 is set in the positions where the standard deviation $\sigma^{(l)}$ or correlation coefficient $r^{(l)}$ was not obtained $(l = 1, 2, \cdots, m)$.

## 5.3.2 DFCRCZ, RFCRCZ Cross Correlation Coefficients (Zero Means)

(1) **Function**

Given the following two sets of time series data for which the means are zero:

$$x_1, x_2, \cdots, x_n$$

$$y_1, y_2, \cdots, y_n$$

the DFCRCZ or RFCRCZ obtains an approximation ${r_{xy}}^{(l)}$ $(l = 1, 2, \cdots, m)$ of the cross correlation coefficient, which is defined by the following equation.

$${r_{xy}}^{(l)} = \frac{n}{n-l} \frac{\sum_{i=1}^{n-l} x_i y_{i+l}}{\sqrt{\sum_{i=1}^{n} (x_i)^2} \sqrt{\sum_{i=1}^{n} (y_i)^2}} \quad (l = 1, 2, \cdots, m; n \gg m)$$

(2) **Usage**

Double precision:
CALL DFCRCZ (X, N, Y, M, CRR, IERR)
Single precision:
CALL RFCRCZ (X, N, Y, M, CRR, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: $\left\{ \begin{array}{l} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{array} \right\}$

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | X | $\left\{\begin{array}{c} D \\ R \end{array}\right\}$ | N | Input | Time series data $x_i$ |
| 2 | N | I | 1 | Input | Number observed values $n$ |
| 3 | Y | $\left\{\begin{array}{c} D \\ R \end{array}\right\}$ | N | Input | Time series data $y_i$ |
| 4 | M | I | 1 | Input | Number of cross correlation coefficients to be obtained $m$ |
| 5 | CRR | $\left\{\begin{array}{c} D \\ R \end{array}\right\}$ | M | Output | Cross correlation coefficient approximation ${r_{xy}}^{(l)}$ |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $1 \leq \mathrm{M} \leq \mathrm{N}$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | The sum of squares of $x_i$ or sum of squares of $y_i$ was smaller than the unit for determining error. | ${r_{xy}}^{(l)} = 0.0$ was set for all $l$. |
| 1100 | For some $l$ the cross correlation coefficient ${r_{xy}}^{(l)}$ was larger than 1. | Processing continues. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |

(6) **Notes**

(a) The subroutine 5.3.3 $\begin{Bmatrix}\text{DFCRCS}\\\text{RFCRCS}\end{Bmatrix}$ must be used for time series data for which the mean is not zero.

(b) When IERR=1100, the correlation coefficients ${r_{xy}}^{(l)}$ are all calculated. However, since the correlation coefficient definition is not satisfied for those having a value greater than or equal to 1, they cannot be used as correlation coefficients.

### 5.3.3 DFCRCS, RFCRCS Cross Correlation Coefficients

(1) **Function**

Let two sets of time series data be represented as follows.

$$x_1, x_2, \cdots, x_n$$

$$y_1, y_2, \cdots, y_n$$

The DFCRCS or RFCRCS obtains the cross correlation coefficients ${r_{xy}}^{(l)}$ and ${r_{yx}}^{(l)}$ $(l = 1, 2, \cdots, m; m \leq n)$, which are defined by the following equations.

$${r_{xy}}^{(l)} = \frac{\sum_{i=1}^{n-l}(x_i - {\mu_x}^{(l)})(y_{i+l} - {\nu_y}^{(l)})}{\sqrt{\sum_{i=1}^{n-l}(x_i - {\mu_x}^{(l)})^2}\sqrt{\sum_{i=1}^{n-l}(y_{i+l} - {\nu_y}^{(l)})^2}} \quad (l = 1, 2, \cdots, m)$$

$${r_{yx}}^{(l)} = \frac{\sum_{i=1}^{n-l}(y_i - {\mu_y}^{(l)})(x_{i+l} - {\nu_x}^{(l)})}{\sqrt{\sum_{i=1}^{n-l}(y_i - {\mu_y}^{(l)})^2}\sqrt{\sum_{i=1}^{n-l}(x_{i+l} - {\nu_x}^{(l)})^2}} \quad (l = 1, 2, \cdots, m)$$

Here, ${\mu_x}^{(l)}$, ${\nu_x}^{(l)}$, ${\mu_y}^{(l)}$, and ${\nu_y}^{(l)}$, which represent the means for $x_i$ and $y_i$ $(i = 1, 2, \cdots, n)$ of the first and last $n - l$ data values, respectively, among these sets of time series data where $(l = 1, 2, \cdots, m; m \leq n)$, are defined as follows.

$${\mu_x}^{(l)} = \frac{\sum_{i=1}^{n-l} x_i}{(n-l)} \quad (l = 1, 2, \cdots, m)$$

$${\nu_x}^{(l)} = \frac{\sum_{i=1}^{n-l} x_{i+l}}{(n-l)} \quad (l = 1, 2, \cdots, m)$$

$${\mu_y}^{(l)} = \frac{\sum_{i=1}^{n-l} y_i}{(n-l)} \quad (l = 1, 2, \cdots, m)$$

$${\nu_y}^{(l)} = \frac{\sum_{i=1}^{n-l} y_{i+l}}{(n-l)} \quad (l = 1, 2, \cdots, m)$$

(2) **Usage**

Double precision:

CALL DFCRCS (X, N, Y, M, RX, RY, IERR)

Single precision:

CALL RFCRCS (X, N, Y, M, RX, RY, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: {INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer}

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | X | {D / R} | N | Input | Time series data $x_i$ |
| 2 | N | I | 1 | Input | Number of time series data observed values $n$ |
| 3 | Y | {D / R} | N | Input | Time series data $y_i$ |
| 4 | M | I | 1 | Input | Number of cross correlation coefficients to be obtained $m$ |
| 5 | RX | {D / R} | M | Output | Cross correlation coefficient ${r_{xy}}^{(l)}$ |
| 6 | RY | {D / R} | M | Output | Cross correlation coefficient ${r_{yx}}^{(l)}$ |
| 7 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $\mathrm{N} \geq 2$

(b) $1 \leq \mathrm{M} \leq \mathrm{N}$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | For some $l$, the correlation coefficient ${r_{xy}}^{(l)}$ or ${r_{yx}}^{(l)}$ was not obtained. | 0.0 is set for the value of the correlation coefficient that was not obtained. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3100 | Restriction (b) was not satisfied. | |

(6) **Notes**

None

## 5.4 SMOOTHING AND DEMAND FORECASTING

### 5.4.1 DFASMA, RFASMA
### Moving Averages

(1) **Function**

Using $n$ given time series data values represented as follows:

$$x_1, x_2, \cdots, x_n$$

and $m$ specified weights represented as follows:

$$w_1, w_2, \cdots, w_m$$

the DFASMA or RFASMA obtains the weighted moving average $M_k^w$. The weighted moving average $M_k^w$ is defined as follows.

$$M_k^w = \sum_{j=1}^{m} \frac{(x_{k+j-1} \cdot w_j)}{\sum_{j=1}^{m} w_j} \ (k = 1, 2, \cdots, n - m + 1)$$

$m$ can be called the smoothing bandwidth.

(2) **Usage**

Double precision:

CALL DFASMA (A, N, M, WA, AV, ISW, IERR)

Single precision:

CALL RFASMA (A, N, M, WA, AV, ISW, IERR)

## (3) Arguments

D:Double precision real　Z:Double precision complex　I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real　C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{matrix}D\\R\end{matrix}\right\}$ | N | Input | Time series data $x_i$ |
| 2 | N | I | 1 | Input | Number of time series data $n$ |
| 3 | M | I | 1 | Input | Smoothing bandwidth $m$ |
| 4 | WA | $\left\{\begin{matrix}D\\R\end{matrix}\right\}$ | M | Input | Weight $w_j$ |
| 5 | AV | $\left\{\begin{matrix}D\\R\end{matrix}\right\}$ | $\mathrm{N}-\mathrm{M}+1$ | Output | Moving average $M_k^w$ |
| 6 | ISW | I | 1 | Input | Weight specification switch<br>ISW=0: Input weight $w_j$<br>ISW=1: Set all weights to 1.0 |
| 7 | IERR | I | 1 | Output | Error indicator |

## (4) Restrictions

(a) $\mathrm{N} \geq 2$

(b) $0 < \mathrm{M} \leq \mathrm{N}$

(c) $\mathrm{WA}(1) + \cdots + \mathrm{WA}(\mathrm{M}) > 0$

(d) $\mathrm{ISW} \in \{0, 1\}$

## (5) Error indicator

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (d) was not satisfied. | All weights are set to 1.0 and processing continues. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |

## (6) Notes

None

## (7) Example

(a) Problem
Obtain the moving averages for the following time series data:

| | | | |
|---|---|---|---|
| A( 1) = | 71.8 | A(13) = | 62.3 |
| A( 2) = | 73.2 | A(14) = | 51.2 |
| A( 3) = | 63.8 | A(15) = | 48.2 |
| A( 4) = | 60.0 | A(16) = | 29.8 |
| A( 5) = | 58.3 | A(17) = | 34.3 |
| A( 6) = | 57.2 | A(18) = | 24.0 |
| A( 7) = | 48.5 | A(19) = | 22.9 |
| A( 8) = | 53.5 | A(20) = | 31.8 |
| A( 9) = | 69.3 | A(21) = | 70.0 |
| A(10) = | 68.7 | A(22) = | 106.7 |
| A(11) = | 73.4 | A(23) = | 138.5 |
| A(12) = | 74.9 | A(24) = | 146.1 |

to which the following weights are assigned:

| | |
|---|---|
| WA( 1) = | −3.0 |
| WA( 2) = | 12.0 |
| WA( 3) = | 17.0 |
| WA( 4) = | 12.0 |
| WA( 5) = | −3.0 |

(b) Input data

Array A and WA, N=24, M=5 and ISW=0.

(c) Main program

```
      PROGRAM BFASMA
! *** EXAMPLE OF DFASMA ***
      IMPLICIT REAL(8) (A-H,O-Z)
      PARAMETER(N=24,M=5,ISW=0)
      DIMENSION A(N),WA(M),AV(N-M+1)
!
      READ(5,*) (A(I),I=1,N)
      READ(5,*) (WA(I),I=1,M)
      WRITE(6,1000) N,M,ISW
      WRITE(6,2000)
      DO 10 I=1,N
        WRITE(6,2100) I,A(I)
   10 CONTINUE
      WRITE(6,3000)
      DO 20 I=1,M
        WRITE(6,3100) I,WA(I)
   20 CONTINUE
      CALL DFASMA(A,N,M,WA,AV,ISW,IERR)
      WRITE(6,4000) IERR
      DO 30 I=1,N-M+1
        WRITE(6,5000) I,AV(I)
   30 CONTINUE
      STOP
!
 1000 FORMAT(1X,' *** DFASMA ***',/,/,'   ** INPUT **',/,/,7X,'N   = ',I4,&
             /,7X,'M   = ',I4,/,7X,'ISW = ',I4)
 2000 FORMAT(1X,/,7X,'TIME SERIES DATA',/)
 2100 FORMAT(8X,'A(',I2,')  = ',D18.10)
 3000 FORMAT(1X,/,7X,'WEIGHT',/)
 3100 FORMAT(8X,'WA(',I2,') = ',D18.10)
 4000 FORMAT(1X,/,1X,'  ** OUTPUT **',/,/,7X,'IERR = ',I4,/)
 5000 FORMAT(7X,'AV(',I2,') = ',D18.10)
!
      END
```

(d) Output results

```
 *** DFASMA ***

  ** INPUT **

      N   =   24
      M   =    5
      ISW =    0

      TIME SERIES DATA

       A( 1)  =   0.7180000000D+02
       A( 2)  =   0.7320000000D+02
       A( 3)  =   0.6380000000D+02
       A( 4)  =   0.6000000000D+02
```

```
    A( 5)  =   0.5830000000D+02
    A( 6)  =   0.5720000000D+02
    A( 7)  =   0.4850000000D+02
    A( 8)  =   0.5350000000D+02
    A( 9)  =   0.6930000000D+02
    A(10)  =   0.6870000000D+02
    A(11)  =   0.7340000000D+02
    A(12)  =   0.7490000000D+02
    A(13)  =   0.6230000000D+02
    A(14)  =   0.5120000000D+02
    A(15)  =   0.4820000000D+02
    A(16)  =   0.2980000000D+02
    A(17)  =   0.3430000000D+02
    A(18)  =   0.2400000000D+02
    A(19)  =   0.2290000000D+02
    A(20)  =   0.3180000000D+02
    A(21)  =   0.7000000000D+02
    A(22)  =   0.1067000000D+03
    A(23)  =   0.1385000000D+03
    A(24)  =   0.1461000000D+03

   WEIGHT

    WA( 1) =  -0.3000000000D+01
    WA( 2) =   0.1200000000D+02
    WA( 3) =   0.1700000000D+02
    WA( 4) =   0.1200000000D+02
    WA( 5) =  -0.3000000000D+01

** OUTPUT **

   IERR =    0

   AV( 1) =   0.6550571429D+02
   AV( 2) =   0.5982857143D+02
   AV( 3) =   0.5887428571D+02
   AV( 4) =   0.5467142857D+02
   AV( 5) =   0.5057428571D+02
   AV( 6) =   0.5558285714D+02
   AV( 7) =   0.6510857143D+02
   AV( 8) =   0.7128857143D+02
   AV( 9) =   0.7360571429D+02
   AV(10) =   0.7262857143D+02
   AV(11) =   0.6307142857D+02
   AV(12) =   0.5378000000D+02
   AV(13) =   0.4290285714D+02
   AV(14) =   0.3631428571D+02
   AV(15) =   0.2901142857D+02
   AV(16) =   0.2598857143D+02
   AV(17) =   0.2131428571D+02
   AV(18) =   0.3609428571D+02
   AV(19) =   0.6765142857D+02
   AV(20) =   0.1080628571D+03
```

### 5.4.2 DFDPES, RFDPES Single Exponential Smoothing

(1) **Function**

The single exponential smoothing equation for a given time series $\cdots, x_{n-1}, x_n$ ($x_n$ is the most recent data) is defined as follows.

$$S_t = \alpha x_t + (1-\alpha)S_{t-1} \ (t = M, M+1, \cdots, n; M \to -\infty)$$

Here, $S_t$ is the smoothing value of $x_t$ and $\alpha$ is the smoothing constant. Now, assume the forecasting model has the following kind of structure.

$$x_t = a + \varepsilon_t$$

Here, $a$ is a constant, and $\varepsilon_t$ is an error term that independently obeys $N(0, \sigma^2)$. At this time, the mathematical expectation value $E_t$ at time $t$ and the forecast value $E_{t+L}$ at $L$ periods after $t$ are as follows.

$$E_{t+L} = E_t = S_t$$

The DFDPES or RFDPES obtains the smoothing value $S_j$ related to the following given time series data:

$$x_1, x_2, \cdots, x_n$$

(where, $x_n$ is the most recent data). The definition of $S_j$ is as follows.

$$S_1 = \frac{1}{m}\sum_{i=1}^{m} x_i \ \text{(initial value)}$$

$$S_j = \alpha x_{j+m-1} + (1-\alpha)S_{j-1} \ (j = 2, 3, \cdots, n-m+1)$$

(2) **Usage**

Double precision:
CALL DFDPES (A, N, ALH, IN, EV, IERR)
Single precision:
CALL RFDPES (A, N, ALH, IN, EV, IERR)

### (3) Arguments

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | N | Input | Time series data $x_j$ |
| 2 | N | I | 1 | Input | Number of time series data $n$ |
| 3 | ALH | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Smoothing constant $\alpha$ |
| 4 | IN | I | 1 | Input | Average number of terms for the initial value setting $m$<br>Default value: 2 (when IN=0) |
| 5 | EV | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | $\mathrm{N}-\mathrm{IN}+1$ | Output | Smoothing value $S_j$ |
| 6 | IERR | I | 1 | Output | Error indicator |

### (4) Restrictions

(a) $\mathrm{N} > 0$

(b) $0 \leq \mathrm{IN} \leq \mathrm{N}$

(c) $0.0 < \mathrm{ALH} < 1.0$

### (5) Error indicator

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | $\mathrm{N} = \mathrm{IN}$ (when $\mathrm{N} \neq 0$) | Only EV(1) is calculated. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3100 | Restriction (b) was not satisfied. | |
| 3200 | Restriction (c) was not satisfied. | |

### (6) Notes

None

### 5.4.3 DFDPED, RFDPED Double Exponential Smoothing

(1) **Function**

The double exponential smoothing equation for a given time series $\cdots, x_{n-1}, x_n$ ($x_n$ is the most recent data) is defined as follows.

$$S_t = \alpha x_t + (1-\alpha)S_{t-1}$$

$$D_t = \alpha S_t + (1-\alpha)D_{t-1} \ (t = M, M+1, \cdots, n; M \to -\infty)$$

Here, $S_t$ is the single exponential smoothing value of $x_t$, $D_t$ is the double exponential smoothing value of $x_t$, and $\alpha$ is the smoothing constant. Now, assume the forecasting model has the following kind of structure.

$$x_t = a + bt + \varepsilon_t$$

Here, $a$ and $b$ are constants, and $\varepsilon_t$ is an error term that independently obeys $N(0, \sigma^2)$. At this time, the mathematical expectation value $E_t$ at time $t$ and the forecast value $E_{t+L}$ at $L$ periods after $t$ are as follows.

$$E_t = 2S_t - D_t$$

$$E_{t+L} = (2S_t - D_t) + B_t \cdot L$$

Here, $B_t$, which represents the linear trend estimate, is defined by the following equation.

$$B_t = \frac{\alpha}{1-\alpha}(S_t - D_t)$$

The DFDPED or RFDPED obtains the smoothing value $E_k$, forecast value $E_{k+L}$, and linear trend estimate $B_k$ related to the following given time series data:

$$x_1, x_2, \cdots, x_n$$

(where, $x_n$ is the most recent data). The definitions of these quantities are as follows.

$$S_1 = \frac{1}{m}\sum_{i=1}^{m} x_i \text{ (initial value)}$$

$$D_1 = S_1 \text{ (initial value)}$$

$$S_j = \alpha x_{j+m-1} + (1-\alpha)S_{j-1}$$

$$D_j = \alpha S_j + (1-\alpha)D_{j-1} \ (j = 2, 3, \cdots, n-m+1)$$

$$E_k = 2S_k - D_k$$

$$B_k = \frac{\alpha}{1-\alpha}(S_k - D_k)$$

$$E_{k+L} = E_k + B_k L \ (k = 1, 2, \cdots, n-m+1)$$

(2) **Usage**

Double precision:

CALL DFDPED (A, N, ALH, IN, M, EV, AV, TR, IERR)

Single precision:

CALL RFDPED (A, N, ALH, IN, M, EV, AV, TR, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | N | Input | Time series data $x_j$ |
| 2 | N | I | 1 | Input | Number of time series data $n$ |
| 3 | ALH | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Smoothing constant $\alpha$ |
| 4 | IN | I | 1 | Input | Average number of terms for the initial value setting $m$<br>Default value: 2 (when IN=0) |
| 5 | M | I | 1 | Input | Time lag $L$ |
| 6 | EV | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | $\text{N} - \text{IN} + 1$ | Output | Smoothing value $E_k$ |
| 7 | AV | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | $\text{N} - \text{IN} + 1$ | Output | Forecast value $E_{k+L}$ |
| 8 | TR | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | $\text{N} - \text{IN} + 1$ | Output | Linear trend estimate $B_k$ |
| 9 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $\text{N} > 0$

(b) $0 \leq \text{IN} \leq \text{N}$

(c) $0.0 < \text{ALH} < 1.0$

(d) $\text{M} \geq 0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | N = IN | EV(1) and AV(1) are calculated, and 0.0 is set for TR(1). |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3100 | Restriction (b) was not satisfied. | |
| 3200 | Restriction (c) was not satisfied. | |
| 3300 | Restriction (d) was not satisfied. | |

(6) **Notes**

(a) When a model is forecasted having this structure pattern, at least two data values are required. A forecast performed using data for one point is statistically meaningless.

### 5.4.4 DFDPET, RFDPET
Triple Exponential Smoothing

(1) **Function**

The triple exponential smoothing equation for a given time series $\cdots, x_{n-1}, x_n$ ($x_n$ is the most recent data) is defined as follows.

$$S_t = \alpha x_t + (1-\alpha)S_{t-1}$$

$$D_t = \alpha S_t + (1-\alpha)D_{t-1}$$

$$T_t = \alpha D_t + (1-\alpha)T_{t-1} \ (t = M, M+1, \cdots, n; M \to -\infty)$$

Here, $S_t$ is the single exponential smoothing value of $x_t$, $D_t$ is the double exponential smoothing value of $x_t$, $T_t$ is the triple exponential smoothing value of $x_t$, and $\alpha$ is the smoothing constant. Now, assume the forecasting model has the following kind of structure.

$$x_t = a + bt + \frac{c}{2}t^2 + \varepsilon_t$$

Here, $a$, $b$ and $c$ are constants, and $\varepsilon_t$ is an error term that independently obeys $N(0, \sigma^2)$. At this time, the mathematical expectation value $E_t$ at time $t$ and the forecast value $E_{t+L}$ at $L$ periods after $t$ are as follows.

$$E_t = 3S_t - 3D_t + T_t$$

$$E_{t+L} = E_t + B_t L + \frac{C_t}{2}L^2$$

Here, $B_t$, which represents the linear trend estimate, and $C_t$, which represents the quadratic trend estimate, are defined by the following equations.

$$B_t = \frac{\alpha}{2(1-\alpha)^2}\{(6-5\alpha)S_t - 2(5-4\alpha)D_t + (4-3\alpha)T_t\}$$

$$C_t = \frac{(\alpha)^2}{(1-\alpha)^2}(S_t - 2D_t + T_t)$$

The DFDPET or RFDPET obtains the smoothing value $E_k$, forecast value $E_{k+L}$, linear trend estimate $B_k$, and quadratic trend estimate $C_k$ related to the following given time series data:

$$x_1, x_2, \cdots, x_n$$

(where, $x_n$ is the most recent data). The definitions of these quantities are as follows.

$$S_1 = \frac{1}{m}\sum_{i=1}^{m} x_i \ \text{(initial value)}$$

$$D_1 = S_1 = T_1 \ \text{(initial value)}$$

$$S_j = \alpha x_{j+m-1} + (1-\alpha)S_{j-1}$$

$$D_j = \alpha S_j + (1-\alpha)D_{j-1}$$

$$T_j = \alpha D_j + (1-\alpha)T_{j-1} \ (j = 2, 3, \cdots, n-m+1)$$

$$E_k = 3S_k - 3D_k + T_k$$

$$B_k = \frac{\alpha}{2(1-\alpha)^2}\{(6-5\alpha)S_k - 2(5-4\alpha)D_k + (4-3\alpha)T_k\}$$

$$C_k = \frac{(\alpha)^2}{(1-\alpha)^2}(S_k - 2D_k + T_k)$$

$$E_{k+L} = E_k + B_k L + \frac{C_k}{2}L^2 \ (k = 1, 2, \cdots, n-m+1)$$

(2) **Usage**

Double precision:

CALL DFDPET (A, N, ALH, IN, M, EV, AV, TR, QR, IERR)

Single precision:

CALL RFDPET (A, N, ALH, IN, M, EV, AV, TR, QR, IERR)

(3) **Arguments**

| | | |
|---|---|---|
| D:Double precision real | Z:Double precision complex | I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$ |
| R:Single precision real | C:Single precision complex | |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | N | Input | Time series data $x_j$ |
| 2 | N | I | 1 | Input | Number of time series data $n$ |
| 3 | ALH | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Smoothing constant $\alpha$ |
| 4 | IN | I | 1 | Input | Average number of terms for the initial value setting<br>Default value: 2 (When IN=0) |
| 5 | M | I | 1 | Input | Time lag $L$ |
| 6 | EV | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | $\text{N} - \text{IN} + 1$ | Output | Smoothing value $E_k$ |
| 7 | AV | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | $\text{N} - \text{IN} + 1$ | Output | Forecast value $E_{k+L}$ |
| 8 | TR | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | $\text{N} - \text{IN} + 1$ | Output | Linear trend estimate $B_k$ |
| 9 | QR | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | $\text{N} - \text{IN} + 1$ | Output | Quadratic trend estimate $C_k$ |
| 10 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $N > 0$

(b) $0 \leq IN \leq N$

(c) $0.0 < ALH < 1.0$

(d) $M \geq 0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | N = IN | EV(1) and AV(1) are calculated, and 0.0 is set for QR(1) and TR(1). |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3100 | Restriction (b) was not satisfied. | |
| 3200 | Restriction (c) was not satisfied. | |
| 3300 | Restriction (d) was not satisfied. | |

(6) **Notes**

(a) When a model is forecasted having this structure pattern, at least three data values are required. A forecast performed using data for two or fewer points is statistically meaningless.

# Chapter 6

# TESTS AND ESTIMATES

## 6.1 INTRODUCTION

A random sample of size $n$ sampled (using sampling with replacement) from a population is theoretically represented as follows by a set of $n$ random variables having an equal probability distribution since they are mutually independent.

$$S = (X_1, X_2, \cdots, X_n)$$

$S$ is called a sample of size $n$, and the common probability distribution of the $X_i$ is called the population distribution. The constants that define the population distribution are called parameters, and a parameter for which the value is unknown is called an unknown parameter. Statistical analysis deals with estimates of the values of these kinds of unknown parameters (**statistical estimates**) and decisions on whether or not hypotheses related to the values of unknown parameters can be established (**statistical hypothesis testing**). This library provides functions for calculating the following kinds of estimates and tests.

- Interval estimations
  - Interval estimation of the population ratio according to one set of samples
  - Interval estimation of the population mean according to one set of samples
  - Interval estimation of the difference of the population means according to two sets of independent samples
  - Interval estimation of the population variance due to one set of samples
  - Interval estimation of the population correlation coefficient according to one set of samples
  - Interval estimation of the difference of the population correlation coefficient according to two sets of independent samples
  - Interval estimation in the simple linear regression
- Tests
  - Test of the population ratio according to one set of samples
  - Test of the difference of the population ratios according to two sets of independent samples
  - Test of the population mean according to one set of samples
  - Test of the difference of the population means according to two sets of independent samples
  - Test of the population variance due to one set of samples
  - Test of the population correlation coefficient according to one set of samples
  - Test of the difference of the population correlation coefficients according to two sets of independent samples
  - Test in the simple linear regression

### 6.1.1 Explanation

(1) **Interval estimation of the population ratio according to one set of samples**

When the number of data having the observed characteristic in the one set of sample data of size $n$ is $m$, obtain the confidence interval of the population ratio when the confidence level $1-\alpha$ is specified. The confidence interval $(p_1, p_2)$ is defined as follows.

$$p_1 = \frac{m}{(n-m+1)F_1 + m}$$

$$p_2 = \frac{(m+1)F_2}{(m+1)F_2 + (n-m)}$$

Here,

$$\frac{\alpha}{2} = 1 - P(F_1|2(n-m+1), 2m) = 1 - P(F_2|2(m+1), 2(n-m))$$

$P(F|n_1, n_2)$ is the cumulative distribution function (c.d.f.) of a $F$ distribution having numbers of degrees of freedom $n_1$ and $n_2$. If the sample ratio $\hat{p}$ is given as $\hat{p} = \frac{m}{n}$,

- when $\hat{p} = 0$:
  the upper bound of one-sided confidence interval in the confidence level $1-\alpha$ is given as follows
  $$p_2 = \frac{F_\alpha}{n + F_\alpha}$$
  where
  $$\alpha = 1 - P(F_\alpha|2, 2n)$$
  $P(F|n_1, n_2)$ is the cumulative distribution function (c.d.f.) of a $F$ distribution having numbers of degrees of freedom $n_1$ and $n_2$.
- when $\hat{p} = 1$:
  the lower bound of one-sided confidence interval in the confidence level $1-\alpha$ is given as follows
  $$p_1 = \frac{n}{n + F_\alpha}$$
  where
  $$\alpha = 1 - P(F_\alpha|2, 2n)$$

(2) **Interval estimation of the population mean according to one set of samples**

From the mean $\mu$ and variance (or population variance) $\sigma^2$ of one set of sample data of size $n$, obtain the confidence interval of the population mean when the confidence level $1-\alpha$ is specified. The confidence interval $(t_1, t_2)$ is defined as follows.

$$t_1 = \mu - z_{\frac{\alpha}{2}}\sqrt{\beta}$$

$$t_2 = \mu + z_{\frac{\alpha}{2}}\sqrt{\beta}$$

(a) When the population variance is known

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

Here, $P(x)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

$$\beta = \frac{\sigma^2}{n}$$

$\sigma^2$: Population variance

(b) When the population variance is unknown

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}}, n-1)$$

Here, $P(t, n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

$$\beta = \frac{\sigma^2}{n}$$

$\sigma^2$ is an unbiased estimate of the population variance.

(3) **Interval estimation of the difference of the population means according to two sets of independent samples**

From the means $\mu_1$ and $\mu_2$ and variances (or population variances) ${\sigma_1}^2$ and ${\sigma_2}^2$ of two sets of independent sample data of sizes $n_1$ and $n_2$, respectively, obtain the confidence interval of the difference of the population means when the confidence level $1-\alpha$ is specified. The confidence interval $(t_1, t_2)$ is defined as follows.

$$t_1 = (\mu_1 - \mu_2) - z_{\frac{\alpha}{2}}\sqrt{\beta_1 + \beta_2}$$

$$t_2 = (\mu_1 - \mu_2) + z_{\frac{\alpha}{2}}\sqrt{\beta_1 + \beta_2}$$

(a) When the population variances are known

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

Here, $P(x)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

$$\beta_1 = \frac{{\sigma_1}^2}{n_1}, \beta_2 = \frac{{\sigma_2}^2}{n_2}$$

${\sigma_1}^2$, ${\sigma_2}^2$: Population variances of the two sets

(b) When the population variances of the two sets are equal and that value is unknown

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}}, n_1 + n_2 - 2)$$

Here, $P(t, n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

$$\beta_1 = \frac{{s_p}^2}{n_1}, \beta_2 = \frac{{s_p}^2}{n_2}$$

${s_p}^2$ is defined as follows.

$${s_p}^2 = \frac{(n_1 - 1){\sigma_1}^2 + (n_2 - 1){\sigma_2}^2}{n_1 + n_2 - 2}$$

${\sigma_1}^2$ and ${\sigma_2}^2$ are unbiased estimates of the population variances.

(c) When the population variances of the two sets are not equal and those values are unknown

$$z_{\frac{\alpha}{2}} = \frac{\beta_1 t^{(1)}_{\frac{\alpha}{2}} + \beta_2 t^{(2)}_{\frac{\alpha}{2}}}{\beta_1 + \beta_2}$$

$$\frac{\alpha}{2} = 1 - P(t^{(1)}_{\frac{\alpha}{2}}, n_1 - 1) = 1 - P(t^{(2)}_{\frac{\alpha}{2}}, n_2 - 1)$$

Here, $P(t, n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

$$\beta_1 = \frac{{\sigma_1}^2}{n_1}, \beta_2 = \frac{{\sigma_2}^2}{n_2}$$

${\sigma_1}^2$ and ${\sigma_2}^2$ are unbiased estimates of the population variances.

(4) **Interval estimation of the population variance due to one set of samples**
From the variance (or population variance) $\sigma^2$ of one set of sample data of size $n$, obtain the confidence interval of the population variance when the confidence level $(t_1, t_2)$ is specified. The confidence interval $(t_1, t_2)$ is defined as follows.

$$t_1 = \frac{\sigma^2(n-1)}{{\chi_1}^2}$$

$$t_2 = \frac{\sigma^2(n-1)}{{\chi_2}^2}$$

$\sigma^2$ is an unbiased estimate of the population variance. Also,

$$\frac{\alpha}{2} = P({\chi_1}^2; n-1) = 1 - P({\chi_2}^2; n-1)$$

Here, $P(x, y)$ is the cumulative distribution function (c.d.f.) of a $\chi^2$ distribution with $y$ degrees of freedom.

(5) **Interval estimation of the population correlation coefficient according to one set of samples**
From the sample correlation coefficient $r$ of one set of sample data of size $n$, obtain the confidence interval of the population correlation coefficient $\rho$ when the confidence level $1-\alpha$ is specified. The confidence interval $(t_1, t_2)$ is defined as follows.

$$t_1 = \frac{e^{a-b}-1}{e^{a-b}+1}$$

$$t_2 = \frac{e^{a+b}-1}{e^{a+b}+1}$$

Here,

$$a = \log_e \frac{1+r}{1-r}$$

$$b = \frac{2z_{\frac{\alpha}{2}}}{\sqrt{n-3}}$$

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

$P(z)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

(6) **Interval estimation of the difference of the population correlation coefficient according to two sets of independent samples**
From the correlation coefficients $r_1$ and $r_2$ of two sets of independent sample data of sizes $n_1$ and $n_2$, respectively, obtain the confidence interval of the difference of the population correlation coefficients $\rho_1$ and $\rho_2$ when the confidence level $1-\alpha$ is specified. If $\rho_1 = \rho_2 = \rho$, the confidence level of $\rho$ is obtained.

(a) When $\rho_1 = \rho_2 = \rho$
The confidence interval $(t_1, t_2)$ of $\rho$ is defined as follows.

$$t_1 = \frac{e^{a-b}-1}{e^{a-b}+1}$$

$$t_2 = \frac{e^{a+b} - 1}{e^{a+b} + 1}$$

Here,

$$a = \frac{2(n_1 - 3)z_1 + 2(n_2 - 3)z_2}{n_1 + n_2 - 6}$$

$$b = \frac{2z_{\frac{\alpha}{2}}}{\sqrt{n_1 + n_2 - 6}}$$

$$z_1 = \frac{1}{2} \log_e \frac{1 + r_1}{1 - r_1}$$

$$z_2 = \frac{1}{2} \log_e \frac{1 + r_2}{1 - r_2}$$

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

$P(z)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

(b) When $\rho_1 \neq \rho_2$

The confidence interval $(t_1, t_2)$ of $\rho_1 - \rho_2$ is defined as follows.

$$t_1 = \frac{e^{a-b} - 1}{e^{a-b} + 1}$$

$$t_2 = \frac{e^{a+b} - 1}{e^{a+b} + 1}$$

Here,

$$a = \log_e \frac{1 + r_1}{1 - r_1} - \log_e \frac{1 + r_2}{1 - r_2}$$

$$b = 2z_{\frac{\alpha}{2}} \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$$

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

$P(z)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

(7) **Interval estimation in the simple linear regression**

For regression coefficient $a$, constant term $b$, and a given specific data $x_0$ in the following simple linear regression expression (or regression line) related to one set of sample data $\{x_i,\, y_i\}$ $(1, \cdots, n)$ of size $n$

$$\hat{y}_i = ax_i + b$$

obtain the estimate $\hat{y}_0$ and the confidence interval of the confidence level $1 - \alpha$ of the theoretical value $Ax_0 - B$. Assume that $y_i$ corresponding to each $x_i$ is the random sample from the normal population having the mean $Ax_i - B$ and the variance $\sigma^2$. Obtain the regression coefficient $a$ and constant term $b$ of the sample data from the following normal equations.

$$\begin{cases} \displaystyle\sum_{i=1}^{n} y_i = a \sum_{i=1}^{n} x_i + bn \\ \displaystyle\sum_{i=1}^{n} x_i y_i = a \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_i \end{cases}$$

The confidence interval $(t_1, t_2)$ is defined as follows.

(a) Regression coefficient

$$t_1 = a - t_{\frac{\alpha}{2}} \cdot s_a$$

$$t_2 = a + t_{\frac{\alpha}{2}} \cdot s_a$$

Here,

$$s_a = \sqrt{\frac{\sigma^2}{\sum(x_i - \mu_x)^2}}$$

i. When the population variance is known
   $\sigma^2$ : Population variance
   $$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}})$$
   Here, $P(t)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

ii. When the population variance is unknown
   $\sigma^2$ : Unbiased variance of error variation
   $$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}}|n-2)$$
   Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

(b) Constant term

$$t_1 = a - t_{\frac{\alpha}{2}} \cdot s_b$$

$$t_2 = a + t_{\frac{\alpha}{2}} \cdot s_b$$

Here,

$$s_b = \sqrt{\sigma^2\left[\frac{1}{n} + \frac{\mu_x^2}{\sum(x_i - \mu_x)^2}\right]}$$

i. When the population variance is known
   $\sigma^2$ : Population variance
   $$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}})$$
   Here, $P(t)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

ii. When the population variance is unknown
   $\sigma^2$ : Unbiased variance of error variation
   $$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}}|n-2)$$
   Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

(c) Estimated value

$$t_1 = \hat{y}_0 - t_{\frac{\alpha}{2}} \cdot s_y$$

$$t_2 = \hat{y}_0 + t_{\frac{\alpha}{2}} \cdot s_y$$

Here,

$$s_y = \sqrt{\sigma^2\left[1 + \frac{1}{n} + \frac{(x_0 - \mu_x)^2}{\sum(x_i - \mu_x)^2}\right]}$$

i. When the population variance is known
$\sigma^2$ : Population variance
$$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}})$$
Here, $P(t)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

ii. When the population variance is unknown
$\sigma^2$ : Unbiased variance of error variation
$$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}}|n-2)$$
Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

(d) Theoretical value

$$t_1 = \hat{y}_0 - t_{\frac{\alpha}{2}} \cdot s_0$$

$$t_2 = \hat{y}_0 + t_{\frac{\alpha}{2}} \cdot s_0$$

Here,

$$s_0 = \sqrt{\sigma^2 \left[ \frac{1}{n} + \frac{(x_0 - \mu_x)^2}{\sum (x_i - \mu_x)^2} \right]}$$

i. When the population variance is known
$\sigma^2$ : Population variance
$$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}})$$
Here, $P(t)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

ii. When the population variance is unknown
$\sigma^2$ : Unbiased variance of error variation
$$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}}|n-2)$$
Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

(8) **Test of the population rate according to one set of samples**
When the number of data having the observed characteristic in the one set of sample data of size $n$ is $m$, test the hypothesis $p = p_0$ related to the ratio $p$ in the population to which the respective set of sample data belong with the confidence level $1 - \alpha$. The test criteria are as follows.

(a) When the alternative hypothesis is $p \neq p_0$
For $f_1$ and $f_2$ defined as follows

$$f_1 = \frac{2(n-m)p_0}{2(m+1)(1-p_0)}$$

$$f_2 = \frac{2m(1-p_0)}{2(n-m+1)p_0}$$

$$\begin{cases} \text{If } f_1 \geq F_1 \text{ or } f_2 \geq F_2, \text{ reject} \\ \text{If } f_1 < F_1 \text{ and } f_2 < F_2, \text{ accept} \end{cases}$$

Here,
$$\frac{\alpha}{2} = 1 - P(F_1|2(n-m+1), 2m) = 1 - P(F_2|2(m+1), 2(n-m))$$
$P(F|n_1, n_2)$ is the cumulative distribution function (c.d.f.) of a $F$ distribution having numbers of degrees of freedom $n_1$ and $n_2$.

(b) When the alternative hypothesis is $p < p_0$
For $f_1$ defined as follows

$$f_1 = \frac{2(n-m)p_0}{2(m+1)(1-p_0)}$$

$$\begin{cases} \text{If } f_1 \geq F_1^*, \text{ reject} \\ \text{If } f_1 < F_1^*, \text{ accept} \end{cases} \text{Here,}$$

$$\alpha = 1 - P(F_1^*|2(m+1), 2(n-m))$$

$P(F|n_1, n_2)$ is the cumulative distribution function (c.d.f.) of a $F$ distribution having numbers of degrees of freedom $n_1$ and $n_2$.

(c) When the alternative hypothesis is $p > p_0$
For $f_2$ defined as follows

$$f_2 = \frac{2m(1-p_0)}{2(n-m+1)(1-p_0)}$$

$$\begin{cases} \text{If } f_2 \geq F_2^*, \text{ reject} \\ \text{If } f_2 < F_2^*, \text{ accept} \end{cases} \text{Here,}$$

$$\alpha = 1 - P(F_2^*|2(n-m+1), 2m)$$

$P(F|n_1, n_2)$ is the cumulative distribution function (c.d.f.) of a $F$ distribution having numbers of degrees of freedom $n_1$ and $n_2$.

(9) **Test of the difference of the population ratios according to two sets of independent samples**
When the number of data having the observed characteristic in the two sets of independent sample data of sizes $n_1$ and $n_2$ are $m_1$ and $m_2$, respectively, test the hypothesis $p_1 = p_2$ related to the ratios $p_1$ and $p_2$ in the population to which the respective sets of sample data belong with the confidence level $1 - \alpha$.
Let the sample ratios in the two sets of independent samples, respectively, be $\hat{p}_1$ and $\hat{p}_2$ as shown below.

$$\hat{p}_1 = \frac{m_1}{n_1}, \hat{p}_2 = \frac{m_2}{n_2}$$

The test criteria are as follows.

(a) When no continuity correction is performed

i. When the alternative hypothesis is $p_1 \neq p_2$
For $z$ defined as follows

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$\begin{cases} \text{If } |z| \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |z| < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where, $\hat{p}$ and $\frac{\alpha}{2}$ are as follows.

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{m_1 + m_2}{n_1 + n_2}$$

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

ii. When the alternative hypothesis is $p_1 < p_2$

For $z$ defined as follows

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$\begin{cases} \text{If } z \leq -z_\alpha, \text{ reject} \\ \text{If } z > -z_\alpha, \text{ accept} \end{cases}$$

where, $\hat{p}$ and $\alpha$ are as follows.

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{m_1 + m_2}{n_1 + n_2}$$

$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

iii. When the alternative hypothesis is $p_1 > p_2$

For $z$ defined as follows

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$\begin{cases} \text{If } z \geq z_\alpha, \text{ reject} \\ \text{If } z < z_\alpha, \text{ accept} \end{cases}$$

where, $\hat{p}$ and $\alpha$ are as follows.

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{m_1 + m_2}{n_1 + n_2}$$

$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

(b) When continuity correction is performed

i. When the alternative hypothesis is $p_1 \neq p_2$

For $z$ defined as follows

$$\begin{cases} z = \dfrac{\hat{p}_1 - \hat{p}_2 - 0.5(\frac{1}{n_1} + \frac{1}{n_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} (\text{when } \hat{p}_1 \geq \hat{p}_2) \\ z = \dfrac{\hat{p}_1 - \hat{p}_2 + 0.5(\frac{1}{n_1} + \frac{1}{n_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} (\text{when } \hat{p}_1 < \hat{p}_2) \end{cases}$$

$$\begin{cases} \text{If } |z| \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |z| < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where, $\hat{p}$ and $\frac{\alpha}{2}$ are as follows.

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{m_1 + m_2}{n_1 + n_2}$$

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

ii. When the alternative hypothesis is $p_1 < p_2$

For $z$ defined as follows

$$\begin{cases} z = \dfrac{\hat{p}_1 - \hat{p}_2 - 0.5(\frac{1}{n_1} + \frac{1}{n_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} (\text{when } \hat{p}_1 \geq \hat{p}_2) \\ z = \dfrac{\hat{p}_1 - \hat{p}_2 + 0.5(\frac{1}{n_1} + \frac{1}{n_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} (\text{when } \hat{p}_1 < \hat{p}_2) \end{cases}$$

$$\begin{cases} \text{If } z \leq -z_\alpha, \text{ reject} \\ \text{If } z > -z_\alpha, \text{ accept} \end{cases}$$

where, $\hat{p}$ and $\alpha$ are as follows.

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{m_1 + m_2}{n_1 + n_2}$$
$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

iii. When the alternative hypothesis is $p_1 > p_2$

For $z$ defined as follows

$$\begin{cases} z = \dfrac{\hat{p}_1 - \hat{p}_2 - 0.5(\frac{1}{n_1} + \frac{1}{n_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} (\text{when } \hat{p}_1 \geq \hat{p}_2) \\ z = \dfrac{\hat{p}_1 - \hat{p}_2 + 0.5(\frac{1}{n_1} + \frac{1}{n_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} (\text{when } \hat{p}_1 < \hat{p}_2) \end{cases}$$

$$\begin{cases} \text{If } z \geq z_\alpha, \text{ reject} \\ \text{If } z < z_\alpha, \text{ accept} \end{cases}$$

where, $\hat{p}$ and $\alpha$ are as follows.

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{m_1 + m_2}{n_1 + n_2}$$
$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

(10) **Test of the population mean according to one set of samples**

From the mean $\mu_x$ and variance (or population variance) $\sigma^2$ of one set of sample data of size $n$, test the hypothesis $\mu = \mu_0$ with the confidence level $1 - \alpha$. The test criteria are as follows.

(a) When the population variance is known

i. When the alternative hypothesis is $\mu \neq \mu_0$

For $z$ defined as follows

$$z = \frac{\mu_x - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

$$\begin{cases} \text{If } |z| \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |z| < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where,

$\sigma^2$ : Population variance

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

$P(z)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

ii. When the alternative hypothesis is $\mu < \mu_0$

For $z$ defined as follows

$$z = \frac{\mu_x - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

$$\begin{cases} \text{If } z \leq -z_\alpha, \text{ reject} \\ \text{If } z > -z_\alpha, \text{ accept} \end{cases}$$

where,

$\sigma^2$ : Population variance

$$\alpha = 1 - P(z_\alpha)$$

$P(z)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

iii. When the alternative hypothesis is $\mu > \mu_0$
For $z$ defined as follows

$$z = \frac{\mu_x - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

$$\begin{cases} \text{If } z \geq z_\alpha, \text{ reject} \\ \text{If } z < z_\alpha, \text{ accept} \end{cases}$$

where,
$\sigma^2$ : Population variance

$$\alpha = 1 - P(z_\alpha)$$

$P(z)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

(b) When the population variance is unknown

i. When the alternative hypothesis is $\mu \neq \mu_0$
For $t$ defined as follows

$$t = \frac{\mu_x - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

$$\begin{cases} \text{If } |t| \geq t_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < t_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where,
$\sigma^2$ : Unbiased estimate of population variance

$$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}}|n-1)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

ii. When the alternative hypothesis is $\mu < \mu_0$
For $t$ defined as follows

$$t = \frac{\mu_x - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

$$\begin{cases} \text{If } t \leq -t_\alpha, \text{ reject} \\ \text{If } t > -t_\alpha, \text{ accept} \end{cases}$$

where,
$\sigma^2$ : Unbiased estimate of population variance

$$\alpha = 1 - P(t_\alpha|n-1)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

iii. When the alternative hypothesis is $\mu > \mu_0$ For $t$ defined as follows

$$t = \frac{\mu_x - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

$$\begin{cases} \text{If } t \geq t_\alpha, \text{ reject} \\ \text{If } t < t_\alpha, \text{ accept} \end{cases}$$

where,
$\sigma^2$ : Unbiased estimate of population variance

$$\alpha = 1 - P(t_\alpha|n-1)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

(11) **Test of the difference of the population means according to two sets of independent samples**
From the means $\mu_{x_1}$ and $\mu_{x_2}$ and variances (or population variances) $\sigma_1^2$ and $\sigma_2^2$ of two sets of independent sample data of sizes $n_1$ and $n_2$, respectively, test the hypothesis $\mu_1 = \mu_2$ related to the means $\mu_1$ and $\mu_2$ in the population to which the respective sets of sample data belong with the confidence level $1 - \alpha$. The test criteria are as follows.

(a) When the population variances are known

i. When the alternative hypothesis is $\mu_1 \neq \mu_2$ For $z$ defined as follows

$$z = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$\begin{cases} \text{If } |z| \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |z| < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $\frac{\alpha}{2}$ is as follows.

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.
$\sigma_1^2$, $\sigma_2^2$ : Population variances of the two sets

ii. When the alternative hypothesis is $\mu_1 < \mu_2$
For $z$ defined as follows

$$z = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$\begin{cases} \text{If } z \leq -z_{\alpha}, \text{ reject} \\ \text{If } z > -z_{\alpha}, \text{ accept} \end{cases}$$

where $\alpha$ is as follows.

$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.
$\sigma_1^2$, $\sigma_2^2$ : Population variances of the two sets

iii. When the alternative hypothesis is $\mu_1 > \mu_2$
For $z$ defined as follows

$$z = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$\begin{cases} \text{If } z \geq z_{\alpha}, \text{ reject} \\ \text{If } z < z_{\alpha}, \text{ accept} \end{cases}$$

where $\alpha$ is as follows.

$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.
$\sigma_1^2$, $\sigma_2^2$ : Population variances of the two sets

(b) When the population variances of the two sets are equal and that value is unknown

i. When the alternative hypothesis is $\mu_1 \neq \mu_2$
For $t$ defined as follows

$$t = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$\begin{cases} \text{If } |t| \geq t_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < t_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $s_p^2$ and $\frac{\alpha}{2}$ are as follows.

$$s_p^2 = \frac{(n_1-1)\sigma_1^2 + (n_2-1)\sigma_2^2}{n_1+n_2-2}$$

$$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}}|n_1+n_2-2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom. $\sigma_1^2$, $\sigma_2^2$ : Unbiased estimates of the population variances.

ii. When the alternative hypothesis is $\mu_1 < \mu_2$

For $t$ defined as follows

$$t = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$\begin{cases} \text{If } t \leq -t_\alpha, \text{ reject} \\ \text{If } t > -t_\alpha, \text{ accept} \end{cases}$$

where $s_p^2$ and $\alpha$ are as follows.

$$s_p^2 = \frac{(n_1-1)\sigma_1^2 + (n_2-1)\sigma_2^2}{n_1+n_2-2}$$

$$\alpha = 1 - P(t_\alpha|n_1+n_2-2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

$\sigma_1^2$, $\sigma_2^2$ : Unbiased estimates of the population variances.

iii. When the alternative hypothesis is $\mu_1 > \mu_2$

For $t$ defined as follows

$$t = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$\begin{cases} \text{If } t \geq t_\alpha, \text{ reject} \\ \text{If } t < t_\alpha, \text{ accept} \end{cases}$$

where $s_p^2$ and $\alpha$ are as follows.

$$s_p^2 = \frac{(n_1-1)\sigma_1^2 + (n_2-1)\sigma_2^2}{n_1+n_2-2}$$

$$\alpha = 1 - P(t_\alpha|n_1+n_2-2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

$\sigma_1^2$, $\sigma_2^2$ : Unbiased estimates of the population variances.

(c) When the population variances of the two sets are not equal and those values are unknown

i. When the alternative hypothesis is $\mu_1 \neq \mu_2$

For $t$ defined as follows

$$t = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$t^*_{\frac{\alpha}{2}} = \frac{\beta_1 t^{(1)}_{\frac{\alpha}{2}} + \beta_2 t^{(2)}_{\frac{\alpha}{2}}}{\beta_1 + \beta_2}$$

$$\begin{cases} \text{If } |t| \geq t^*_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < t^*_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $\frac{\alpha}{2}$ is as follows.

$$\frac{\alpha}{2} = 1 - P(t^{(1)}_{\frac{\alpha}{2}}|n_1-1) = 1 - P(t^{(2)}_{\frac{\alpha}{2}}|n_2-1)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

$$\beta_1 = \frac{\sigma_1^2}{n_1}, \beta_2 = \frac{\sigma_2^2}{n_2}$$

$\sigma_1^2$, $\sigma_2^2$ : Unbiased estimates of the population variances

ii. When the alternative hypothesis is $\mu_1 < \mu_2$

For $t$ defined as follows

$$t = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$t_\alpha^* = \frac{\beta_1 t_\alpha^{(1)} + \beta_2 t_\alpha^{(2)}}{\beta_1 + \beta_2}$$

$$\begin{cases} \text{If } t \leq -t_\alpha^*, \text{ reject} \\ \text{If } t > -t_\alpha^*, \text{ accept} \end{cases}$$

where $\alpha$ is as follows.

$$\alpha = 1 - P(t_\alpha^{(1)}|n_1 - 1) = 1 - P(t_\alpha^{(2)}|n_2 - 1)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

$$\beta_1 = \frac{\sigma_1^2}{n_1}, \beta_2 = \frac{\sigma_2^2}{n_2}$$

$\sigma_1^2$, $\sigma_2^2$ : Unbiased estimates of the population variances

iii. When the alternative hypothesis is $\mu_1 > \mu_2$

For $t$ defined as follows

$$t = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$t_\alpha^* = \frac{\beta_1 t_\alpha^{(1)} + \beta_2 t_\alpha^{(2)}}{\beta_1 + \beta_2}$$

$$\begin{cases} \text{If } t \geq t_\alpha^*, \text{ reject} \\ \text{If } t < t_\alpha^*, \text{ accept} \end{cases}$$

where $\alpha$ is as follows.

$$\alpha = 1 - P(t_\alpha^{(1)}|n_1 - 1) = 1 - P(t_\alpha^{(2)}|n_2 - 1)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

$$\beta_1 = \frac{\sigma_1^2}{n_1}, \beta_2 = \frac{\sigma_2^2}{n_2}$$

$\sigma_1^2$, $\sigma_2^2$ : Unbiased estimates of the population variances

(12) **Test of the population variance due to one set of samples**

From the variance (or population variance) $s^2$ of one set of sample data of size $n$, test the hypothesis $\sigma^2 = \sigma_0^2$ related to the population variance $\sigma^2$ in the population to which the respective set of sample data belong with the confidence level $1 - \alpha$. The test criteria are as follows.

(a) When the alternative hypothesis is $\sigma^2 \neq \sigma_0^2$

For $\chi^2$ defined as follows

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

$$\begin{cases} \text{If } \chi^2 \leq \chi^2_{1-\frac{\alpha}{2}} \text{ or } \chi^2 \geq \chi^2_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } \chi^2_{1-\frac{\alpha}{2}} < \chi^2 < \chi^2_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $\frac{\alpha}{2}$ is as follows.

$$\frac{\alpha}{2} = 1 - P(\chi^2_{\frac{\alpha}{2}}|n-1) = P(\chi^2_{1-\frac{\alpha}{2}}|n-1)$$

$P(\chi^2|n)$ is the cumulative distribution function (c.d.f.) of a $\chi^2$ distribution having number of degrees of freedom $n$.

$s^2$ : Unbiased estimate of population variance

(b) When the alternative hypothesis is $\sigma^2 < \sigma_0^2$

For $\chi^2$ defined as follows

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

$$\begin{cases} \text{If } \chi^2 \leq \chi^2_{1-\alpha}, \text{ reject} \\ \text{If } \chi^2 > \chi^2_{1-\alpha}, \text{ accept} \end{cases}$$

where $\alpha$ is as follows.

$$\alpha = P(\chi^2_{1-\alpha}|n-1)$$

$P(\chi^2|n)$ is the cumulative distribution function (c.d.f.) of a $\chi^2$ distribution having number of degrees of freedom $n$.

$s^2$ : Unbiased estimate of population variance

(c) When the alternative hypothesis is $\sigma^2 > \sigma_0^2$

For $\chi^2$ defined as follows

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

$$\begin{cases} \text{If } \chi^2 \geq \chi^2_{\alpha}, \text{ reject} \\ \text{If } \chi^2 < \chi^2_{\alpha}, \text{ accept} \end{cases}$$

where $\alpha$ is as follows.

$$\alpha = P(\chi^2_{\alpha}|n-1)$$

$P(\chi^2|n)$ is the cumulative distribution function (c.d.f.) of a $\chi^2$ distribution having number of degrees of freedom $n$.

$s^2$ : Unbiased estimate of population variance

(13) **Test of the population correlation coefficient according to one set of samples**

From the sample correlation coefficient $r$ of one set of sample data of size $n$, test the hypothesis $\rho = \rho_0$ related to the population correlation coefficient $\rho$ in the population to which the respective set of sample data belong with the confidence level $1 - \alpha$. The test criteria are as follows.

(a) Hypothesis: $\rho = 0$

i. When the alternative hypothesis is $\rho \neq 0$

For $t$ defined as follows

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

$$\begin{cases} \text{If } |t| \geq t_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < t_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where, $\frac{\alpha}{2}$ is as follows.

$$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}}|n-2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

ii. When the alternative hypothesis is $\rho < 0$
For $t$ defined as follows
$$t = r\sqrt{\frac{n-2}{1-r^2}}$$
$$\begin{cases} \text{If } t \geq -t_\alpha, \text{ reject} \\ \text{If } t < -t_\alpha, \text{ accept} \end{cases}$$
where, $\frac{\alpha}{2}$ is as follows.
$$\frac{\alpha}{2} = 1 - P(t_\alpha|n-2)$$
Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

iii. When the alternative hypothesis is $\rho > 0$
For $t$ defined as follows
$$t = r\sqrt{\frac{n-2}{1-r^2}}$$
$$\begin{cases} \text{If } t \geq t_\alpha, \text{ reject} \\ \text{If } t < t_\alpha, \text{ accept} \end{cases}$$
where, $\frac{\alpha}{2}$ is as follows.
$$\frac{\alpha}{2} = 1 - P(t_\alpha|n-2)$$
Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

(b) Hypothesis: $\rho = \rho_0$

i. When the alternative hypothesis is $\rho \neq \rho_0$
For $t$ defined as follows
$$t = (z - z_0)\sqrt{n-3}$$
$$\begin{cases} \text{If } |t| \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$
where, $z$, $z_0$ and $\frac{\alpha}{2}$ are as follows.
$$z = \frac{1}{2}\log_e \frac{1+r}{1-r}$$
$$z_0 = \frac{1}{2}\log_e \frac{1+\rho_0}{1-\rho_0}$$
$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$
Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

ii. When the alternative hypothesis is $\rho < \rho_0$
For $t$ defined as follows
$$t = (z - z_0)\sqrt{n-3}$$
$$\begin{cases} \text{If } t \leq -z_\alpha, \text{ reject} \\ \text{If } t > -z_\alpha, \text{ accept} \end{cases}$$
where, $z$, $z_0$ and $\frac{\alpha}{2}$ are as follows.
$$z = \frac{1}{2}\log_e \frac{1+r}{1-r}$$
$$z_0 = \frac{1}{2}\log_e \frac{1+\rho_0}{1-\rho_0}$$

$$\frac{\alpha}{2} = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

iii. When the alternative hypothesis is $\rho > \rho_0$

For $t$ defined as follows

$$t = (z - z_0)\sqrt{n-3}$$

$$\begin{cases} \text{If } t \geq z_\alpha, \text{ reject} \\ \text{If } t < z_\alpha, \text{ accept} \end{cases}$$

where, $z$, $z_0$ and $\frac{\alpha}{2}$ are as follows.

$$z = \frac{1}{2}\log_e \frac{1+r}{1-r}$$

$$z_0 = \frac{1}{2}\log_e \frac{1+\rho_0}{1-\rho_0}$$

$$\frac{\alpha}{2} = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

(14) **Test of the difference of the population correlation coefficients according to two sets of independent samples**

From the correlation coefficients $r_1$ and $r_2$ of two sets of independent sample data of sizes $n_1$ and $n_2$, respectively, test the hypothesis $\rho_1 = \rho_2$ related to the population correlation coefficients $\rho_1$ and $\rho_2$ in the population to which the respective set of sample data belong with the confidence level $1 - \alpha$. The test criteria are as follows.

(a) When the alternative hypothesis is $\rho_1 \neq \rho_2$

For $t$ defined as follows

$$t = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$$

$$\begin{cases} \text{If } |t| \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where, $z_1$, $z_2$ and $\frac{\alpha}{2}$ are as follows.

$$z_1 = \frac{1}{2}\log_e \frac{1+r_1}{1-r_1}$$

$$z_2 = \frac{1}{2}\log_e \frac{1+r_2}{1-r_2}$$

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

(b) When the alternative hypothesis is $\rho_1 < \rho_2$

For $t$ defined as follows

$$t = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$$

$$\begin{cases} \text{If } t \leq -z_\alpha, \text{ reject} \\ \text{If } t > -z_\alpha, \text{ accept} \end{cases}$$

where, $z_1$, $z_2$ and $\alpha$ are as follows.

$$z_1 = \frac{1}{2}\log_e \frac{1+r_1}{1-r_1}$$

$$z_2 = \frac{1}{2} \log_e \frac{1 + r_2}{1 - r_2}$$

$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

(c) When the alternative hypothesis is $\rho_1 > \rho_2$
For $t$ defined as follows

$$t = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$$

$$\begin{cases} \text{If } t \geq z_\alpha, \text{ reject} \\ \text{If } t < z_\alpha, \text{ accept} \end{cases}$$

where, $z_1$, $z_2$ and $\alpha$ are as follows.

$$z_1 = \frac{1}{2} \log_e \frac{1 + r_1}{1 - r_1}$$

$$z_2 = \frac{1}{2} \log_e \frac{1 + r_2}{1 - r_2}$$

$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

(15) **Test in the simple linear regression**

For regression coefficient $a$ and constant term $b$ in the following simple linear regression expression (or regression line) related to one set of sample data $\{x_i, y_i\}$ $(1, \cdots, n)$ of size $n$

$$\hat{y}_i = ax_i + b$$

test the hypothesis related to the regression coefficient $A$ and constant term $B$ in the population to which the respective set of sample data belong with the confidence level $1 - \alpha$. Assume that $y_i$ corresponding to each $x_i$ is the random sample from the normal population having the mean $Ax_i - B$ and the variance $\sigma^2$. Obtain the regression coefficient $a$ and constant term $b$ of the sample data from the following normal equations.

$$\begin{cases} \displaystyle\sum_{i=1}^{n} y_i = a \sum_{i=1}^{n} x_i + bn \\ \displaystyle\sum_{i=1}^{n} x_i y_i = a \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_i \end{cases}$$

The test criteria are as follows.

(a) Regression coefficient
Hypothesis: $A = A_0$

i. When the population variance is known

A. When the alternative hypothesis is $A \neq A_0$
For $t$ defined as follows

$$t = \frac{a - A_0}{s_a}$$

$$\begin{cases} \text{If } |t| \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $s_a$ is as follows.

$$s_a = \sqrt{\frac{\sigma^2}{\sum(x_i - \mu_x)^2}}$$

$\sigma^2$ : Population variance

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

B. When the alternative hypothesis is $A < A_0$

For $t$ defined as follows

$$t = \frac{a - A_0}{s_a}$$

$$\begin{cases} \text{If } t \geq -z_{\alpha}, \text{ reject} \\ \text{If } t < -z_{\alpha}, \text{ accept} \end{cases}$$

where $s_a$ is as follows.

$$s_a = \sqrt{\frac{\sigma^2}{\sum(x_i - \mu_x)^2}}$$

$\sigma^2$ : Population variance

$$\alpha = 1 - P(z_{\alpha})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

C. When the alternative hypothesis is $A > A_0$

For $t$ defined as follows

$$t = \frac{a - A_0}{s_a}$$

$$\begin{cases} \text{If } t \geq z_{\alpha}, \text{ reject} \\ \text{If } t < z_{\alpha}, \text{ accept} \end{cases}$$

where $s_a$ is as follows.

$$s_a = \sqrt{\frac{\sigma^2}{\sum(x_i - \mu_x)^2}}$$

$\sigma^2$ : Population variance

$$\alpha = 1 - P(z_{\alpha})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

ii. When the population variance is unknown

A. When the alternative hypothesis is $A < A_0$

For $t$ defined as follows

$$t = \frac{a - A_0}{s_a}$$

$$\begin{cases} \text{If } |t| \geq t_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < t_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $s_a$ is as follows.

$$s_a = \sqrt{\frac{\sigma^2}{\sum(x_i - \mu_x)^2}}$$

$\sigma^2$ : Unbiased variance of error variation

$$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}}|n - 2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

B. When the alternative hypothesis is $A < A_0$
For $t$ defined as follows
$$t = \frac{a - A_0}{s_a}$$
$$\begin{cases} \text{If } t \geq -t_\alpha, \text{ reject} \\ \text{If } t < -t_\alpha, \text{ accept} \end{cases}$$
where $s_a$ is as follows.
$$s_a = \sqrt{\frac{\sigma^2}{\sum(x_i - \mu_x)^2}}$$
$\sigma^2$ : Unbiased variance of error variation
$$\alpha = 1 - P(t_\alpha | n - 2)$$
Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

C. When the alternative hypothesis is $A > A_0$
For $t$ defined as follows
$$t = \frac{a - A_0}{s_a}$$
$$\begin{cases} \text{If } t \geq t_\alpha, \text{ reject} \\ \text{If } t < t_\alpha, \text{ accept} \end{cases}$$
where $s_a$ is as follows.
$$s_a = \sqrt{\frac{\sigma^2}{\sum(x_i - \mu_x)^2}}$$
$\sigma^2$ : Unbiased variance of error variation
$$\alpha = 1 - P(t_\alpha | n - 2)$$
Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

(b) Constant term
Hypothesis: $B = B_0$

i. When the population variance is known

A. When the alternative hypothesis is $B \neq B_0$
For $t$ defined as follows
$$t = \frac{b - B_0}{s_b}$$
$$\begin{cases} \text{If } |t| \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$
where $s_b$ is as follows.
$$s_b = \sqrt{\sigma^2 \left[ \frac{1}{n} + \frac{\mu_x^2}{\sum(x_i - \mu_x)^2} \right]}$$
$\sigma^2$ : Population variance
$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$
Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

B. When the alternative hypothesis is $B < B_0$
For $t$ defined as follows
$$t = \frac{b - B_0}{s_b}$$

$$\begin{cases} \text{If } t \geq -z_\alpha, \text{ reject} \\ \text{If } t < -z_\alpha, \text{ accept} \end{cases}$$

where $s_b$ is as follows.

$$s_b = \sqrt{\sigma^2 \left[ \frac{1}{n} + \frac{\mu_x^2}{\sum(x_i - \mu_x)^2} \right]}$$

$\sigma^2$ : Population variance

$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

C. When the alternative hypothesis is $B > B_0$

For $t$ defined as follows

$$t = \frac{b - B_0}{s_b}$$

$$\begin{cases} \text{If } t \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } t < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $s_b$ is as follows.

$$s_b = \sqrt{\sigma^2 \left[ \frac{1}{n} + \frac{\mu_x^2}{\sum(x_i - \mu_x)^2} \right]}$$

$\sigma^2$ : Population variance

$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

ii. When the population variance is unknown

A. When the alternative hypothesis is $B \neq B_0$

For $t$ defined as follows

$$t = \frac{b - B_0}{s_b}$$

$$\begin{cases} \text{If } |t| \geq t_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < t_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $s_b$ is as follows.

$$s_b = \sqrt{\sigma^2 \left[ \frac{1}{n} + \frac{\mu_x^2}{\sum(x_i - \mu_x)^2} \right]}$$

$\sigma^2$ : Unbiased variance of error variation

$$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}} | n - 2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

B. When the alternative hypothesis is $B < B_0$

For $t$ defined as follows

$$t = \frac{b - B_0}{s_b}$$

$$\begin{cases} \text{If } t \geq -t_\alpha, \text{ reject} \\ \text{If } t < -t_\alpha, \text{ accept} \end{cases}$$

where $s_b$ is as follows.

$$s_b = \sqrt{\sigma^2 \left[ \frac{1}{n} + \frac{\mu_x^2}{\sum(x_i - \mu_x)^2} \right]}$$

$\sigma^2$ : Unbiased variance of error variation

$$\alpha = 1 - P(t_\alpha | n - 2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

C. When the alternative hypothesis is $B > B_0$

For $t$ defined as follows

$$t = \frac{b - B_0}{s_b}$$

$$\begin{cases} \text{If } t \geq t_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } t < t_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $s_b$ is as follows.

$$s_b = \sqrt{\sigma^2 \left[ \frac{1}{n} + \frac{\mu_x^2}{\sum (x_i - \mu_x)^2} \right]}$$

$\sigma^2$ : Unbiased variance of error variation

$$\alpha = 1 - P(t_\alpha | n - 2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

### 6.1.2 Reference Bibliography

(1) Lehmann, E. L. , "Testing statistical hypotheses", John Wiley and Sons, New York (1959, 2nd ed. 1986)

(2) Snedecar, G. W. , "Statistical methods", Ames, Iowa (1940)

## 6.2 INTERVAL ESTIMATES

### 6.2.1 D3IERA, R3IERA
### Interval Estimation of the Population Ratio According to One Set of Samples

(1) **Function**

The D3IERA or R3IERA obtains the confidence interval of confidence coefficient $1-\alpha$ related to the population ratio $p$ when the number of samples having the characteristic being observed among one set of sample data of size $n$ is assumed to be $m$. The confidence interval $(p_1, p_2)$ is defined as follows.

$$p_1 = \frac{m}{(n-m+1)F^{(1)}_{\frac{\alpha}{2}} + m}$$

$$p_2 = \frac{(m+1)F^{(2)}_{\frac{\alpha}{2}}}{(m+1)F^{(2)}_{\frac{\alpha}{2}} + (n-m)}$$

where,

$$\frac{\alpha}{2} = 1 - P(F^{(1)}_{\frac{\alpha}{2}}|2(n-m+1), 2m) = 1 - P(F^{(2)}_{\frac{\alpha}{2}}|2(m+1), 2(n-m))$$

Here, $P(F|n_1, n_2)$ is the value of the cumulative distribution function (c.d.f.) of the $F$ distribution with $n_1$ and $n_2$ degrees of freedom. Also, assume the sample ratio $\hat{p} = \frac{m}{n}$ is given as follows.

- When $\hat{p} = 0$

  $p_2$, which is given as follows:

  $$p_2 = \frac{F(\alpha|2, 2n)}{n + F(\alpha|2, 2n)}$$

  is the upper bound of the one-tailed confidence interval of confidence level $1-\alpha$.

- When $\hat{p} = 1$

  $p_1$, which is given as follows:

  $$p_1 = \frac{n}{n + F(\alpha|2, 2n)}$$

  is the lower bound of the one-tailed confidence interval of confidence level $1-\alpha$.

(2) **Usage**

Double precision:

CALL D3IERA (N, M, CL, CI, IERR)

Single precision:

CALL R3IERA (N, M, CL, CI, IERR)

(3) **Arguments**

| | | |
|---|---|---|
| D:Double precision real | Z:Double precision complex | I: $\left\{ \begin{array}{l} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{array} \right\}$ |
| R:Single precision real | C:Single precision complex | |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Number of sample data $n$ |
| 2 | M | I | 1 | Input | Number of samples having the characteristic being observed $m$ |
| 3 | CL | $\left\{ \begin{array}{c} \text{D} \\ \text{R} \end{array} \right\}$ | 1 | Input | Confidence level $100(1-\alpha)(\%)$ |
| 4 | CI | $\left\{ \begin{array}{c} \text{D} \\ \text{R} \end{array} \right\}$ | 2 | Output | CI(1): Confidence interval lower bound $p_1$<br>CI(2): Confidence interval upper bound $p_2$ |
| 5 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $\mathrm{N} > 0$

(b) $\mathrm{M} \geq 0$

(c) $\mathrm{N} \geq \mathrm{M}$

(d) $0.0 \leq \mathrm{CL} \leq 100.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | CL=100.0 | CI(1)=0.0 and CI(2)=1.0 are performed. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |

(6) **Notes**
None

(7) **Example**

(a) Problem

When the number of samples having the characteristic being observed among a sample of size 20 is assumed to be 14, obtain the confidence interval of the population ratio with 95% confidence level.

(b) Input data

N=20, M=14 and CL=95.0.

(c) Main program

```
      PROGRAM B3IERA
! *** EXAMPLE OF D3IERA ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER N,M
      REAL(8) CL,CI(2)
!
      N=20
      M=14
      CL=90.0D0
      WRITE(6,1000)
      WRITE(6,2000) N
      WRITE(6,2010) M
      WRITE(6,2020) CL
      WRITE(6,3000)
      CALL D3IERA(N,M,CL,CI,IERR)
      WRITE(6,4000) IERR
      WRITE(6,6000) CI(1),CI(2)
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D3IERA ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'N    = ',I3)
 2010 FORMAT(9X,'M    = ',I3)
 2020 FORMAT(9X,'CL   = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 6000 FORMAT(9X,'CI(1) = ',F8.4,2X,'CI(2) = ',F8.4)
      END
```

(d) Output results

```
     *** D3IERA ***
      ** INPUT **
         N    =  20
         M    =  14
         CL   = 90.0


      ** OUTPUT **
         IERR =    0
         CI(1) =   0.4922  CI(2) =   0.8604
```

### 6.2.2 D3IEME, R3IEME
### Interval Estimation of the Population Mean According to One Set of Samples

(1) **Function**

The D3IEME or R3IEME obtains the confidence interval of the population mean from the mean $\mu$ and variance (or population variance) $\sigma^2$ of one set of sample data of size $n$ when the confidence level $1-\alpha$ is specified. The confidence interval $(t_1, t_2)$ is defined as follows.

$$t_1 = \mu - z_{\frac{\alpha}{2}}\sqrt{\beta}$$

$$t_2 = \mu + z_{\frac{\alpha}{2}}\sqrt{\beta}$$

(a) When the population variance is known

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

Here, $P(x)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

$$\beta = \frac{\sigma^2}{n}$$

$\sigma^2$: Population variance

(b) When the population variance is unknown

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}}, n-1)$$

Here, $P(t,n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

$$\beta = \frac{\sigma^2}{n}$$

$\sigma^2$ is an unbiased estimate of the population variance.

(2) **Usage**

Double precision:

CALL D3IEME (N, XE, XV, CL, CI, ISW, IERR)

Single precision:

CALL R3IEME (N, XE, XV, CL, CI, ISW, IERR)

## (3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{ \begin{array}{l} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{array} \right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Number of sample data $n$ |
| 2 | XE | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 1 | Input | Mean of sample data $\mu$ |
| 3 | XV | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 1 | Input | Variance of sample data or population $\sigma^2$ |
| 4 | CL | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 1 | Input | Confidence level $100(1-\alpha)(\%)$ |
| 5 | CI | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 2 | Output | CI(1): Confidence interval lower bound $t_1$<br>CI(2): Confidence interval upper bound $t_2$ |
| 6 | ISW | I | 1 | Input | ISW=1: The variance of the population is entered for XV<br>ISW=2: The variance (not an unbiased estimate) of the sample data is entered for XV<br>ISW=3: The variance (unbiased estimate) of the sample data is entered for XV |
| 7 | IERR | I | 1 | Output | Error indicator |

## (4) **Restrictions**

(a) ISW $\in \{1, 2, 3\}$

(b) N $\geq 2$

(c) XV $> 0.0$

(d) $0.0 \leq$ CL $\leq 100.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | CL=100.0 | The negative minimum value is set for CI(1) and the positive maximum value is set for CI(2). |
| 1010 | CL=0.0 (Confidence level is 0.0%.) | The mean is set for CI(1) and CI(2). |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |

(6) **Notes**
None

(7) **Example**

(a) Problem
Obtain the confidence interval of the population mean at a 95% confidence level when the number of sample data is 100, the mean is 42.0, and the unbiased variance is 2.25.

(b) Input data
ISW=1, N=100, XE=42.0, XV=2.25 and CL=95.0.

(c) Main program

```
      PROGRAM B3IEME
! *** EXAMPLE OF D3IEME ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER N,ISW
      REAL(8) XE,XV,CL,CI(2)
!
      ISW=1
      N=100
      XE=42.0D0
      XV=2.25D0
      CL=95.0D0
      WRITE(6,1000)
      WRITE(6,2000) ISW
      WRITE(6,2010) N
      WRITE(6,2020) XE
      WRITE(6,2030) XV
      WRITE(6,2040) CL
      WRITE(6,3000)
      CALL D3IEME(N,XE,XV,CL,CI,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) CI(1),CI(2)
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D3IEME ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'ISW = ',I3)
 2010 FORMAT(9X,'N   = ',I3)
 2020 FORMAT(9X,'XE  = ',F4.1)
 2030 FORMAT(9X,'XV  = ',F4.1)
 2040 FORMAT(9X,'CL  = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'INTERVAL = (',D17.10,',',D17.10,')')
      END
```

(d) Output results

```
*** D3IEME ***
 ** INPUT **
    ISW =   1
    N   = 100
    XE  = 42.0
    XV  =  2.3
    CL  = 95.0


 ** OUTPUT **
    IERR =    0
    INTERVAL = ( 0.4170600540D+02, 0.4229399460D+02)
```

### 6.2.3 D3IESU, R3IESU Interval Estimation of the Difference of the Population Means According to Two Sets of Independent Samples

(1) **Function**

The D3IESU or R3IESU obtains the confidence interval of the difference of the population means from the means $\mu_1$ and $\mu_2$ and variances (or population variances) ${\sigma_1}^2$ and ${\sigma_2}^2$ of two sets of independent sample data of sizes $n_1$ and $n_2$, respectively, when the confidence level $1-\alpha$ is specified. The confidence interval $(t_1, t_2)$ is defined as follows.

$$t_1 = (\mu_1 - \mu_2) - z_{\frac{\alpha}{2}}\sqrt{\beta_1 + \beta_2}$$

$$t_2 = (\mu_1 - \mu_2) + z_{\frac{\alpha}{2}}\sqrt{\beta_1 + \beta_2}$$

(a) When the population variances are known

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

Here, $P(x)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

$$\beta_1 = \frac{{\sigma_1}^2}{n_1}, \beta_2 = \frac{{\sigma_2}^2}{n_2}$$

${\sigma_1}^2$, ${\sigma_2}^2$: Population variances of the two sets

(b) When the population variances of the two sets are equal and that value is unknown

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}}, n_1 + n_2 - 2)$$

Here, $P(t, n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

$$\beta_1 = \frac{{s_p}^2}{n_1}, \beta_2 = \frac{{s_p}^2}{n_2}$$

${s_p}^2$ is defined as follows.

$${s_p}^2 = \frac{(n_1 - 1){\sigma_1}^2 + (n_2 - 1){\sigma_2}^2}{n_1 + n_2 - 2}$$

${\sigma_1}^2$ and ${\sigma_2}^2$ are unbiased estimates of the population variances.

(c) When the population variances of the two sets are not equal and those values are unknown

$$z_{\frac{\alpha}{2}} = \frac{\beta_1 t^{(1)}_{\frac{\alpha}{2}} + \beta_2 t^{(2)}_{\frac{\alpha}{2}}}{\beta_1 + \beta_2}$$

$$\frac{\alpha}{2} = 1 - P(t^{(1)}_{\frac{\alpha}{2}}, n_1 - 1) = 1 - P(t^{(2)}_{\frac{\alpha}{2}}, n_2 - 1)$$

Here, $P(t, n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with n degrees of freedom.

$$\beta_1 = \frac{{\sigma_1}^2}{n_1}, \beta_2 = \frac{{\sigma_2}^2}{n_2}$$

${\sigma_1}^2$ and ${\sigma_2}^2$ are unbiased estimates of the population variances.

(2) **Usage**

Double precision:

CALL D3IESU (N1, XE1, XV1, N2, XE2, XV2, CL, CI, ISW, IERR)

Single precision:

CALL R3IESU (N1, XE1, XV1, N2, XE2, XV2, CL, CI, ISW, IERR)

(3) **Arguments**

| | | |
|---|---|---|
| D:Double precision real | Z:Double precision complex | I: $\left\{ \begin{array}{l} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{array} \right\}$ |
| R:Single precision real | C:Single precision complex | |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N1 | I | 1 | Input | Number of sample data $n_1$ |
| 2 | XE1 | $\left\{ \begin{array}{c} \text{D} \\ \text{R} \end{array} \right\}$ | 1 | Input | Mean of sample data $\mu_1$ |
| 3 | XV1 | $\left\{ \begin{array}{c} \text{D} \\ \text{R} \end{array} \right\}$ | 1 | Input | Variance of sample data or population ${\sigma_1}^2$ |
| 4 | N2 | I | 1 | Input | Number of sample data $n_2$ |
| 5 | XE2 | $\left\{ \begin{array}{c} \text{D} \\ \text{R} \end{array} \right\}$ | 1 | Input | Mean of sample data $\mu_2$ |
| 6 | XV2 | $\left\{ \begin{array}{c} \text{D} \\ \text{R} \end{array} \right\}$ | 1 | Input | Variance of sample data or population ${\sigma_2}^2$ |
| 7 | CL | $\left\{ \begin{array}{c} \text{D} \\ \text{R} \end{array} \right\}$ | 1 | Input | Confidence level $100(1-\alpha)(\%)$ |
| 8 | CI | $\left\{ \begin{array}{c} \text{D} \\ \text{R} \end{array} \right\}$ | 2 | Output | CI(1): Confidence interval lower bound $t_1$<br>CI(2): Confidence interval upper bound $t_2$ |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 9 | ISW | I | 1 | Input | ISW=1: The population variances of the two sets are entered for XV1 and XV2<br>ISW=2: The population variances of the two sets are equal, and the variances (not unbiased estimates) of the sample data are entered for XV1 and XV2<br>ISW=3: The population variances of the two sets are equal, and the variances (unbiased estimates) of the sample data are entered for XV1 and XV2<br>ISW=4: The population variances of the two sets are not equal, and the variances (not unbiased estimates) of the sample data are entered for XV1 and XV2<br>ISW=5: The population variances of the two sets are not equal, and the variances (unbiased estimates) of the sample data are entered for XV1 and XV2 |
| 10 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $\mathrm{ISW} \in \{1, 2, 3, 4, 5\}$

(b) $\mathrm{N1}, \mathrm{N2} \geq 2$

(c) $\mathrm{XV1}, \mathrm{XV2} > 0.0$

(d) $0.0 \leq \mathrm{CL} \leq 100.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | CL=100.0 | The negative minimum value is set for CI(1) and the positive maximum value is set for CI(2). |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |

(6) **Notes**
None

(7) **Example**

(a) Problem
Obtain the confidence interval of the difference of the population means at a 95% confidence level

from two sets of independent samples for which the sample variances are not equal and the numbers of sample data, the means, and the unbiased variances are 100, 42.0, 2.25 and 50, 30.0, 3.25, respectively.

(b) Input data

ISW=5, N1=100, XE1=42.0, XV1=2.25, N2=50, XE2=30.0, XV2=3.25 and CL=95.0.

(c) Main program

```
      PROGRAM B3IESU
! *** EXAMPLE OF D3IESU ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER N1,N2,ISW
      REAL(8) XE1,XV1,XE2,XV2,CL,CI(2)
!
      ISW=5
      N1=100
      XE1=42.0D0
      XV1=2.25D0
      N2=50
      XE2=30.0D0
      XV2=3.25D0
      CL=95.0D0
      WRITE(6,1000)
      WRITE(6,2000) ISW
      WRITE(6,*) '  FOR SAMPLE 1'
      WRITE(6,2010) N1
      WRITE(6,2020) XE1
      WRITE(6,2030) XV1
      WRITE(6,*) '  FOR SAMPLE 2'
      WRITE(6,2010) N2
      WRITE(6,2020) XE2
      WRITE(6,2030) XV2
      WRITE(6,2040) CL
      WRITE(6,3000)
      CALL D3IESU(N1,XE1,XV1,N2,XE2,XV2,CL,CI,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) CI(1),CI(2)
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D3IESU ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'ISW = ',I3)
 2010 FORMAT(9X,'N   = ',I3)
 2020 FORMAT(9X,'XE  = ',F4.1)
 2030 FORMAT(9X,'XV  = ',F4.1)
 2040 FORMAT(9X,'CL  = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'INTERVAL = (',D17.10,',',D17.10,')')
      END
```

(d) Output results

```
     *** D3IESU ***
      ** INPUT **
         ISW =   5
  FOR SAMPLE 1
         N   = 100
         XE  = 42.0
         XV  =  2.3
  FOR SAMPLE 2
         N   =  50
         XE  = 30.0
         XV  =  3.3
         CL  = 95.0


      ** OUTPUT **
         IERR =    0
         INTERVAL = ( 0.1140748848D+02, 0.1259251152D+02)
```

### 6.2.4 D3IEVA, R3IEVA Interval Estimation of the Population Variance Due to One Set of Samples

(1) **Function**

The D3IEVA or R3IEVA obtains the confidence interval of the population variance from the variance (or population variance) $\sigma^2$ of one set of sample data of size $n$ when the confidence level $1-\alpha$ is specified. The confidence interval $(t_1, t_2)$ is defined as follows.

$$t_1 = \frac{\sigma^2(n-1)}{{\chi_1}^2}$$

$$t_2 = \frac{\sigma^2(n-1)}{{\chi_2}^2}$$

$\sigma^2$ is an unbiased estimate of the population variance. Also,

$$\frac{\alpha}{2} = P({\chi_1}^2; n-1) = 1 - P({\chi_2}^2; n-1)$$

Here, $P(x, y)$ is the cumulative distribution function (c.d.f.) of a $\chi^2$ distribution with $y$ degrees of freedom.

(2) **Usage**

Double precision:

CALL D3IEVA (N, XV, CL, CI, ISW, IERR)

Single precision:

CALL R3IEVA (N, XV, CL, CI, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Number of sample data $n$ |
| 2 | XV | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Variance of sample data $\sigma^2$ |
| 3 | CL | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Confidence level $100(1-\alpha)(\%)$ |
| 4 | CI | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 2 | Output | CI(1): Confidence interval lower bound $t_1$<br>CI(2): Confidence interval upper bound $t_2$ |
| 5 | ISW | I | 1 | Input | ISW=1: The variance (not an unbiased estimate) of the sample data is entered for XV<br>ISW=2: The variance (unbiased estimate) of the sample data is entered for XV |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{1, 2\}$

(b) N $\geq 2$

(c) XV $> 0.0$

(d) $0.0 <$ CL $\leq 100.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | CL=100.0 | 0.0 is set for CI(1) and the positive maximum value is set for CI(2). |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |

(6) **Notes**

(a) When the confidence interval at a $1 - \alpha$ confidence level for the variance $\sigma^2$ is $(t_1, t_2)$, the confidence interval for the standard deviation is $(\sqrt{t_1}, \sqrt{t_2})$.

(7) **Example**

(a) Problem
Obtain the confidence interval of the population variance at a 95% confidence level when the number of sample data is 25 and the unbiased variance is 29.16.

(b) Input data
ISW=2, N=25, XV=29.16 and CL=95.0.

(c) Main program

```
      PROGRAM B3IEVA
! *** EXAMPLE OF D3IEVA ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER N,ISW
      REAL(8) XV,CL,CI(2)
!
      ISW=2
      N=25
      XV=29.16D0
      CL=95.0D0
      WRITE(6,1000)
      WRITE(6,2000) ISW
      WRITE(6,2010) N
      WRITE(6,2030) XV
      WRITE(6,2040) CL
      WRITE(6,3000)
      CALL D3IEVA(N,XV,CL,CI,ISW,IERR)
      WRITE(6,4000) IERR
      WRITE(6,5000) CI(1),CI(2)
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D3IEVA ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'ISW = ',I3)
 2010 FORMAT(9X,'N   = ',I3)
 2030 FORMAT(9X,'XV  = ',F4.1)
 2040 FORMAT(9X,'CL  = ',F4.1)
 3000 FORMAT(' ',/,/,6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'INTERVAL = (',D17.10,',',D17.10,')')
      END
```

(d) Output results

```
*** D3IEVA ***
 ** INPUT **
    ISW =   2
    N   =  25
    XV  = 29.2
    CL  = 95.0

 ** OUTPUT **
    IERR =    0
    INTERVAL = ( 0.1777864624D+02, 0.5643347494D+02)
```

## 6.2.5 D3IETC, R3IETC Interval Estimation of the Population Correlation Coefficient According to One Set of Samples

(1) **Function**

From the sample correlation coefficient $r$ of one set of sample data of size $n$, obtain the confidence interval of the population correlation coefficient $\rho$ when the confidence level $1-\alpha$ is specified. The confidence interval $(t_1, t_2)$ is defined as follows.

$$t_1 = \frac{e^{a-b}-1}{e^{a-b}+1}$$

$$t_2 = \frac{e^{a+b}-1}{e^{a+b}+1}$$

where $a$, $b$ and $\frac{\alpha}{2}$ are as follows.

$$a = \log_e \frac{1+r}{1-r}$$

$$b = \frac{2z_{\frac{\alpha}{2}}}{\sqrt{n-3}}$$

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

$P(z)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

(2) **Usage**

Double precision:
CALL D3IETC (N, R, CL, T, IERR)
Single precision:
CALL R3IETC (N, R, CL, T, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Number of sample data $n$ |
| 2 | R | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | The sample correlation coefficient $r$ of one set of sample data (See Note (a)) |
| 3 | CL | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Confidence level $100(1-\alpha)(\%)$ |
| 4 | T | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 2 | Output | T(1) : Confidence interval lower bound $t_1$<br>T(2) : Confidence interval upper bound $t_2$ |
| 5 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) N≥2

(b) $-1.0 < \mathrm{R} < 1.0$

(c) 0.0≤CL≤100.0

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | CL=100.0 | $\mathrm{T}(1) = -1.0$ and<br>$\mathrm{T}(2) = 1.0$<br>are performed. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |

(6) **Notes**

(a) The sample correlation coefficient $r$ for $n$ sample data, $\{x_i, y_i\}$ $(i = 1, \cdots, n)$, is defined as follows. (See 4.4.1 $\left\{\begin{matrix}\text{D2CCMT}\\ \text{R2CCMT}\end{matrix}\right\}$)

$$r = \frac{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \mu_x)^2}\sqrt{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(y_i - \mu_y)^2}}$$

(7) **Example**

(a) Problem

From the following one set of sample data of size 10, obtain the confidence interval of the population correlation coefficient with 95% confidence level.

| $x_i$ | $y_i$ |
|---|---|
| 10.129 | 63.4 |
| 12.611 | 60.1 |
| 13.900 | 57.2 |
| 16.532 | 46.5 |
| 20.822 | 43.9 |
| 26.025 | 39.6 |
| 28.283 | 39.7 |
| 29.199 | 39.1 |
| 30.766 | 37.8 |
| 32.664 | 27.8 |

(b) Input data

N=10, CL=95.0 and sample data $\{x_i, y_i\}$.

R : calculate from D2CCMT.

(c) Main program

```
      PROGRAM B3IETC
! *** EXAMPLE OF D3IETC ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER N
      PARAMETER (N=10)
      INTEGER M
      PARAMETER (M=2)
      REAL(8) X(N,2),X1(M),WK(M),RR(N,M)
      INTEGER IERR,NS,KERR,ISW
      REAL(8) R,CL,T(2)
      DATA (X(I,1),I=1,N)&
            /10.129D0, 12.611D0, 13.900D0, 16.532D0, 20.822D0,&
             26.025D0, 28.283D0, 29.199D0, 30.766D0, 32.664D0/,&
           (X(I,2),I=1,N) &
            /63.4D0, 60.1D0, 57.2D0, 46.5D0, 43.9D0,&
             39.6D0, 39.7D0, 39.1D0, 37.8D0, 37.8D0/
!
      CL=95.0D0
!
      ISW=0
!
      CALL D2CCMT(X,N,N,M,NS,X1,RR,N,ISW,WK,KERR)
!
      IF (KERR.NE.0) THEN
        WRITE(6,900)KERR
        STOP
      ENDIF
      R=RR(1,2)
!
      WRITE(6,1000)
      WRITE(6,2010) N
      WRITE(6,2030) CL
      WRITE(6,2100)
      DO 100 I=1,N
         WRITE(6,2110)I,X(I,1),X(I,2)
  100 CONTINUE
      WRITE(6,3000)
!
      CALL D3IETC(N,R,CL,T,IERR)
!
      WRITE(6,4000) IERR
      WRITE(6,5000) R
      WRITE(6,5100) T(1),T(2)
!
      STOP
!
  900 FORMAT(' ',/,5X,'ERROR OCCURED IN D2CCMT',/,1X,'KERR=',I4)
 1000 FORMAT(' ',/,5X,'*** D3IETC ***',/,&
      6X,'** INPUT **')
 2010 FORMAT(9X,'N    = ',I3)
 2030 FORMAT(9X,'CL   = ',F5.1)
 2100 FORMAT(9X,'NO.',1X,'SAMPLE1',1X,'SAMPLE2')
 2110 FORMAT(9X,I3,2F8.3)
 3000 FORMAT(6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'R    = ',F15.8)
 5100 FORMAT(9X,'INTERVAL = (',D18.10,',',D18.10,')')
      END
```

(d) Output results

```
     *** D3IETC ***
      ** INPUT **
         N    =  10
         CL   =  95.0
         NO. SAMPLE1 SAMPLE2
           1  10.129  63.400
           2  12.611  60.100
           3  13.900  57.200
           4  16.532  46.500
           5  20.822  43.900
           6  26.025  39.600
           7  28.283  39.700
           8  29.199  39.100
           9  30.766  37.800
          10  32.664  37.800
      ** OUTPUT **
         IERR =    0
         R    =      -0.94474904
         INTERVAL = ( -0.9871689208D+00, -0.7777713387D+00)
```

### 6.2.6 D3IECD, R3IECD Interval Estimation of the Difference of the Population Correlation Coefficient According to Two Sets of Independent Samples

(1) **Function**

From the correlation coefficients $r_1$ and $r_2$ of two sets of independent sample data of sizes $n_1$ and $n_2$, respectively, obtain the confidence interval of the difference of the population correlation coefficients $\rho_1$ and $\rho_2$ when the confidence level $1-\alpha$ is specified. If $\rho_1 = \rho_2 = \rho$, the confidence level of $\rho$ is obtained.

(a) When $\rho_1 = \rho_2 = \rho$

The confidence interval $(t_1, t_2)$ of $\rho$ is defined as follows.

$$t_1 = \frac{e^{a-b} - 1}{e^{a-b} + 1}$$

$$t_2 = \frac{e^{a+b} - 1}{e^{a+b} + 1}$$

Here,

$$a = \frac{2(n_1 - 3)z_1 + 2(n_2 - 3)z_2}{n_1 + n_2 - 6}$$

$$b = \frac{2z_{\frac{\alpha}{2}}}{\sqrt{n_1 + n_2 - 6}}$$

$$z_1 = \frac{1}{2}\log_e \frac{1 + r_1}{1 - r_1}$$

$$z_2 = \frac{1}{2}\log_e \frac{1 + r_2}{1 - r_2}$$

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

$P(z)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

(b) When $\rho_1 \neq \rho_2$

The confidence interval $(t_1, t_2)$ of $\rho_1 - \rho_2$ is defined as follows.

$$t_1 = \frac{e^{a-b} - 1}{e^{a-b} + 1}$$

$$t_2 = \frac{e^{a+b} - 1}{e^{a+b} + 1}$$

Here,

$$a = \log_e \frac{1 + r_1}{1 - r_1} - \log_e \frac{1 + r_2}{1 - r_2}$$

$$b = 2z_{\frac{\alpha}{2}}\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$$

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

$P(z)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

(2) **Usage**

Double precision:

CALL D3IECD (N1, R1, N2, R2, CL, T, ISW, IERR)

Single precision:

CALL R3IECD (N1, R1, N2, R2, CL, T, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N1 | I | 1 | Input | Number of first sample data $n_1$ |
| 2 | R1 | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | The correlation coefficient $r_1$ of first sample data (See Note (a)) |
| 3 | N2 | I | 1 | Input | Number of second sample data $n_2$ |
| 4 | R2 | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | The correlation coefficient $r_2$ of second sample data (See Note (a)) |
| 5 | CL | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Confidence level $100(1-\alpha)(\%)$ |
| 6 | T | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 3 | Output | T(1) : Confidence interval lower bound $t_1$<br>T(2) : Confidence interval upper bound $t_2$ |
| 7 | ISW | I | 1 | Input | Processing switch<br>ISW=1 : when $\rho_1 = \rho_2$<br>ISW=2 : when $\rho_1 \neq \rho_2$ |
| 8 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{1, 2\}$

(b) N1 $\geq 4$, N2 $\geq 4$

(c) $-1.0 <$ R1 $< 1.0$, $-1.0 <$ R2 $< 1.0$

(d) $0.0 \leq$ CL $\leq 100.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | CL=100.0 | T(1) = −1.0 and T(2) = 1.0 are performed. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |

(6) **Notes**

(a) The sample correlation coefficient $r$ for $n$ sample data, $\{x_i, y_i\}$ $(i = 1, \cdots, n)$, is defined as follows. (See 4.4.1 $\left\{\begin{matrix}\text{D2CCMT}\\ \text{R2CCMT}\end{matrix}\right\}$)

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_y)^2}}$$

(7) **Example**

(a) Problem

For the correlation coefficient $r_1 = 0.8$ obtained from a set of 50 observed values and the correlation coefficient $r_2 = 0.6$ obtained from a set of 15 observed values, test the hypothesis $\rho_1 = \rho_2$ related to the population correlation coefficients of the populations to which the respective observed values belong. Assume that the alternative hypothesis is $\rho_1 \neq \rho_2$.

Moreover, if the hypothesis $\rho_1 = \rho_2$ is accepted, obtain the confidence interval of the correlation coefficient of population $\rho_1 = \rho_2 = \rho$ with 95% confidence level. If the hypothesis $\rho_1 = \rho_2$ is rejected, obtain the difference of the two correlation coefficient of population $\rho_1 - \rho_2$ with 95% confidence level.

(b) Input data

N1=50, R1=0.8, N2=40, R2=0.6 and CL=95.0.

(c) Main program

```
      PROGRAM B3IECD
! *** EXAMPLE OF D3IECD ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERRTS,IERRIE
      INTEGER N1,N2,ISWTS,ISWIE
      REAL(8) R1,R2,CL,T(2),Z(2)
!
      ISWTS=1
      N1=50
      N2=40
      R1=0.8D0
      R2=0.6D0
      CL=95.0D0
!
      WRITE(6,1000)
      WRITE(6,1010)
      WRITE(6,1100) ISWTS
      WRITE(6,2010) N1
      WRITE(6,2015) N2
      WRITE(6,2020) R1
      WRITE(6,2025) R2
      WRITE(6,2030) CL
      WRITE(6,1020)
      WRITE(6,2010) N1
```

```
      WRITE(6,2015) N2
      WRITE(6,2020) R1
      WRITE(6,2025) R2
      WRITE(6,2030) CL
      WRITE(6,3000)
!
      CALL D3TSCD(N1,R1,N2,R2,CL,IR,Z,ISWTS,IERRTS)
!
      WRITE(6,1010)
      WRITE(6,4000) IERRTS
      WRITE(6,5000)
      IF(IR.EQ.0) THEN
         WRITE(6,5010)
      ELSE
         WRITE(6,5020)
      ENDIF
!
      WRITE(6,6000) Z(1),Z(2)
      IF(IR.EQ.0) THEN
         ISWIE=1
      ELSE
         ISWIE=2
      ENDIF
!
      CALL D3IECD(N1,R1,N2,R2,CL,T,ISWIE,IERRIE)
!
      WRITE(6,1020)
      WRITE(6,4010) IERRIE
      WRITE(6,1110) ISWIE
      WRITE(6,7000) T(1),T(2)
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D3IECD ***',/,&
      6X,'** INPUT **')
 1010 FORMAT(9X,'** D3TSCD **')
 1020 FORMAT(9X,'** D3IECD **')
 1100 FORMAT(9X,'ISWTS = ',I3)
 1110 FORMAT(9X,'ISWIE = ',I3)
 2010 FORMAT(9X,'N1    = ',I3)
 2015 FORMAT(9X,'N2    = ',I3)
 2020 FORMAT(9X,'R1    = ',F5.1)
 2025 FORMAT(9X,'R2    = ',F5.1)
 2030 FORMAT(9X,'CL    = ',F5.1)
 3000 FORMAT(6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERRTS = ',I4)
 4010 FORMAT(9X,'IERRIE = ',I4)
 5000 FORMAT(9X,'HYPOTHESIS: RHO1 .EQ. RHO2')
 5010 FORMAT(9X,'HYPOTHESIS IS ACCEPTED.')
 5020 FORMAT(9X,'HYPOTHESIS IS REJECTED.')
 6000 FORMAT(9X,'Z(1)= ',F8.4,/,9X,'Z(2)= ',F8.4)
 7000 FORMAT(9X,'INTERVAL       = (',D18.10,',',D18.10,')')
      END
```

(d) Output results

```
     *** D3IECD ***
      ** INPUT **
         ** D3TSCD **
         ISWTS =   1
         N1    =  50
         N2    =  40
         R1    =   0.8
         R2    =   0.6
         CL    =  95.0
         ** D3IECD **
         N1    =  50
         N2    =  40
         R1    =   0.8
         R2    =   0.6
         CL    =  95.0
      ** OUTPUT **
         ** D3TSCD **
         IERRTS =    0
         HYPOTHESIS: RHO1 .EQ. RHO2
         HYPOTHESIS IS ACCEPTED.
         Z(1)=   1.8449
         Z(2)=   1.9600
         ** D3IECD **
         IERRIE =    0
         ISWIE =   1
         INTERVAL       = (  0.6082663460D+00,  0.8123376258D+00)
```

### 6.2.7 D3IESR, R3IESR Interval Estimation in the Simple Linear Regression

(1) **Function**

For regression coefficient $a$, constant term $b$, and a given specific data $x_0$ in the following simple linear regression expression (or regression line) related to one set of sample data $\{x_i,\, y_i\}$ $(1, \cdots, n)$ of size $n$

$$\hat{y}_i = ax_i + b$$

obtain the estimate $\hat{y}_0$ and the confidence interval of the confidence level $1 - \alpha$ of the theoretical value $Ax_0 - B$. Assume that $y_i$ corresponding to each $x_i$ is the random sample from the normal population having the mean $Ax_i - B$ and the variance $\sigma^2$. Obtain the regression coefficient $a$ and constant term $b$ of the sample data from the following normal equations.

$$\begin{cases} \displaystyle\sum_{i=1}^{n} y_i = a \sum_{i=1}^{n} x_i + bn \\ \displaystyle\sum_{i=1}^{n} x_i y_i = a \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_i \end{cases}$$

The confidence interval $(t_1, t_2)$ is defined as follows.

(a) Regression coefficient

$$t_1 = a - t_{\frac{\alpha}{2}} \cdot s_a$$

$$t_2 = a + t_{\frac{\alpha}{2}} \cdot s_a$$

Here,

$$s_a = \sqrt{\frac{\sigma^2}{\sum(x_i - \mu_x)^2}}$$

i. When the population variance is known
$\sigma^2$ : Population variance
$$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}})$$
Here, $P(t)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

ii. When the population variance is unknown
$\sigma^2$ : Unbiased variance of error variation
$$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}}|n - 2)$$
Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

(b) Constant term

$$t_1 = a - t_{\frac{\alpha}{2}} \cdot s_b$$

$$t_2 = a + t_{\frac{\alpha}{2}} \cdot s_b$$

Here,

$$s_b = \sqrt{\sigma^2 \left[ \frac{1}{n} + \frac{\mu_x^2}{\sum(x_i - \mu_x)^2} \right]}$$

i. When the population variance is known
$\sigma^2$ : Population variance
$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}})$
Here, $P(t)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

ii. When the population variance is unknown
$\sigma^2$ : Unbiased variance of error variation
$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}}|n-2)$
Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

(c) Estimated value

$$t_1 = \hat{y}_0 - t_{\frac{\alpha}{2}} \cdot s_y$$

$$t_2 = \hat{y}_0 + t_{\frac{\alpha}{2}} \cdot s_y$$

Here,

$$s_y = \sqrt{\sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \mu_x)^2}{\sum(x_i - \mu_x)^2}\right]}$$

i. When the population variance is known
$\sigma^2$ : Population variance
$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}})$
Here, $P(t)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

ii. When the population variance is unknown
$\sigma^2$ : Unbiased variance of error variation
$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}}|n-2)$
Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

(d) Theoretical value

$$t_1 = \hat{y}_0 - t_{\frac{\alpha}{2}} \cdot s_0$$

$$t_2 = \hat{y}_0 + t_{\frac{\alpha}{2}} \cdot s_0$$

Here,

$$s_0 = \sqrt{\sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \mu_x)^2}{\sum(x_i - \mu_x)^2}\right]}$$

i. When the population variance is known
$\sigma^2$ : Population variance
$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}})$
Here, $P(t)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

ii. When the population variance is unknown
$\sigma^2$ : Unbiased variance of error variation
$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}}|n-2)$
Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

(2) **Usage**

Double precision:

CALL D3IESR (X, N, Y, YV, X0, CL, T, STAT, ISW1, ISW2, W, IERR)

Single precision:

CALL R3IESR (X, N, Y, YV, X0, CL, T, STAT, ISW1, ISW2, W, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: {INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer}

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | X | {D / R} | N | Input | Values of independent variable X of sample, $x_i (i = 1, n)$ |
| 2 | N | I | 1 | Input | Number of sample data $n$ |
| 3 | Y | {D / R} | N | Input | Values of dependent variable Y of sample, $y_i (i = 1, n)$ |
| 4 | YV | {D / R} | 1 | Input | Variance of the population to which dependent variable Y belongs(if ISW2 = 1) |
| | | | | Output | Unbiased variance of error variation $\sigma^2$ (if ISW2 = 2) (See 10.2.1) |
| 5 | X0 | {D / R} | 1 | Input | Value of $x_0$ to obtain the estimated value or the theoretical value (If ISW1 = 1 or ISW1 = 2, initialization is not necessary.) |
| 6 | CL | {D / R} | 1 | Input | Confidence level $100(1 - \alpha)(\%)$ |
| 7 | T | {D / R} | 2 | Output | T(1) : Confidence interval lower bound $t_1$<br>T(2) : Confidence interval upper bound $t_2$ |
| 8 | STAT | {D / R} | 2 | Output | STAT(1) : The sample regression coefficient<br>STAT(2) : The sample constant term |
| 9 | ISW1 | I | 1 | Input | The switch for selection of the statistic.<br>ISW1=1 : To obtain the confidence interval of the regression coefficient.<br>ISW1=2 : To obtain the confidence interval of the constant term.<br>ISW1=3 : To obtain the confidence interval of the estimated value.<br>ISW1=4 : To obtain the confidence interval of the theoretical value. |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 10 | ISW2 | I | 1 | Input | Switch for variance<br>ISW2=1 : The variance of the population is entered for YV<br>ISW2=2 : The variance (not an unbiased estimate) of the sample data is entered for YV |
| 11 | W | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | 29 | Work | Work area |
| 12 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW1 $\in \{1, 2, 3, 4\}$

(b) ISW2 $\in \{1, 2\}$

(c) N $\geq 3$

(d) $0.0 \leq$ CL $\leq 100.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | CL=100.0 | The negative minimum value is set for T(1) and the positive maximum value is set for T(2). |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 4000 | There are no difference among independent variable X. | |
| 4100 | The Unbiased variance of error variation is 0.0. (See 10.2.1) | |

(6) **Notes**
None

(7) **Example**

(a) Problem

From the following one set of sample data of size 9, obtain the confidence interval regression coefficient of population with 95% confidence level.

| $x_i$ | $y_i$ |
|---|---|
| 1 | 3 |
| 2 | 3 |
| 3 | 5 |
| 4 | 5 |
| 5 | 6 |
| 6 | 7 |
| 7 | 8 |
| 8 | 8 |
| 9 | 9 |

The variance value of the population to which the samples belongs is unknown.

(b) Input data

ISW1=1, ISW2=2 N=9, array X, array Y and CL=95.0.

(c) Main program

```
      PROGRAM B3IESR
! *** EXAMPLE OF D3IESR ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER N
      PARAMETER (N=9)
      REAL(8) X(N),Y(N),W(29)
      INTEGER IERR,ISW1,ISW2
      REAL(8) CL,T(2),STAT(2)
      DATA (X(I),I=1,N) &
            /1.0D0, 2.0D0, 3.0D0, 4.0D0, 5.0D0,&
             6.0D0, 7.0D0, 8.0D0, 9.0D0/,&
           (Y(I),I=1,N) &
            /3.0D0, 3.0D0, 5.0D0, 5.0D0, 6.0D0,&
             7.0D0, 8.0D0, 8.0D0, 9.0D0/
!
      ISW1=1
      ISW2=2
      X0=5.0D0
      CL=95.0D0
!
      WRITE(6,1000)
      WRITE(6,1100)ISW1
      WRITE(6,1150)ISW2
      WRITE(6,2000)N
      WRITE(6,2100)CL
      WRITE(6,2200)X0
      WRITE(6,2500)
      DO 100 I=1,N
         WRITE(6,2510)I,X(I),Y(I)
  100 CONTINUE
!
      CALL D3IESR(X,N,Y,YV,X0,CL,T,STAT,ISW1,ISW2,W,IERR)
!
      WRITE(6,3000)
      WRITE(6,4000) IERR
      WRITE(6,3010)
      WRITE(6,5000) T(1),T(2)
!
!
      WRITE(6,6000) STAT(1)
      WRITE(6,6010) STAT(2)
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D3IESR ***',/,&
      6X,'** INPUT **')
 1100 FORMAT(9X,'ISW1= ',I3)
 1150 FORMAT(9X,'ISW2= ',I3)
 2000 FORMAT(9X,'N   = ',I3)
 2100 FORMAT(9X,'CL  = ',F5.1)
 2200 FORMAT(9X,'X0  = ',F5.1)
 2500 FORMAT(9X,'SAMPLE DATA',/,&
             9X,' I',1X,'X(I)',1X,'Y(I)')
 2510 FORMAT(9X,I2,2F5.1)
 3000 FORMAT(6X,'** OUTPUT **')
```

```
3010 FORMAT(6X,'*** REGRESSION COEFFICIENT ***')
4000 FORMAT(9X,'IERR = ',I4)
5000 FORMAT(9X,'INTERVAL = (',D18.10,',',D18.10,')')
6000 FORMAT(9X,'REGRESSION COEFFICIENT OF SAMPLE= ',D18.10)
6010 FORMAT(9X,'CONSTANT TERM          OF SAMPLE= ',D18.10)
     END
```

(d) Output results

```
*** D3IESR ***
 ** INPUT **
    ISW1=   1
    ISW2=   2
    N   =   9
    CL  =  95.0
    X0  =   5.0
    SAMPLE DATA
     I X(I) Y(I)
     1  1.0  3.0
     2  2.0  3.0
     3  3.0  5.0
     4  4.0  5.0
     5  5.0  6.0
     6  6.0  7.0
     7  7.0  8.0
     8  8.0  8.0
     9  9.0  9.0
 ** OUTPUT **
    IERR =    0
 *** REGRESSION COEFFICIENT ***
    INTERVAL = (  0.6578196590D+00,  0.9088470076D+00)
    REGRESSION COEFFICIENT OF SAMPLE=   0.7833333333D+00
    CONSTANT TERM          OF SAMPLE=   0.2083333333D+01
```

## 6.3 TESTS

### 6.3.1 D3TSRA, R3TSRA
### Test of the Population Ratio According to One Set of Samples

(1) **Function**

When the number of data having the observed characteristic in the one set of sample data of size $n$ is $m$, test the hypothesis $p = p_0$ related to the ratio $p$ in the population to which the respective set of sample data belong with the confidence level $1-\alpha$. The test criteria are as follows.

(a) When the alternative hypothesis is $p \neq p_0$

For $f_1$ and $f_2$ defined as follows

$$f_1 = \frac{2(n-m)p_0}{2(m+1)(1-p_0)}$$

$$f_2 = \frac{2m(1-p_0)}{2(n-m+1)p_0}$$

$$\begin{cases} \text{If } f_1 \geq F_1 \text{ or } f_2 \geq F_2, \text{ reject} \\ \text{If } f_1 < F_1 \text{ and } f_2 < F_2, \text{ accept} \end{cases}$$

Here,

$$\frac{\alpha}{2} = 1 - P(F_1|2(n-m+1), 2m) = 1 - P(F_2|2(m+1), 2(n-m))$$

$P(F|n_1, n_2)$ is the cumulative distribution function (c.d.f.) of a $F$ distribution having numbers of degrees of freedom $n_1$ and $n_2$.

(b) When the alternative hypothesis is $p < p_0$

For $f_1$ defined as follows

$$f_1 = \frac{2(n-m)p_0}{2(m+1)(1-p_0)}$$

$$\begin{cases} \text{If } f_1 \geq F_1^*, \text{ reject} \\ \text{If } f_1 < F_1^*, \text{ accept} \end{cases} \quad \text{Here,}$$

$$\alpha = 1 - P(F_1^*|2(m+1), 2(n-m))$$

$P(F|n_1, n_2)$ is the cumulative distribution function (c.d.f.) of a $F$ distribution having numbers of degrees of freedom $n_1$ and $n_2$.

(c) When the alternative hypothesis is $p > p_0$

For $f_2$ defined as follows

$$f_2 = \frac{2m(1-p_0)}{2(n-m+1)(1-p_0)}$$

$$\begin{cases} \text{If } f_2 \geq F_2^*, \text{ reject} \\ \text{If } f_2 < F_2^*, \text{ accept} \end{cases} \quad \text{Here,}$$

$$\alpha = 1 - P(F_2^*|2(n-m+1), 2m)$$

$P(F|n_1, n_2)$ is the cumulative distribution function (c.d.f.) of a $F$ distribution having numbers of degrees of freedom $n_1$ and $n_2$.

(2) **Usage**

Double precision:

CALL D3TSRA (N, M, CL, P0, IR, F, ISW, IERR)

Single precision:

CALL R3TSRA (N, M, CL, P0, IR, F, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Number of sample data $n$ |
| 2 | M | I | 1 | Input | Number of samples having the characteristic being observed $m$ |
| 3 | CL | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Confidence level $100(1-\alpha)(\%)$ |
| 4 | P0 | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Tested ratio value $p_0$ |
| 5 | IR | I | 1 | Output | Test result<br>IR=0 : Hypothesis $p = p_0$ is accepted<br>IR=1 : Hypothesis $p = p_0$ is rejected |
| 6 | F | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 4 | Output | When ISW=1<br>F(1) : Value of $f_1$<br>F(2) : Value of $f_2$<br>F(3) : Value of $F$ distribution $F_1$<br>F(4) : Value of $F$ distribution $F_2$<br>When ISW=2<br>F(1) : Value of $f_1$<br>F(2) : Value of $F$ distribution $F_1^*$<br>When ISW=3<br>F(1) : Value of $f_2$<br>F(2) : Value of $F$ distribution $F_2^*$ |
| 7 | ISW | I | 1 | Input | The alternative hypothesis switch<br>ISW=1 :<br>When the alternative hypothesis is $p \neq p_0$<br>ISW=2 :<br>When the alternative hypothesis is $p > p_0$<br>ISW=3 :<br>When the alternative hypothesis is $p < p_0$ |
| 8 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{1, 2, 3\}$

(b) N $> 0$

(c) $0 <$ M $<$ N

(d) $0.0 \leq$ P0 $\leq 1.0$

(e) $0.0 \leq$ CL $\leq 100.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | ISW=1 :<br>CL=100.0<br><br>ISW=2 or 3 :<br>CL=0.0 or CL=100.0. | ISW=1 :<br>0.0 or the positive maximum value is set for F(3) and F(4).<br>ISW=2 or 3 :<br>0.0 or the positive maximum value is set for F(2). |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 3040 | Restriction (e) was not satisfied. | |

(6) **Notes**
None

(7) **Example**

(a) Problem
When the number of sample data having the observed characteristic in the independent sample of size 19 is 5, test the hypothesis $p = 0.5$ related to the population ratio $p$ with 95% confidence level. Assume that the alternative hypothesis is $p \neq 0.5$.

(b) Input data
ISW=1, N=19, M=5, P=0.5 and CL=95.0.

(c) Main program

```
      PROGRAM B3TSRA
! *** EXAMPLE OF D3TSRA ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER N,M,IR,ISW
      REAL(8) CL,P0,F(4)
!
      ISW=1
      N=19
      M=5
      P0=0.5D0
      CL=95.0D0
!
      WRITE(6,1000)
      WRITE(6,2000) N
      WRITE(6,2010) M
      WRITE(6,2020) CL
      WRITE(6,2030) ISW
      WRITE(6,3000)
!
      CALL D3TSRA(N,M,CL,P0,IR,F,ISW,IERR)
```

```
!
      WRITE(6,4000) IERR
      WRITE(6,5000) P0
      IF(IR.EQ.0) THEN
         WRITE(6,5010)
      ELSE
         WRITE(6,5020)
      ENDIF
      WRITE(6,6000) F(1),F(3),F(2),F(4)
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D3TSRA ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'N   = ',I3)
 2010 FORMAT(9X,'M   = ',I3)
 2020 FORMAT(9X,'CL  = ',F4.1)
 2030 FORMAT(9X,'ISW = ',I3)
 3000 FORMAT(6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'HYPOTHESIS:  P .EQ.',F5.1)
 5010 FORMAT(9X,'HYPOTHESIS IS ACCEPTED.')
 5020 FORMAT(9X,'HYPOTHESIS IS REJECTED.')
 6000 FORMAT(9X,'F(1) = ',F8.4,2X,'F(3) = ',F8.4,/,&
             9X,'F(2) = ',F8.4,2X,'F(4) = ',F8.4)
      END
```

(d) Output results

```
     *** D3TSRA ***
      ** INPUT **
         N   =  19
         M   =   5
         CL  = 95.0
         ISW =   1
      ** OUTPUT **
         IERR =    0
         HYPOTHESIS:  P .EQ.  0.5
         HYPOTHESIS IS ACCEPTED.
         F(1) =   2.3333  F(3) =   2.4484
         F(2) =   0.3333  F(4) =   3.3110
```

### 6.3.2 D3TSRD, R3TSRD
### Test of the Difference of the Population Ratios According to Two Sets of Independent Samples

(1) **Function**

When the number of data having the observed characteristic in the two sets of independent sample data of sizes $n_1$ and $n_2$ are $m_1$ and $m_2$, respectively, test the hypothesis $p_1 = p_2$ related to the ratios $p_1$ and $p_2$ in the population to which the respective sets of sample data belong with the confidence level $1 - \alpha$.
Let the sample ratios in the two sets of independent samples, respectively, be $\hat{p}_1$ and $\hat{p}_2$ as shown below.

$$\hat{p}_1 = \frac{m_1}{n_1}, \hat{p}_2 = \frac{m_2}{n_2}$$

The test criteria are as follows.

(a) When no continuity correction is performed

i. When the alternative hypothesis is $p_1 \neq p_2$

For $z$ defined as follows

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$\begin{cases} \text{If } |z| \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |z| < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where, $\hat{p}$ and $\frac{\alpha}{2}$ are as follows.

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{m_1 + m_2}{n_1 + n_2}$$

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

ii. When the alternative hypothesis is $p_1 < p_2$

For $z$ defined as follows

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$\begin{cases} \text{If } z \leq -z_\alpha, \text{ reject} \\ \text{If } z > -z_\alpha, \text{ accept} \end{cases}$$

where, $\hat{p}$ and $\alpha$ are as follows.

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{m_1 + m_2}{n_1 + n_2}$$

$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

iii. When the alternative hypothesis is $p_1 > p_2$

For $z$ defined as follows

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$\begin{cases} \text{If } z \geq z_\alpha, \text{ reject} \\ \text{If } z < z_\alpha, \text{ accept} \end{cases}$$

where, $\hat{p}$ and $\alpha$ are as follows.

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{m_1 + m_2}{n_1 + n_2}$$

$\alpha = 1 - P(z_\alpha)$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

(b) When continuity correction is performed

i. When the alternative hypothesis is $p_1 \neq p_2$

For $z$ defined as follows

$$\begin{cases} z = \dfrac{\hat{p}_1 - \hat{p}_2 - 0.5(\frac{1}{n_1} + \frac{1}{n_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} (\text{when } \hat{p}_1 \geq \hat{p}_2) \\ z = \dfrac{\hat{p}_1 - \hat{p}_2 + 0.5(\frac{1}{n_1} + \frac{1}{n_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} (\text{when } \hat{p}_1 < \hat{p}_2) \end{cases}$$

$$\begin{cases} \text{If } |z| \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |z| < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where, $\hat{p}$ and $\frac{\alpha}{2}$ are as follows.

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{m_1 + m_2}{n_1 + n_2}$$

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

ii. When the alternative hypothesis is $p_1 < p_2$

For $z$ defined as follows

$$\begin{cases} z = \dfrac{\hat{p}_1 - \hat{p}_2 - 0.5(\frac{1}{n_1} + \frac{1}{n_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} (\text{when } \hat{p}_1 \geq \hat{p}_2) \\ z = \dfrac{\hat{p}_1 - \hat{p}_2 + 0.5(\frac{1}{n_1} + \frac{1}{n_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} (\text{when } \hat{p}_1 < \hat{p}_2) \end{cases}$$

$$\begin{cases} \text{If } z \leq -z_\alpha, \text{ reject} \\ \text{If } z > -z_\alpha, \text{ accept} \end{cases}$$

where, $\hat{p}$ and $\alpha$ are as follows.

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{m_1 + m_2}{n_1 + n_2}$$

$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

iii. When the alternative hypothesis is $p_1 > p_2$

For $z$ defined as follows

$$\begin{cases} z = \dfrac{\hat{p}_1 - \hat{p}_2 - 0.5(\frac{1}{n_1} + \frac{1}{n_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} (\text{when } \hat{p}_1 \geq \hat{p}_2) \\ z = \dfrac{\hat{p}_1 - \hat{p}_2 + 0.5(\frac{1}{n_1} + \frac{1}{n_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} (\text{when } \hat{p}_1 < \hat{p}_2) \end{cases}$$

$$\begin{cases} \text{If } z \geq z_\alpha, \text{ reject} \\ \text{If } z < z_\alpha, \text{ accept} \end{cases}$$

where, $\hat{p}$ and $\alpha$ are as follows.

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{m_1 + m_2}{n_1 + n_2}$$

$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

(2) **Usage**

Double precision:

CALL D3TSRD (N1, M1, N2, M2, CL, IR, Z, ISW1, ISW2, IERR)

Single precision:

CALL R3TSRD (N1, M1, N2, M2, CL, IR, Z, ISW1, ISW2, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N1 | I | 1 | Input | Number of first sample data $n_1$ |
| 2 | M1 | I | 1 | Input | Number of data having the observed characteristic in first sample data $m_1$ |
| 3 | N2 | I | 1 | Input | Number of second sample data $n_2$ |
| 4 | M2 | I | 1 | Input | Number of data having the observed characteristic in second sample data $m_2$ |
| 5 | CL | $\left\{\begin{matrix}\text{D}\\ \text{R}\end{matrix}\right\}$ | 1 | Input | Confidence level $100(1-\alpha)(\%)$ |
| 6 | IR | I | 1 | Output | Test result<br>IR=0 : Hypothesis $p_1 = p_2$ is accepted<br>IR=1 : Hypothesis $p_1 = p_2$ is rejected |
| 7 | Z | $\left\{\begin{matrix}\text{D}\\ \text{R}\end{matrix}\right\}$ | 2 | Output | When ISW2=1<br>Z(1) : Values of $z$<br>Z(2) : Values of a standard normal distribution $z_{\frac{\alpha}{2}}$<br>When ISW2=2<br>Z(1) : Values of $z$<br>Z(2) : Values of a standard normal distribution $-z_{\alpha}$<br>When ISW2=3<br>Z(1) : Values of $z$<br>Z(2) : Values of a standard normal distribution $z_{\alpha}$ |
| 8 | ISW1 | I | 1 | Input | Processing switch<br>ISW1=1 : Continuity correction is not performed<br>ISW1=2 : Continuity correction is performed |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 9 | ISW2 | I | 1 | Input | The alternative hypothesis switch<br>ISW2=1 :<br>When the alternative hypothesis is $p \neq p_0$<br>ISW2=2 :<br>When the alternative hypothesis is $p > p_0$<br>ISW2=3 :<br>When the alternative hypothesis is $p < p_0$ |
| 10 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW1 $\in \{1, 2\}$

(b) ISW2 $\in \{1, 2, 3\}$

(c) N1 $>$ 0, N2 $>$ 0

(d) 0 $<$ M1 $<$ N1, 0 $<$ M2 $<$ N2

(e) 0.0 $\leq$ CL $\leq$ 100.0

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | ISW2=1 :<br>CL=100.0<br><br>ISW2=2 or 3 :<br>CL=0.0 or CL=100.0. | ISW2=1 :<br>The positive maximum value is set for Z(2).<br>ISW2=2 or 3 :<br>The positive maximum value or the negative minimum value is set for Z(2). |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 3040 | Restriction (e) was not satisfied. | |

(6) **Notes**
None

(7) **Example**

(a) Problem

When the numbers of data having the observed characteristic in the two sets of independent samples of size 200 and 250 are 140 and 150, respectively, test the hypothesis $p_1 = p_2$ related to the population ratios $p_1$ and $p_2$ with 95% confidence level and perform continuity correction. Assume that the alternative hypothesis is $p_1 \neq p_2$.

(b) Input data

ISW1=2, ISW2=1 N1=200, M1=140, N2=250, M2=150 and CL=95.0.

(c) Main program

```
      PROGRAM B3TSRD
! *** EXAMPLE OF D3TSRD ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER N1,N2,M1,M2,IR,ISW1,ISW2
      REAL(8) CL,Z(2)
!
      ISW1=2
      ISW2=1
      N1=200
      M1=140
      N2=250
      M2=150
      CL=95.0D0
!
      WRITE(6,1000)
      WRITE(6,2000) N1
      WRITE(6,2005) M1
      WRITE(6,2010) N2
      WRITE(6,2015) M2
      WRITE(6,2030) CL
      WRITE(6,2040) ISW1
      WRITE(6,2045) ISW2
      WRITE(6,3000)
!
      CALL D3TSRD(N1,M1,N2,M2,CL,IR,Z,ISW1,ISW2,IERR)
!
      WRITE(6,4000) IERR
      WRITE(6,5000)
      IF(IR.EQ.0) THEN
         WRITE(6,5010)
      ELSE
         WRITE(6,5020)
      ENDIF
      WRITE(6,6000) Z(1),Z(2)
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D3TSRD ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'N1  = ',I3)
 2005 FORMAT(9X,'M1  = ',I3)
 2010 FORMAT(9X,'N2  = ',I3)
 2015 FORMAT(9X,'M2  = ',I3)
 2030 FORMAT(9X,'CL  = ',F4.1)
 2040 FORMAT(9X,'ISW1= ',I3)
 2045 FORMAT(9X,'ISW2= ',I3)
 3000 FORMAT(6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'HYPOTHESIS:  P1 .EQ. P2')
 5010 FORMAT(9X,'HYPOTHESIS IS ACCEPTED.')
 5020 FORMAT(9X,'HYPOTHESIS IS REJECTED.')
 6000 FORMAT(9X,'Z(1) = ',F8.4,2X,'Z(2) = ',F8.4)
      END
```

(d) Output results

```
     *** D3TSRD ***
      ** INPUT **
         N1  = 200
         M1  = 140
         N2  = 250
         M2  = 150
         CL  = 95.0
         ISW1=   2
         ISW2=   1
      ** OUTPUT **
         IERR =    0
         HYPOTHESIS:  P1 .EQ. P2
         HYPOTHESIS IS REJECTED.
         Z(1) =   2.1030  Z(2) =   1.9600
```

### 6.3.3 D3TSME, R3TSME Test of the Population Mean According to One Set of Samples

(1) **Function**

From the mean $\mu_x$ and variance (or population variance) $\sigma^2$ of one set of sample data of size $n$, test the hypothesis $\mu = \mu_0$ with the confidence level $1 - \alpha$. The test criteria are as follows.

(a) When the population variance is known

i. When the alternative hypothesis is $\mu \neq \mu_0$

For $z$ defined as follows

$$z = \frac{\mu_x - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

$$\begin{cases} \text{If } |z| \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |z| < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where,

$\sigma^2$ : Population variance

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

$P(z)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

ii. When the alternative hypothesis is $\mu < \mu_0$

For $z$ defined as follows

$$z = \frac{\mu_x - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

$$\begin{cases} \text{If } z \leq -z_\alpha, \text{ reject} \\ \text{If } z > -z_\alpha, \text{ accept} \end{cases}$$

where,

$\sigma^2$ : Population variance

$$\alpha = 1 - P(z_\alpha)$$

$P(z)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

iii. When the alternative hypothesis is $\mu > \mu_0$

For $z$ defined as follows

$$z = \frac{\mu_x - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

$$\begin{cases} \text{If } z \geq z_\alpha, \text{ reject} \\ \text{If } z < z_\alpha, \text{ accept} \end{cases}$$

where,

$\sigma^2$ : Population variance

$$\alpha = 1 - P(z_\alpha)$$

$P(z)$ is the cumulative distribution function (c.d.f.) of a standard normal distribution.

(b) When the population variance is unknown

i. When the alternative hypothesis is $\mu \neq \mu_0$

For $t$ defined as follows

$$t = \frac{\mu_x - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

$$\begin{cases} \text{If } |t| \geq t_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < t_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where,

$\sigma^2$ : Unbiased estimate of population variance

$$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}}|n-1)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

ii. When the alternative hypothesis is $\mu < \mu_0$

For $t$ defined as follows

$$t = \frac{\mu_x - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

$$\begin{cases} \text{If } t \leq -t_{\alpha}, \text{ reject} \\ \text{If } t > -t_{\alpha}, \text{ accept} \end{cases}$$

where,

$\sigma^2$ : Unbiased estimate of population variance

$$\alpha = 1 - P(t_{\alpha}|n-1)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

iii. When the alternative hypothesis is $\mu > \mu_0$ For $t$ defined as follows

$$t = \frac{\mu_x - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

$$\begin{cases} \text{If } t \geq t_{\alpha}, \text{ reject} \\ \text{If } t < t_{\alpha}, \text{ accept} \end{cases}$$

where,

$\sigma^2$ : Unbiased estimate of population variance

$$\alpha = 1 - P(t_{\alpha}|n-1)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

(2) **Usage**

Double precision:

CALL D3TSME (N, XE, XV, CL, XI, IR, Z, ISW1, ISW2, IERR)

Single precision:

CALL R3TSME (N, XE, XV, CL, XI, IR, Z, ISW1, ISW2, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Number of sample data $n$ |
| 2 | XE | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Mean of sample data $\bar{x}$ |
| 3 | XV | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Variance of sample data or population $\sigma^2$ |
| 4 | CL | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Confidence level $100(1-\alpha)(\%)$ |
| 5 | XI | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Tested population mean $\mu_0$ |
| 6 | IR | I | 1 | Output | Test result<br>IR=0 : Hypothesis $\mu=\mu_0$ is accepted<br>IR=1 : Hypothesis $\mu=\mu_0$ is rejected |
| 7 | Z | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 2 | Output | When ISW2=1<br>Z(1) : Value of $z$ or $t$<br>Z(2) : Value of a normal distribution $z_{\frac{\alpha}{2}}$ or value of a $t$ distribution $t_{\frac{\alpha}{2}}$<br>When ISW2=2<br>Z(1) : Value of $z$ or $t$<br>Z(2) : Value of a normal distribution $-z_\alpha$ or value of a $t$ distribution $-t_\alpha$<br>When ISW2=3<br>Z(1) : Value of $z$ or $t$<br>Z(2) : Value of a normal distribution $z_\alpha$ or value of a $t$ distribution $t_\alpha$ |
| 8 | ISW1 | I | 1 | Input | Switch for variance<br>ISW1=1 : The variance of the population is entered for XV<br>ISW1=2 : he variance (not an unbiased estimate) of the sample data is entered for XV<br>ISW1=3 : The variance (unbiased estimate) of the sample data is entered for XV |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 9 | ISW2 | I | 1 | Input | The alternative hypothesis switch<br>ISW2=1 :<br>When the alternative hypothesis is $\mu \neq \mu_0$<br>ISW2=2 :<br>When the alternative hypothesis is $\mu > \mu_0$<br>ISW2=3 :<br>When the alternative hypothesis is $\mu < \mu_0$ |
| 10 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW1 $\in \{1, 2, 3\}$

(b) ISW2 $\in \{1, 2, 3\}$

(c) N $\geq 2$

(d) XV $> 0.0$

(e) $0.0 \leq$ CL $\leq 100.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | ISW2=1 :<br>CL=100.0<br><br>ISW2=2 or 3 :<br>CL=0.0 or CL=100.0. | ISW2=1 :<br>The positive maximum value is set for Z(2).<br>ISW2=2 or 3 :<br>The positive maximum value or the negative minimum value is set for Z(2). |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 3040 | Restriction (e) was not satisfied. | |

(6) **Notes**
None

(7) **Example**

(a) Problem

When the number of sample data is 10, the sample mean is 20.54 and the population variance is 0.08, test the hypothesis $\mu = 20.52$ related to the population mean $\mu$ with 95% confidence level. Assume that the alternative hypothesis is $\mu \neq 20.52$.

(b) Input data

ISW1=1, ISW2=1, N=10, XE=20.54, XV=0.08, XI=20.52 and CL=95.0.

(c) Main program

```
      PROGRAM B3TSME
! *** EXAMPLE OF D3TSME ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER N,ISW1,ISW2,IR
      REAL(8) XE,XV,CL,XI,Z(2)
!
      ISW1=1
      ISW2=1
      N=10
      XE=20.54D0
      XV=0.08D0
      XI=20.52D0
      CL=95.0D0
!
      WRITE(6,1000)
      WRITE(6,1100) ISW1
      WRITE(6,1200) ISW2
      WRITE(6,2010) N
      WRITE(6,2020) XE
      WRITE(6,2030) XV
      WRITE(6,2040) CL
      WRITE(6,3000)
!
      CALL D3TSME(N,XE,XV,CL,XI,IR,Z,ISW1,ISW2,IERR)
!
      WRITE(6,4000) IERR
      WRITE(6,5000) XI
      IF(IR.EQ.0) THEN
         WRITE(6,5010)
      ELSE
         WRITE(6,5020)
      ENDIF
      WRITE(6,6000) Z(1),Z(2)
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D3TSME ***',/,&
      6X,'** INPUT **')
 1100 FORMAT(9X,'ISW1= ',I3)
 1200 FORMAT(9X,'ISW2= ',I3)
 2010 FORMAT(9X,'N    = ',I3)
 2020 FORMAT(9X,'XE   = ',F6.2)
 2030 FORMAT(9X,'XV   = ',F6.2)
 2040 FORMAT(9X,'CL   = ',F6.2)
 3000 FORMAT(6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'HYPOTHESIS:  MU .EQ.',F6.2)
 5010 FORMAT(9X,'HYPOTHESIS IS ACCEPT.')
 5020 FORMAT(9X,'HYPOTHESIS IS REJECT.')
 6000 FORMAT(9X,'Z(1)= ',F8.4,/,9X,'Z(2)= ',F8.4)
      END
```

(d) Output results

```
     *** D3TSME ***
      ** INPUT **
         ISW1=   1
         ISW2=   1
         N   =  10
         XE  =  20.54
         XV  =   0.08
         CL  =  95.00
      ** OUTPUT **
         IERR =    0
         HYPOTHESIS:  MU .EQ. 20.52
         HYPOTHESIS IS ACCEPT.
         Z(1)=   0.2236
         Z(2)=   1.9600
```

### 6.3.4 D3TSSU, R3TSSU
### Test of the Difference of the Population Means According to Two Sets of Independent Samples

(1) **Function**

From the means $\mu_{x_1}$ and $\mu_{x_2}$ and variances (or population variances) $\sigma_1^2$ and $\sigma_2^2$ of two sets of independent sample data of sizes $n_1$ and $n_2$, respectively, test the hypothesis $\mu_1 = \mu_2$ related to the means $\mu_1$ and $\mu_2$ in the population to which the respective sets of sample data belong with the confidence level $1 - \alpha$. The test criteria are as follows.

(a) When the population variances are known

i. When the alternative hypothesis is $\mu_1 \neq \mu_2$ For $z$ defined as follows

$$z = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$\begin{cases} \text{If } |z| \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |z| < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $\frac{\alpha}{2}$ is as follows.

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

$\sigma_1^2$, $\sigma_2^2$ : Population variances of the two sets

ii. When the alternative hypothesis is $\mu_1 < \mu_2$

For $z$ defined as follows

$$z = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$\begin{cases} \text{If } z \leq -z_{\alpha}, \text{ reject} \\ \text{If } z > -z_{\alpha}, \text{ accept} \end{cases}$$

where $\alpha$ is as follows.

$$\alpha = 1 - P(z_{\alpha})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

$\sigma_1^2$, $\sigma_2^2$ : Population variances of the two sets

iii. When the alternative hypothesis is $\mu_1 > \mu_2$

For $z$ defined as follows

$$z = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$\begin{cases} \text{If } z \geq z_{\alpha}, \text{ reject} \\ \text{If } z < z_{\alpha}, \text{ accept} \end{cases}$$

where $\alpha$ is as follows.

$$\alpha = 1 - P(z_{\alpha})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

$\sigma_1^2$, $\sigma_2^2$ : Population variances of the two sets

(b) When the population variances of the two sets are equal and that value is unknown

i. When the alternative hypothesis is $\mu_1 \neq \mu_2$

For $t$ defined as follows

$$t = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$\begin{cases} \text{If } |t| \geq t_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < t_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $s_p^2$ and $\frac{\alpha}{2}$ are as follows.

$$s_p^2 = \frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}$$

$$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}} | n_1 + n_2 - 2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom. $\sigma_1^2$, $\sigma_2^2$ : Unbiased estimates of the population variances.

ii. When the alternative hypothesis is $\mu_1 < \mu_2$

For $t$ defined as follows

$$t = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$\begin{cases} \text{If } t \leq -t_{\alpha}, \text{ reject} \\ \text{If } t > -t_{\alpha}, \text{ accept} \end{cases}$$

where $s_p^2$ and $\alpha$ are as follows.

$$s_p^2 = \frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}$$

$$\alpha = 1 - P(t_{\alpha} | n_1 + n_2 - 2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

$\sigma_1^2$, $\sigma_2^2$ : Unbiased estimates of the population variances.

iii. When the alternative hypothesis is $\mu_1 > \mu_2$

For $t$ defined as follows

$$t = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$\begin{cases} \text{If } t \geq t_{\alpha}, \text{ reject} \\ \text{If } t < t_{\alpha}, \text{ accept} \end{cases}$$

where $s_p^2$ and $\alpha$ are as follows.

$$s_p^2 = \frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}$$

$$\alpha = 1 - P(t_{\alpha} | n_1 + n_2 - 2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

$\sigma_1^2$, $\sigma_2^2$ : Unbiased estimates of the population variances.

(c) When the population variances of the two sets are not equal and those values are unknown

i. When the alternative hypothesis is $\mu_1 \neq \mu_2$

For $t$ defined as follows

$$t = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$t^*_{\frac{\alpha}{2}} = \frac{\beta_1 t^{(1)}_{\frac{\alpha}{2}} + \beta_2 t^{(2)}_{\frac{\alpha}{2}}}{\beta_1 + \beta_2}$$

$$\begin{cases} \text{If } |t| \geq t^*_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < t^*_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $\frac{\alpha}{2}$ is as follows.

$$\frac{\alpha}{2} = 1 - P(t^{(1)}_{\frac{\alpha}{2}} | n_1 - 1) = 1 - P(t^{(2)}_{\frac{\alpha}{2}} | n_2 - 1)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

$$\beta_1 = \frac{\sigma_1^2}{n_1}, \beta_2 = \frac{\sigma_2^2}{n_2}$$

$\sigma_1^2$, $\sigma_2^2$ : Unbiased estimates of the population variances

ii. When the alternative hypothesis is $\mu_1 < \mu_2$

For $t$ defined as follows

$$t = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$t_\alpha^* = \frac{\beta_1 t_\alpha^{(1)} + \beta_2 t_\alpha^{(2)}}{\beta_1 + \beta_2}$$

$$\begin{cases} \text{If } t \leq -t_\alpha^*, \text{ reject} \\ \text{If } t > -t_\alpha^*, \text{ accept} \end{cases}$$

where $\alpha$ is as follows.

$$\alpha = 1 - P(t_\alpha^{(1)}|n_1 - 1) = 1 - P(t_\alpha^{(2)}|n_2 - 1)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

$$\beta_1 = \frac{\sigma_1^2}{n_1}, \beta_2 = \frac{\sigma_2^2}{n_2}$$

$\sigma_1^2$, $\sigma_2^2$ : Unbiased estimates of the population variances

iii. When the alternative hypothesis is $\mu_1 > \mu_2$

For $t$ defined as follows

$$t = \frac{\mu_{x_1} - \mu_{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$t_\alpha^* = \frac{\beta_1 t_\alpha^{(1)} + \beta_2 t_\alpha^{(2)}}{\beta_1 + \beta_2}$$

$$\begin{cases} \text{If } t \geq t_\alpha^*, \text{ reject} \\ \text{If } t < t_\alpha^*, \text{ accept} \end{cases}$$

where $\alpha$ is as follows.

$$\alpha = 1 - P(t_\alpha^{(1)}|n_1 - 1) = 1 - P(t_\alpha^{(2)}|n_2 - 1)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

$$\beta_1 = \frac{\sigma_1^2}{n_1}, \beta_2 = \frac{\sigma_2^2}{n_2}$$

$\sigma_1^2$, $\sigma_2^2$ : Unbiased estimates of the population variances

(2) **Usage**

Double precision:

CALL D3TSSU (N1, XE1, XV1, N2, XE2, XV2, CL, IR, Z, ISW1, ISW2, IERR)

Single precision:

CALL R3TSSU (N1, XE1, XV1, N2, XE2, XV2, CL, IR, Z, ISW1, ISW2, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N1 | I | 1 | Input | Number of sample data $n_1$ |
| 2 | XE1 | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Mean of sample data $\mu_{x_1}$ |
| 3 | XV1 | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Variance of sample data or population ${\sigma_1}^2$ |
| 4 | N2 | I | 1 | Input | Number of sample data $n_2$ |
| 5 | XE2 | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Mean of sample data $\mu_{x_2}$ |
| 6 | XV2 | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Variance of sample data or population ${\sigma_2}^2$ |
| 7 | CL | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Input | Confidence level $100(1-\alpha)(\%)$ |
| 8 | IR | I | 1 | Output | Test result<br>IR=0 : Hypothesis $\mu_1 = \mu_2$ is accepted<br>IR=1 : Hypothesis $\mu_1 = \mu_2$ is rejected |
| 9 | Z | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 2 | Output | When ISW2=1<br>Z(1) : Value of $z$ or $t$<br>Z(2) : Value of a standard normal distribution $z_{\frac{\alpha}{2}}$ or value of $t$ distribution $t_{\frac{\alpha}{2}}$ or $t^*_{\frac{\alpha}{2}}$<br>When ISW2=2<br>Z(1) : Value of $z$ or $t$<br>Z(2) : Value of a standard normal distribution $-z_\alpha$ or value of $t$ distribution $-t_\alpha$ or $-t^*_\alpha$<br>When ISW2=3<br>Z(1) : Value of $z$ or $t$<br>Z(2) : Value of a standard normal distribution $z_\alpha$ or value of $t$ distribution $t_\alpha$ or $t^*_\alpha$ |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 10 | ISW1 | I | 1 | Input | Switch for variance<br>ISW1=1 : The population variances of the two sets are entered for XV1 and XV2<br>ISW1=2 : The population variances of the two sets are equal, and the variances (not unbiased estimates) of the sample data are entered for XV1 and XV2<br>ISW1=3 : The population variances of the two sets are equal, and the variances (unbiased estimates) of the sample data are entered for XV1 and XV2<br>ISW1=4 : The population variances of the two sets are not equal, and the variances (not unbiased estimates) of the sample data are entered for XV1 and XV2<br>ISW1=5 : The population variances of the two sets are not equal, and the variances (unbiased estimates) of the sample data are entered for XV1 and XV2 |
| 11 | ISW2 | I | 1 | Input | The alternative hypothesis switch<br>ISW2=1 :<br>When the alternative hypothesis is $\mu_1 \neq \mu_2$<br>ISW2=2 :<br>When the alternative hypothesis is $\mu_1 < \mu_2$<br>ISW2=3 :<br>When the alternative hypothesis is $\mu_1 > \mu_2$ |
| 12 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW1 $\in \{1, 2, 3, 4, 5\}$

(b) ISW2 $\in \{1, 2, 3\}$

(c) N1$\geq$ 2, N2$\geq$ 2

(d) XV1$\geq$ 0.0, XV2$\geq$ 0.0

(e) 0.0 $\leq$ CL $\leq$ 100.0

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | ISW2=1 :<br>CL=100.0<br><br>ISW2=2 or 3 :<br>CL=0.0 or CL=100.0. | ISW2=1 :<br>The positive maximum value is set for Z(2).<br>ISW2=2 or 3 :<br>The positive maximum value or the negative minimum value is set for Z(2). |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 3040 | Restriction (e) was not satisfied. | |

(6) **Notes**

None

(7) **Example**

(a) Problem

When the number of sample data, the sample mean and the population variance in the two sets of independent samples are 20, 62.0, 64.0 and 25, 67.0, 81.0, respectively, test the hypothesis $\mu_1 = \mu_2$ related to the each population mean $\mu_1$ and $\mu_2$ with 95% confidence level. Assume that the alternative hypothesis is $\mu_1 \neq \mu_2$.

(b) Input data

ISW1=1, ISW2=1, N1=20, XE1=62.0, XV1=64.0, N2=25, XE2=67.0, XV2=81.0 and CL=95.0.

(c) Main program

```
      PROGRAM B3TSSU
! *** EXAMPLE OF D3TSSU ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER N1,N2,ISW1,ISW2,IR
      REAL(8) XE1,XE2,XV1,XV2,CL,Z(2)
!
      ISW1=1
      ISW2=1
      N1=20
      XE1=62.0D0
      XV1=64.0D0
      N2=25
      XE2=67.0D0
      XV2=81.0D0
      CL=95.0D0
!
      WRITE(6,1000)
      WRITE(6,2000) ISW1
      WRITE(6,2005) ISW2
      WRITE(6,2010) N1
      WRITE(6,2020) XE1
      WRITE(6,2030) XV1
      WRITE(6,2015) N2
      WRITE(6,2025) XE2
      WRITE(6,2035) XV2
      WRITE(6,2050) CL
      WRITE(6,3000)
!
      CALL D3TSSU(N1,XE1,XV1,N2,XE2,XV2,CL,IR,Z,ISW1,ISW2,IERR)
!
      WRITE(6,4000) IERR
      WRITE(6,5000)
      IF(IR.EQ.0) THEN
         WRITE(6,5010)
      ELSE
```

```
            WRITE(6,5020)
         ENDIF
         WRITE(6,6000) Z(1),Z(2)
!
         STOP
!
    1000 FORMAT(' ',/,5X,'*** D3TSSU ***',/,&
         6X,'** INPUT **')
    2000 FORMAT(9X,'ISW1= ',I3)
    2005 FORMAT(9X,'ISW2= ',I3)
    2010 FORMAT(9X,'N1  = ',I3)
    2015 FORMAT(9X,'N2  = ',I3)
    2020 FORMAT(9X,'XE1 = ',F4.1)
    2025 FORMAT(9X,'XE2 = ',F4.1)
    2030 FORMAT(9X,'XV1 = ',F4.1)
    2035 FORMAT(9X,'XV2 = ',F4.1)
    2050 FORMAT(9X,'CL  = ',F4.1)
    3000 FORMAT(6X,'** OUTPUT **')
    4000 FORMAT(9X,'IERR = ',I4)
    5000 FORMAT(9X,'HYPOTHESIS: MU1 .EQ. M2')
    5010 FORMAT(9X,'HYPOTHESIS IS ACCEPTED.')
    5020 FORMAT(9X,'HYPOTHESIS IS REJECTED.')
    6000 FORMAT(9X,'Z(1)= ',F10.4,/,9X,'Z(2)= ',F10.4)
         END
```

(d) Output results

```
     *** D3TSSU ***
      ** INPUT **
         ISW1=   1
         ISW2=   1
         N1  =  20
         XE1 = 62.0
         XV1 = 64.0
         N2  =  25
         XE2 = 67.0
         XV2 = 81.0
         CL  = 95.0
      ** OUTPUT **
         IERR =    0
         HYPOTHESIS: MU1 .EQ. M2
         HYPOTHESIS IS REJECTED.
         Z(1)=    -1.9703
         Z(2)=     1.9600
```

### 6.3.5 D3TSVA, R3TSVA Test of the Population Variance Due to One Set of Samples

(1) **Function**

From the variance (or population variance) $s^2$ of one set of sample data of size $n$, test the hypothesis $\sigma^2 = \sigma_0^2$ related to the population variance $\sigma^2$ in the population to which the respective set of sample data belong with the confidence level $1 - \alpha$. The test criteria are as follows.

(a) When the alternative hypothesis is $\sigma^2 \neq \sigma_0^2$

For $\chi^2$ defined as follows

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

$$\begin{cases} \text{If } \chi^2 \leq \chi^2_{1-\frac{\alpha}{2}} \text{ or } \chi^2 \geq \chi^2_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } \chi^2_{1-\frac{\alpha}{2}} < \chi^2 < \chi^2_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $\frac{\alpha}{2}$ is as follows.

$$\frac{\alpha}{2} = 1 - P(\chi^2_{\frac{\alpha}{2}}|n-1) = P(\chi^2_{1-\frac{\alpha}{2}}|n-1)$$

$P(\chi^2|n)$ is the cumulative distribution function (c.d.f.) of a $\chi^2$ distribution having number of degrees of freedom $n$.

$s^2$ : Unbiased estimate of population variance

(b) When the alternative hypothesis is $\sigma^2 < \sigma_0^2$

For $\chi^2$ defined as follows

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

$$\begin{cases} \text{If } \chi^2 \leq \chi^2_{1-\alpha}, \text{ reject} \\ \text{If } \chi^2 > \chi^2_{1-\alpha}, \text{ accept} \end{cases}$$

where $\alpha$ is as follows.

$$\alpha = P(\chi^2_{1-\alpha}|n-1)$$

$P(\chi^2|n)$ is the cumulative distribution function (c.d.f.) of a $\chi^2$ distribution having number of degrees of freedom $n$.

$s^2$ : Unbiased estimate of population variance

(c) When the alternative hypothesis is $\sigma^2 > \sigma_0^2$

For $\chi^2$ defined as follows

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

$$\begin{cases} \text{If } \chi^2 \geq \chi^2_{\alpha}, \text{ reject} \\ \text{If } \chi^2 < \chi^2_{\alpha}, \text{ accept} \end{cases}$$

where $\alpha$ is as follows.

$$\alpha = P(\chi^2_{\alpha}|n-1)$$

$P(\chi^2|n)$ is the cumulative distribution function (c.d.f.) of a $\chi^2$ distribution having number of degrees of freedom $n$.

$s^2$ : Unbiased estimate of population variance

(2) **Usage**

Double precision:

CALL D3TSVA (N, XV, CL, XI, IR, Z, ISW1, ISW2, IERR)

Single precision:

CALL R3TSVA (N, XV, CL, XI, IR, Z, ISW1, ISW2, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: {INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer}
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Number of sample data $n$ |
| 2 | XV | {D / R} | 1 | Input | Variance of sample data $s^2$ |
| 3 | CL | {D / R} | 1 | Input | Confidence level $100(1-\alpha)(\%)$ |
| 4 | XI | {D / R} | 1 | Input | Tested variance $\sigma_0$ |
| 5 | IR | I | 1 | Output | Test result<br>IR=0 : Hypothesis $\sigma^2 = \sigma_0^2$ is accepted<br>IR=1 : Hypothesis $\sigma^2 = \sigma_0^2$ is rejected |
| 6 | Z | {D / R} | 3 | Output | Z(1) : Value of $\chi^2$<br>Z(2) : Value of $\chi^2$ distribution $\chi^2_{\frac{\alpha}{2}}$<br>Z(3) : Value of $\chi^2$ distribution $\chi^2_{1-\frac{\alpha}{2}}$ |
| 7 | ISW1 | I | 1 | Input | ISW1=1 : The variance (not an unbiased estimate) of the sample data is entered for XV<br>ISW1=2 : The variance (unbiased estimate) of the sample data is entered for XV |
| 8 | ISW2 | I | 1 | Input | The alternative hypothesis switch<br>ISW2=1 :<br>When the alternative hypothesis is $\sigma^2 \neq \sigma_0^2$<br>ISW2=2 :<br>When the alternative hypothesis is $\sigma^2 < \sigma_0^2$<br>ISW2=3 :<br>When the alternative hypothesis is $\sigma^2 > \sigma_0^2$ |
| 9 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW1 $\in \{1, 2\}$

(b) ISW2 $\in \{1, 2, 3\}$

(c) N $\geq 2$

(d) XV $> 0.0$

(e) XI $> 0.0$

(f) $0.0 \leq$ CL $\leq 100.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | ISW2=1 :<br>CL=100.0<br><br>ISW2=2 or 3 :<br>CL=0.0 or CL=100.0. | ISW2=1 :<br>The positive maximum value is set for Z(2).<br>ISW2=2 or 3 :<br>0.0 or the positive maximum value is set for Z(2). |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 3040 | Restriction (e) was not satisfied. | |

(6) **Notes**
None

(7) **Example**

(a) Problem
When the number of sample data is 25 and the sample variance (unbiased estimate) is 182.25, test the hypothesis $\sigma^2 = 100.0$ related to the population variance with 95% confidence level. Assume that the alternative hypothesis is $\sigma^2 \neq 100.0$.

(b) Input data
ISW1=2, ISW2=1, N=25, XV=182.25, XI=100.0 and CL=95.0.

(c) Main program

```
      PROGRAM B3TSVA
! *** EXAMPLE OF D3TSVA ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER IERR
      INTEGER N,ISW1,ISW2,IR
      REAL(8) XV,CL,XI,Z(3)
!
      ISW1=2
      ISW2=1
      N=25
      XV=182.25
      XI=100
      CL=95
!
      WRITE(6,1000)
      WRITE(6,2000) ISW1
      WRITE(6,2005) ISW2
      WRITE(6,2010) N
```

```
      WRITE(6,2030) XV
      WRITE(6,2040) CL
      WRITE(6,3000)
!
      CALL D3TSVA(N,XV,CL,XI,IR,Z,ISW1,ISW2,IERR)
!
      WRITE(6,4000) IERR
      WRITE(6,5000) XI
      IF(IR.EQ.0) THEN
         WRITE(6,5010)
      ELSE
         WRITE(6,5020)
      ENDIF
      WRITE(6,6000) Z(1),Z(2),Z(3)
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D3TSVA ***',/,&
      6X,'** INPUT **')
 2000 FORMAT(9X,'ISW1= ',I3)
 2005 FORMAT(9X,'ISW2= ',I3)
 2010 FORMAT(9X,'N    = ',I3)
 2030 FORMAT(9X,'XV   = ',F9.4)
 2040 FORMAT(9X,'CL   = ',F9.4)
 3000 FORMAT(6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'HYPOTHESIS:  SIGMA2 .EQ.',F5.1)
 5010 FORMAT(9X,'HYPOTHESIS IS ACCEPTED.')
 5020 FORMAT(9X,'HYPOTHESIS IS REJECTED.')
 6000 FORMAT(9X,'Z(1)= ',F8.4,/,&
             9X,'Z(2)= ',F8.4,2X,'Z(3)= ',F8.4)
      END
```

(d) Output results

```
     *** D3TSVA ***
      ** INPUT **
         ISW1=   2
         ISW2=   1
         N    =  25
         XV   =  182.2500
         CL   =   95.0000
      ** OUTPUT **
         IERR =    0
         HYPOTHESIS:  SIGMA2 .EQ.100.0
         HYPOTHESIS IS REJECTED.
         Z(1)=  43.7400
         Z(2)=  39.3641  Z(3)=  12.4012
```

### 6.3.6 D3TSTC, R3TSTC Test of the Population Correlation Coefficient According to One Set of Samples

(1) **Function**

From the sample correlation coefficient $r$ of one set of sample data of size $n$, test the hypothesis $\rho = \rho_0$ related to the population correlation coefficient $\rho$ in the population to which the respective set of sample data belong with the confidence level $1 - \alpha$. The test criteria are as follows.

(a) Hypothesis: $\rho = 0$

i. When the alternative hypothesis is $\rho \neq 0$

For $t$ defined as follows

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

$$\begin{cases} \text{If } |t| \geq t_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < t_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where, $\frac{\alpha}{2}$ is as follows.

$$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}}|n-2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

ii. When the alternative hypothesis is $\rho < 0$

For $t$ defined as follows

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

$$\begin{cases} \text{If } t \geq -t_{\alpha}, \text{ reject} \\ \text{If } t < -t_{\alpha}, \text{ accept} \end{cases}$$

where, $\frac{\alpha}{2}$ is as follows.

$$\frac{\alpha}{2} = 1 - P(t_{\alpha}|n-2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

iii. When the alternative hypothesis is $\rho > 0$

For $t$ defined as follows

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

$$\begin{cases} \text{If } t \geq t_{\alpha}, \text{ reject} \\ \text{If } t < t_{\alpha}, \text{ accept} \end{cases}$$

where, $\frac{\alpha}{2}$ is as follows.

$$\frac{\alpha}{2} = 1 - P(t_{\alpha}|n-2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution with $n$ degrees of freedom.

(b) Hypothesis: $\rho = \rho_0$

i. When the alternative hypothesis is $\rho \neq \rho_0$

For $t$ defined as follows

$$t = (z - z_0)\sqrt{n-3}$$

$$\begin{cases} \text{If } |t| \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where, $z$, $z_0$ and $\frac{\alpha}{2}$ are as follows.

$$z = \frac{1}{2}\log_e \frac{1+r}{1-r}$$

$$z_0 = \frac{1}{2}\log_e \frac{1+\rho_0}{1-\rho_0}$$

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

ii. When the alternative hypothesis is $\rho < \rho_0$

For $t$ defined as follows

$$t = (z - z_0)\sqrt{n-3}$$

$$\begin{cases} \text{If } t \leq -z_{\alpha}, \text{ reject} \\ \text{If } t > -z_{\alpha}, \text{ accept} \end{cases}$$

where, $z$, $z_0$ and $\frac{\alpha}{2}$ are as follows.

$$z = \frac{1}{2}\log_e \frac{1+r}{1-r}$$

$$z_0 = \frac{1}{2}\log_e \frac{1+\rho_0}{1-\rho_0}$$

$$\frac{\alpha}{2} = 1 - P(z_{\alpha})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

iii. When the alternative hypothesis is $\rho > \rho_0$

For $t$ defined as follows

$$t = (z - z_0)\sqrt{n-3}$$

$$\begin{cases} \text{If } t \geq z_{\alpha}, \text{ reject} \\ \text{If } t < z_{\alpha}, \text{ accept} \end{cases}$$

where, $z$, $z_0$ and $\frac{\alpha}{2}$ are as follows.

$$z = \frac{1}{2}\log_e \frac{1+r}{1-r}$$

$$z_0 = \frac{1}{2}\log_e \frac{1+\rho_0}{1-\rho_0}$$

$$\frac{\alpha}{2} = 1 - P(z_{\alpha})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

(2) **Usage**

Double precision:

CALL D3TSTC (N, R, CL, R0, IR, Z, ISW, IERR)

Single precision:

CALL R3TSTC (N, R, CL, R0, IR, Z, ISW, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\\text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Number of sample data $n$ |
| 2 | R | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | The sample correlation coefficient $r$ of one set of sample data (See Note (a)) |
| 3 | CL | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Confidence level $100(1-\alpha)$ |
| 4 | R0 | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Input | Value of tested correlation coefficient $\rho_0$ |
| 5 | IR | I | 1 | Output | Test result<br>IR=0 : Hypothesis $\rho = \rho_0$ is accepted<br>IR=1 : Hypothesis $\rho = \rho_0$ is rejected |
| 6 | Z | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 2 | Output | When ISW=1<br>Z(1) : Value of $t$<br>Z(2) : Value of $t$ distribution $t_{\frac{\alpha}{2}}$ or value of a standard normal distribution $z_{\frac{\alpha}{2}}$<br>When ISW=2<br>Z(1) : Value of $t$<br>Z(2) : Value of $t$ distribution $-t_\alpha$ or value of a standard normal distribution $-z_\alpha$ |
| 7 | ISW | I | 1 | Input | The alternative hypothesis switch<br>ISW=1 :<br>When the alternative hypothesis is $\rho \neq \rho_0$<br>ISW=2 :<br>When the alternative hypothesis is $\rho < \rho_0$<br>ISW=3 :<br>When the alternative hypothesis is $\rho > \rho_0$ |
| 8 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in \{1, 2, 3\}$

(b) N$\geq$4

(c) $-1.0 <$ R $< 1.0$

(d) $-1.0 <$ R0 $< 1.0$

(e) $0.0 \leq$ CL $\leq 100.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | ISW=1 :<br>CL=100.0<br><br>ISW=2 or 3 :<br>CL=0.0 or CL=100.0. | ISW=1 :<br>The positive maximum value is set for Z(2).<br>ISW=2 or 3 :<br>The positive maximum value or the negative minimum value is set for Z(2). |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 3040 | Restriction (e) was not satisfied. | |

(6) **Notes**

(a) The sample correlation coefficient $r$ for $n$ sample data, $\{x_i, y_i\}$ $(i = 1, \cdots, n)$, is defined as follows. (See 4.4.1 $\left\{\begin{array}{c}\text{D2CCMT}\\ \text{R2CCMT}\end{array}\right\}$)

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_y)^2}}$$

(7) **Example**

(a) Problem

From the following one set of sample data of size 10, test the hypothesis $\rho = 0.0$ related to the population correlation coefficient $\rho$ with 95% confidence level.

| $x_i$ | $y_i$ |
|---|---|
| 10.129 | 63.4 |
| 12.611 | 60.1 |
| 13.900 | 57.2 |
| 16.532 | 46.5 |
| 20.822 | 43.9 |
| 26.025 | 39.6 |
| 28.283 | 39.7 |
| 29.199 | 39.1 |
| 30.766 | 37.8 |
| 32.664 | 27.8 |

Assume that the alternative hypothesis is $\rho \neq 0.0$.

(b) Input data

N=10, R0=0.0 and CL=95.0.

(c) Main program

```
      PROGRAM B3TSTC
! *** EXAMPLE OF D3TSTC ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER N
      PARAMETER (N=10)
      INTEGER M
      PARAMETER (M=2)
      REAL(8) X(N,2),X1(M),WK(M),RR(N,M)
      INTEGER IERR,ISW,IR,NS
      REAL(8) R,R0,CL,Z(2)
      DATA (X(I,1),I=1,N)&
            /10.129D0, 12.611D0, 13.900D0, 16.532D0, 20.822D0,&
             26.025D0, 28.283D0, 29.199D0, 30.766D0, 32.664D0/,&
           (X(I,2),I=1,N) &
            /63.4D0, 60.1D0, 57.2D0, 46.5D0, 43.9D0,&
             39.6D0, 39.7D0, 39.1D0, 37.8D0, 37.8D0/
!
      ISW=1
      R0=0.0D0
      CL=95.0D0
!
      CALL D2CCMT(X,N,N,M,NS,X1,RR,N,0,WK,IERR)
      R=RR(1,2)
!
      WRITE(6,1000)
      WRITE(6,1100) ISW
      WRITE(6,2010) N
      WRITE(6,2020) R
      WRITE(6,2030) CL
      WRITE(6,2100)
      DO 100 I=1,N
         WRITE(6,2110)I,X(I,1),X(I,2)
  100 CONTINUE
      WRITE(6,3000)
!
      CALL D3TSTC(N,R,CL,R0,IR,Z,ISW,IERR)
!
      WRITE(6,4000) IERR
      WRITE(6,5000) R0
      IF(IR.EQ.0) THEN
         WRITE(6,5010)
      ELSE
         WRITE(6,5020)
      ENDIF
      WRITE(6,6000) Z(1),Z(2)
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D3TSTC ***',/,&
      6X,'** INPUT **')
 1100 FORMAT(9X,'ISW = ',I3)
 2010 FORMAT(9X,'N   = ',I3)
 2020 FORMAT(9X,'R   = ',F5.1)
 2030 FORMAT(9X,'CL  = ',F5.1)
 2100 FORMAT(9X,'NO.',1X,'SAMPLE1',1X,'SAMPLE2')
 2110 FORMAT(9X,I3,2F8.3)
 3000 FORMAT(6X,'** OUTPUT **')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'HYPOTHESIS:  RHO .EQ.',F5.1)
 5010 FORMAT(9X,'HYPOTHESIS IS ACCEPTED.')
 5020 FORMAT(9X,'HYPOTHESIS IS REJECTED.')
 6000 FORMAT(9X,'Z(1)= ',F8.4,/,9X,'Z(2)= ',F8.4)
      END
```

(d) Output results

```
     *** D3TSTC ***
      ** INPUT **
         ISW =   1
         N   =  10
         R   =  -0.9
         CL  =  95.0
         NO. SAMPLE1 SAMPLE2
           1  10.129  63.400
           2  12.611  60.100
           3  13.900  57.200
           4  16.532  46.500
           5  20.822  43.900
           6  26.025  39.600
           7  28.283  39.700
           8  29.199  39.100
           9  30.766  37.800
          10  32.664  37.800
      ** OUTPUT **
         IERR =    0
         HYPOTHESIS:  RHO .EQ.  0.0
         HYPOTHESIS IS REJECTED.
         Z(1)=  -8.1519
         Z(2)=   2.3060
```

### 6.3.7 D3TSCD, R3TSCD Test of the Difference of the Population Correlation Coefficients According to Two Sets of Independent Samples

(1) **Function**

From the correlation coefficients $r_1$ and $r_2$ of two sets of independent sample data of sizes $n_1$ and $n_2$, respectively, test the hypothesis $\rho_1 = \rho_2$ related to the population correlation coefficients $\rho_1$ and $\rho_2$ in the population to which the respective set of sample data belong with the confidence level $1 - \alpha$. The test criteria are as follows.

(a) When the alternative hypothesis is $\rho_1 \neq \rho_2$

For $t$ defined as follows

$$t = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$$

$$\begin{cases} \text{If } |t| \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where, $z_1$, $z_2$ and $\frac{\alpha}{2}$ are as follows.

$$z_1 = \frac{1}{2} \log_e \frac{1+r_1}{1-r_1}$$

$$z_2 = \frac{1}{2} \log_e \frac{1+r_2}{1-r_2}$$

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

(b) When the alternative hypothesis is $\rho_1 < \rho_2$

For $t$ defined as follows

$$t = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$$

$$\begin{cases} \text{If } t \leq -z_{\alpha}, \text{ reject} \\ \text{If } t > -z_{\alpha}, \text{ accept} \end{cases}$$

where, $z_1$, $z_2$ and $\alpha$ are as follows.

$$z_1 = \frac{1}{2} \log_e \frac{1+r_1}{1-r_1}$$

$$z_2 = \frac{1}{2} \log_e \frac{1+r_2}{1-r_2}$$

$$\alpha = 1 - P(z_{\alpha})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

(c) When the alternative hypothesis is $\rho_1 > \rho_2$

For $t$ defined as follows

$$t = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$$

$$\begin{cases} \text{If } t \geq z_\alpha, \text{ reject} \\ \text{If } t < z_\alpha, \text{ accept} \end{cases}$$

where, $z_1$, $z_2$ and $\alpha$ are as follows.

$$z_1 = \frac{1}{2} \log_e \frac{1+r_1}{1-r_1}$$

$$z_2 = \frac{1}{2} \log_e \frac{1+r_2}{1-r_2}$$

$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

(2) **Usage**

Double precision:

CALL D3TSCD (N1, R1, N2, R2, CL, IR, Z, ISW, IERR)

Single precision:

CALL R3TSCD (N1, R1, N2, R2, CL, IR, Z, ISW, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\begin{Bmatrix} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{Bmatrix}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/Output | Contents |
|---|---|---|---|---|---|
| 1 | N1 | I | 1 | Input | Number of first sample data $n_1$ |
| 2 | R1 | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | 1 | Input | The correlation coefficient $r_1$ of first sample data (See Note (a)) |
| 3 | N2 | I | 1 | Input | Number of second sample data $n_2$ |
| 4 | R2 | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | 1 | Input | The correlation coefficient $r_2$ of second sample data (See Note (a)) |
| 5 | CL | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | 1 | Input | Confidence level $100(1-\alpha)$ |
| 6 | IR | I | 1 | Output | Test result<br>IR=0 : Hypothesis $\rho_1 = \rho_2$ is accepted<br>IR=1 : Hypothesis $\rho_1 = \rho_2$ is rejected |
| 7 | Z | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | 2 | Output | When ISW=1<br>Z(1) : Value of $t$<br>Z(2) : Value of a standard normal distribution $z_{\frac{\alpha}{2}}$<br>When ISW=2<br>Z(1) : Value of $t$<br>Z(2) : Value of a standard normal distribution $-z_\alpha$<br>When ISW=3<br>Z(1) : Value of $t$<br>Z(2) : Value of a standard normal distribution $z_\alpha$ |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 8 | ISW | I | 1 | Input | The alternative hypothesis switch<br>ISW=1 :<br>When the alternative hypothesis is $\rho_1 \neq \rho_2$<br>ISW=2 :<br>When the alternative hypothesis is $\rho_1 < \rho_2$<br>ISW=3 :<br>When the alternative hypothesis is $\rho_1 > \rho_2$ |
| 9 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $\in$ {1, 2, 3}

(b) N1 $\geq$ 4, N2 $\geq$ 4

(c) $-1.0 <$ R1 $< 1.0$, $-1.0 <$ R2 $< 1.0$

(d) $0.0 \leq$ CL $\leq 100.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | ISW=1 :<br>CL=100.0<br><br>ISW=2 or 3 :<br>CL=0.0 or CL=100.0. | ISW=1 :<br>The positive maximum value is set for Z(2).<br>ISW=2 or 3 :<br>The positive maximum value or the negative minimum value is set for Z(2). |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |

(6) **Notes**

(a) The sample correlation coefficient $r$ for $n$ sample data, $\{x_i, y_i\}$ $(i = 1, \cdots, n)$, is defined as follows. (See 4.4.1 $\left\{\begin{matrix}\text{D2CCMT}\\\text{R2CCMT}\end{matrix}\right\}$)

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_y)^2}}$$

(7) **Example**

See the example in Section 6.2.6 (7).

### 6.3.8 D3TSSR, R3TSSR
### Test in the Simple Linear Regression

(1) **Function**

For regression coefficient $a$ and constant term $b$ in the following simple linear regression expression (or regression line) related to one set of sample data $\{x_i,\, y_i\}$ $(1, \cdots, n)$ of size $n$

$$\hat{y}_i = ax_i + b$$

test the hypothesis related to the regression coefficient $A$ and constant term $B$ in the population to which the respective set of sample data belong with the confidence level $1-\alpha$. Assume that $y_i$ corresponding to each $x_i$ is the random sample from the normal population having the mean $Ax_i - B$ and the variance $\sigma^2$. Obtain the regression coefficient $a$ and constant term $b$ of the sample data from the following normal equations.

$$\begin{cases} \displaystyle\sum_{i=1}^{n} y_i = a\sum_{i=1}^{n} x_i + bn \\ \displaystyle\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i \end{cases}$$

The test criteria are as follows.

(a) Regression coefficient
Hypothesis: $A = A_0$

i. When the population variance is known

A. When the alternative hypothesis is $A \neq A_0$
For $t$ defined as follows

$$t = \frac{a - A_0}{s_a}$$

$$\begin{cases} \text{If } |t| \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $s_a$ is as follows.

$$s_a = \sqrt{\frac{\sigma^2}{\sum(x_i - \mu_x)^2}}$$

$\sigma^2$ : Population variance

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

B. When the alternative hypothesis is $A < A_0$
For $t$ defined as follows

$$t = \frac{a - A_0}{s_a}$$

$$\begin{cases} \text{If } t \geq -z_{\alpha}, \text{ reject} \\ \text{If } t < -z_{\alpha}, \text{ accept} \end{cases}$$

where $s_a$ is as follows.

$$s_a = \sqrt{\frac{\sigma^2}{\sum(x_i - \mu_x)^2}}$$

$\sigma^2$ : Population variance

$$\alpha = 1 - P(z_{\alpha})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

C. When the alternative hypothesis is $A > A_0$

For $t$ defined as follows

$$t = \frac{a - A_0}{s_a}$$

$$\begin{cases} \text{If } t \geq z_\alpha, \text{ reject} \\ \text{If } t < z_\alpha, \text{ accept} \end{cases}$$

where $s_a$ is as follows.

$$s_a = \sqrt{\frac{\sigma^2}{\sum(x_i - \mu_x)^2}}$$

$\sigma^2$ : Population variance

$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

ii. When the population variance is unknown

A. When the alternative hypothesis is $A < A_0$

For $t$ defined as follows

$$t = \frac{a - A_0}{s_a}$$

$$\begin{cases} \text{If } |t| \geq t_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < t_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $s_a$ is as follows.

$$s_a = \sqrt{\frac{\sigma^2}{\sum(x_i - \mu_x)^2}}$$

$\sigma^2$ : Unbiased variance of error variation

$$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}}|n-2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

B. When the alternative hypothesis is $A < A_0$

For $t$ defined as follows

$$t = \frac{a - A_0}{s_a}$$

$$\begin{cases} \text{If } t \geq -t_\alpha, \text{ reject} \\ \text{If } t < -t_\alpha, \text{ accept} \end{cases}$$

where $s_a$ is as follows.

$$s_a = \sqrt{\frac{\sigma^2}{\sum(x_i - \mu_x)^2}}$$

$\sigma^2$ : Unbiased variance of error variation

$$\alpha = 1 - P(t_\alpha|n-2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

C. When the alternative hypothesis is $A > A_0$

For $t$ defined as follows

$$t = \frac{a - A_0}{s_a}$$

$$\begin{cases} \text{If } t \geq t_\alpha, \text{ reject} \\ \text{If } t < t_\alpha, \text{ accept} \end{cases}$$

where $s_a$ is as follows.

$$s_a = \sqrt{\frac{\sigma^2}{\sum(x_i - \mu_x)^2}}$$

$\sigma^2$ : Unbiased variance of error variation

$$\alpha = 1 - P(t_\alpha | n-2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

(b) Constant term
Hypothesis: $B = B_0$

i. When the population variance is known

A. When the alternative hypothesis is $B \neq B_0$

For $t$ defined as follows

$$t = \frac{b - B_0}{s_b}$$

$$\begin{cases} \text{If } |t| \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $s_b$ is as follows.

$$s_b = \sqrt{\sigma^2 \left[ \frac{1}{n} + \frac{\mu_x^2}{\sum(x_i - \mu_x)^2} \right]}$$

$\sigma^2$ : Population variance

$$\frac{\alpha}{2} = 1 - P(z_{\frac{\alpha}{2}})$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

B. When the alternative hypothesis is $B < B_0$

For $t$ defined as follows

$$t = \frac{b - B_0}{s_b}$$

$$\begin{cases} \text{If } t \geq -z_\alpha, \text{ reject} \\ \text{If } t < -z_\alpha, \text{ accept} \end{cases}$$

where $s_b$ is as follows.

$$s_b = \sqrt{\sigma^2 \left[ \frac{1}{n} + \frac{\mu_x^2}{\sum(x_i - \mu_x)^2} \right]}$$

$\sigma^2$ : Population variance

$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

C. When the alternative hypothesis is $B > B_0$

For $t$ defined as follows

$$t = \frac{b - B_0}{s_b}$$

$$\begin{cases} \text{If } t \geq z_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } t < z_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $s_b$ is as follows.

$$s_b = \sqrt{\sigma^2 \left[ \frac{1}{n} + \frac{\mu_x^2}{\sum(x_i - \mu_x)^2} \right]}$$

$\sigma^2$ : Population variance

$$\alpha = 1 - P(z_\alpha)$$

Here, $P(z)$ is the c.d.f. value of the standard normal distribution.

ii. When the population variance is unknown

A. When the alternative hypothesis is $B \neq B_0$

For $t$ defined as follows

$$t = \frac{b - B_0}{s_b}$$

$$\begin{cases} \text{If } |t| \geq t_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } |t| < t_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $s_b$ is as follows.

$$s_b = \sqrt{\sigma^2 \left[\frac{1}{n} + \frac{\mu_x^2}{\sum(x_i - \mu_x)^2}\right]}$$

$\sigma^2$ : Unbiased variance of error variation

$$\frac{\alpha}{2} = 1 - P(t_{\frac{\alpha}{2}}|n-2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

B. When the alternative hypothesis is $B < B_0$

For $t$ defined as follows

$$t = \frac{b - B_0}{s_b}$$

$$\begin{cases} \text{If } t \geq -t_{\alpha}, \text{ reject} \\ \text{If } t < -t_{\alpha}, \text{ accept} \end{cases}$$

where $s_b$ is as follows.

$$s_b = \sqrt{\sigma^2 \left[\frac{1}{n} + \frac{\mu_x^2}{\sum(x_i - \mu_x)^2}\right]}$$

$\sigma^2$ : Unbiased variance of error variation

$$\alpha = 1 - P(t_\alpha|n-2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

C. When the alternative hypothesis is $B > B_0$

For $t$ defined as follows

$$t = \frac{b - B_0}{s_b}$$

$$\begin{cases} \text{If } t \geq t_{\frac{\alpha}{2}}, \text{ reject} \\ \text{If } t < t_{\frac{\alpha}{2}}, \text{ accept} \end{cases}$$

where $s_b$ is as follows.

$$s_b = \sqrt{\sigma^2 \left[\frac{1}{n} + \frac{\mu_x^2}{\sum(x_i - \mu_x)^2}\right]}$$

$\sigma^2$ : Unbiased variance of error variation

$$\alpha = 1 - P(t_\alpha|n-2)$$

Here, $P(t|n)$ is the cumulative distribution function (c.d.f.) of a $t$ distribution having number of degrees of freedom $n$.

(2) **Usage**

Double precision:

CALL D3TSSR (X, N, Y, YV, X0, CL, IR, Z, STAT, ISW1, ISW2, ISW3, W, IERR)

Single precision:

CALL R3TSSR (X, N, Y, YV, X0, CL, IR, Z, STAT, ISW1, ISW2, ISW3, W, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: {INTEGER(4) as for 32bit Integer, INTEGER(8) as for 64bit Integer}
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | X | {D, R} | N | Input | Values of independent variable X of sample, $x_i (i = 1, n)$ |
| 2 | N | I | 1 | Input | Number of sample data $n$ |
| 3 | Y | {D, R} | N | Input | Values of dependent variable Y of sample, $y_i (i = 1, n)$ |
| 4 | YV | {D, R} | 1 | Input | Variance of the population to which dependent variable Y belongs(if ISW2 = 1) (See 10.2.1) |
| | | | | Output | Unbiased variance of error variation $\sigma^2$ (When ISW2=2) |
| 5 | X0 | {D, R} | 1 | Input | When ISW1=1<br>Tested regression coefficient value $A_0$<br>When ISW1=2<br>Tested constant term value $B_0$ |
| 6 | CL | {D, R} | 1 | Input | Confidence level $100(1-\alpha)(\%)$ |
| 7 | IR | {D, R} | 1 | Output | Test result<br>When ISW1=1,<br>IR=0 : Hypothesis $A = A_0$ is accepted<br>IR=1 : Hypothesis $A = A_0$ is rejected<br>When ISW1=2,<br>IR=0 : Hypothesis $B = B_0$ is accepted<br>IR=1 : Hypothesis $B = B_0$ is rejected |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 8 | Z | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | 2 | Output | When ISW2=1 and ISW3=1<br>Z(1) : Value of $t$<br>Z(2) : Values of a standard normal distribution $z_{\frac{\alpha}{2}}$<br>When ISW2=1 and ISW3=2<br>Z(1) : Value of $t$<br>Z(2) : Values of a standard normal distribution $-z_\alpha$<br>When ISW2=1 and ISW3=3<br>Z(1) : Value of $t$<br>Z(2) : Values of a standard normal distribution $z_\alpha$<br>When ISW2=2 and ISW3=1<br>Z(1) : Value of $t$<br>Z(2) : Value of $t$ distribution $t_{\frac{\alpha}{2}}$<br>When ISW2=2 and ISW3=2<br>Z(1) : Value of $t$<br>Z(2) : Value of $t$ distribution $-t_\alpha$<br>When ISW2=2 and ISW3=3<br>Z(1) : Value of $t$<br>Z(2) : Value of $t$ distribution $t_\alpha$ |
| 9 | STAT | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | 2 | Output | STAT(1) : The sample regression coefficient<br>STAT(2) : The sample constant term |
| 10 | ISW1 | I | 1 | Input | The switch for selection of the statistic.<br>ISW1=1 : Testing the population regression coefficient<br>ISW1=2 : Testing the population constant term |
| 11 | ISW2 | I | 1 | Input | Switch for variance<br>ISW2=1 : The variance of the population is entered for YV<br>ISW2=2 : The variance (not an unbiased estimate) of the sample data is entered for YV |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 12 | ISW3 | I | 1 | Input | The alternative hypothesis switch<br>For ISW1=1,<br>ISW3=1 :<br>When the alternative hypothesis is $A \neq A_0$<br>ISW3=2 :<br>When the alternative hypothesis is $A < A_0$<br>ISW3=3 :<br>When the alternative hypothesis is $A > A_0$<br>For ISW1=2,<br>ISW3=1 :<br>When the alternative hypothesis is $B \neq B_0$<br>ISW3=2 :<br>When the alternative hypothesis is $B < B_0$<br>ISW3=3 :<br>When the alternative hypothesis is $B > B_0$ |
| 13 | W | $\begin{Bmatrix} \mathrm{D} \\ \mathrm{R} \end{Bmatrix}$ | 29 | Work | Work area |
| 14 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW1 $\in \{1, 2\}$

(b) ISW2 $\in \{1, 2\}$

(c) ISW3 $\in \{1, 2, 3\}$

(d) N $\geq 3$

(e) $0.0 \leq$ CL $\leq 100.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | ISW3=1 :<br>CL=100.0<br><br>ISW3=2 or 3 :<br>CL=0.0 or CL=100.0. | ISW3=1 :<br>The positive maximum value is set for Z(2).<br>ISW3=2 or 3 :<br>The positive maximum value or the negative minimum value is set for Z(2). |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 3040 | Restriction (e) was not satisfied. | |
| 4000 | There are no difference among independent variable X. | |
| 4100 | The Unbiased variance of error variation is 0.0. (See 10.2.1) | |

(6) **Notes**
None

(7) **Example**

(a) Problem
From the following one set of sample data of size 9, test the hypothesis $A = 0$ related to the population regression coefficient with 95% confidence level.

| $x_i$ | $y_i$ |
|---|---|
| 1 | 3 |
| 2 | 3 |
| 3 | 5 |
| 4 | 5 |
| 5 | 6 |
| 6 | 7 |
| 7 | 8 |
| 8 | 8 |
| 9 | 9 |

Assume that the alternative hypothesis is $A \neq 0$ and the population variance value is unknown.

(b) Input data
ISW1=1 ISW2=2, ISW3=1, N=9, array X, array Y, X0=0.0 and CL=95.0.

(c) Main program

```
      PROGRAM B3TSSR
! *** EXAMPLE OF D3TSSR ***
      IMPLICIT REAL(8) (A-H,O-Z)
      INTEGER N
      PARAMETER (N=9)
      REAL(8) X(N),Y(N),W(29)
      INTEGER IERR,ISW1,ISW2
      REAL(8) CL,Z(2),STAT(2)
```

```
      DATA (X(I),I=1,N) &
            /1.0D0, 2.0D0, 3.0D0, 4.0D0, 5.0D0,&
             6.0D0, 7.0D0, 8.0D0, 9.0D0/,&
           (Y(I),I=1,N) &
            /3.0D0, 3.0D0, 5.0D0, 5.0D0, 6.0D0,&
             7.0D0, 8.0D0, 8.0D0, 9.0D0/
!
      ISW1=1
      ISW2=2
      ISW3=1
      X0=0.0D0
      CL=95.0D0
!
      WRITE(6,1000)
      WRITE(6,1100)ISW1
      WRITE(6,1110)ISW2
      WRITE(6,1120)ISW3
      WRITE(6,2000)N
      WRITE(6,2100)CL
      WRITE(6,2200)
      DO 100 I=1,N
         WRITE(6,2210)I,X(I),Y(I)
  100 CONTINUE
!
      CALL D3TSSR(X,N,Y,YV,X0,CL,IR,Z,STAT,ISW1,ISW2,ISW3,W,IERR)
!
      WRITE(6,3000)
      WRITE(6,4000) IERR
      WRITE(6,3010)
      WRITE(6,5000)X0
      IF(IR.EQ.0) THEN
         WRITE(6,5010)
      ELSE
         WRITE(6,5020)
      ENDIF
      WRITE(6,6000) Z(1),Z(2)
      WRITE(6,7000) STAT(1)
      WRITE(6,7100) STAT(2)
!
      STOP
!
 1000 FORMAT(' ',/,5X,'*** D3TSSR ***',/,&
      6X,'** INPUT **')
 1100 FORMAT(9X,'ISW1= ',I3)
 1110 FORMAT(9X,'ISW2= ',I3)
 1120 FORMAT(9X,'ISW3= ',I3)
 2000 FORMAT(9X,'N   = ',I3)
 2100 FORMAT(9X,'CL  = ',F5.1)
 2200 FORMAT(9X,'SAMPLE DATA',/,&
             9X,' I',1X,'X(I)',1X,'Y(I)')
 2210 FORMAT(9X,I2,2F5.1)
 3000 FORMAT(6X,'** OUTPUT **')
 3010 FORMAT(6X,'*** REGRESSION COEFFICIENT ***')
 4000 FORMAT(9X,'IERR = ',I4)
 5000 FORMAT(9X,'HYPOTHESIS:  A .EQ.',F5.1)
 5010 FORMAT(9X,'HYPOTHESIS IS ACCEPTED.')
 5020 FORMAT(9X,'HYPOTHESIS IS REJECTED.')
 6000 FORMAT(9X,'Z(1)= ',F8.4,/,9X,'Z(2)= ',F8.4)
 7000 FORMAT(9X,'REGRESSION COEFFICIENT OF SAMPLE= ',D18.10)
 7100 FORMAT(9X,'CONSTANT TERM          OF SAMPLE= ',D18.10)
      END
```

(d) Output results

```
     *** D3TSSR ***
      ** INPUT **
         ISW1=   1
         ISW2=   2
         ISW3=   1
         N   =   9
         CL  =  95.0
         SAMPLE DATA
          I X(I) Y(I)
          1  1.0  3.0
          2  2.0  3.0
          3  3.0  5.0
          4  4.0  5.0
          5  5.0  6.0
          6  6.0  7.0
          7  7.0  8.0
          8  8.0  8.0
          9  9.0  9.0
      ** OUTPUT **
         IERR =    0
      *** REGRESSION COEFFICIENT ***
         HYPOTHESIS:  A .EQ.  0.0
         HYPOTHESIS IS REJECTED.
         Z(1)=  14.7577
         Z(2)=   2.3646
         REGRESSION COEFFICIENT OF SAMPLE=   0.7833333333D+00
         CONSTANT TERM          OF SAMPLE=   0.2083333333D+01
```

# Chapter 7

# ANALYSIS OF VARIANCE AND DESIGN OF EXPERIMENTS

## 7.1 INTRODUCTION

**Design of experiments** is one of the fields of inductive statistics. With design of experiments, an actual measurement value $x$ is considered to be the realized value of the random variable $X$, and the internal structure of $X$ corresponding to the experiment conditions is considered and analyzed through the actual measurement values. To perform the analysis, the experiment must be designed in advance. However, to analyze the actual measurement values, the within class variation or between class variation and error variation quantities are systematically sought by using tables called analysis of variance tables. This type of analysis method is called **analysis of variance**.
This library provides the following functions for performing design of experiments and analysis of variance.

- One-way Layout
- Two-way Layout
- Multiple-way Layout
- Randomized Block Design
- Greco-Latin Square Method
- Cumulative Method
- Balanced Incomplete Block Design (BIBD)

### 7.1.1 Explanation

(1) **One-way Layout**

With the one-way layout analysis of variance method, given one-way layout data $\{x_{ij}\}(i = 1, \cdots, n_j; j = 1, \cdots, m)$ consisting of $m$ levels so that the number of repetitions in each level is $n_j$, the problem is to test whether or not there is a difference among the means of $m$ population distributions based on this data. At this time, the structure equation of the one-way layout observed values is defined as follows.

$$x_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

$\mu$ represents the total mean over all observed values, $\alpha_i$ represents the effect in level $i$, and $\epsilon_{ij}$ represents the observation error, and the observed values are assumed to be mutually independent values that obey a normal distribution $N(0, \sigma^2)$. The following kinds of data are calculated in the analysis. Mean of each level:

$$\bar{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} \qquad j = 1, \cdots, m$$

Variance of each level :

$$V_j = \frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{\alpha_j} \qquad j = 1, \cdots, m$$

Total mean :

$$\bar{x} = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n_j} x_{ij}}{\sum_{j=1}^{m} n_j}$$

Here, $\alpha_j$ is $n_j$ when a sample variance is used, and $\alpha_j$ is $n_j - 1$ when an unbiased variance is used.
Variation :

- Total variation
$$S_T = \sum_{j=1}^{m}\sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2$$
- Between class variation
$$S_A = \sum_{j=1}^{m} (\bar{x}_j - \bar{x})^2$$
- Error variation
$$S_E = S_T - S_A$$

Degrees of freedom :

- Degrees of freedom of total variation
$$\phi_T = \sum_{j=1}^{m} n_j - 1$$

- Degrees of freedom of between class variation

$$\phi_A = m - 1$$

- Degrees of freedom of error variation

$$\phi_E = \sum_{j=1}^{m}(n_j - 1)$$

Unbiased variance :

- Unbiased variance of between class variation

$$V_A = \frac{S_A}{\phi_A}$$

- Unbiased variance of error variation

$$V_E = \frac{S_E}{\phi_E}$$

Variance ratio :

$$F_A = \frac{V_A}{V_E}$$

Contribution ratio :

- Contribution ratio of between class variance

$$P_A = \frac{S_A - \phi_A \cdot V_E}{S_T}$$

- Contribution ratio of error variance

$$P_E = 1 - P_A$$

The test is performed by using the critical value of an $F$ distribution for which the variance ratio $F_A$ has degrees of freedom $\phi_A$ and $\phi_E$.

(2) **Two-way Layout**
Given factors A and B consisting of $m_a$ and $m_b$ levels respectively and two-way layout data $\{x_{kij}\}(k = 1, \cdots, n_{ij}; i = 1, \cdots, m_a; j = 1, \cdots, m_b)$ for which the number of repetitions in the combination of each level is $n_{ij}$, the problem is to test the effect $\alpha_i$ in level $i$ factor A, the effect $\beta_j$ in level $j$ of factor B, and the interaction effect $\gamma_{ij}$ when the structure equation of the two-way layout observed values is defined as follows based on this data.

$$x_{kij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{kij}$$

$\mu$ represents the total mean over all observed values, $\epsilon_{ij}$ represents the observation error, and the observed values are assumed to be mutually independent values that obey a normal distribution $N(0, \sigma^2)$. Since each effect is a constant, this kind of model is called a **parameter model**. For details, refer to the reference bibliography.

(3) **Multiple-way Layout**
This method is a generalization of the two-way layout method in which the number of factors is increased. The problem is to test the effects of each factor and the interaction effects between the factors. For details, refer to the reference bibliography.

(4) **Randomized Block Design**
Randomized block design is one type of experimental design. Generally, when performing an experiment, there may be factors that tend to increase the efficiency of the experiment if they are taken into account in advance even though they are secondary factors from the viewpoint of conditions (**factors**) having to do with the immediate objective. To eliminate the effect of these kinds of factors, **randomized block design** uses blocks that group the factors so that the effect of these secondary factors is reduced. Generally, the blocks are arranged so that the secondary conditions increase between different blocks and decrease within the same block. With randomized block design, equal subdivisions are provided in the levels of factors in each block, and for each block, the various levels of factors are randomly assigned to each subdivision.

(5) **Greco-Latin Square Method**
The $n$ Latin characters $A, B, \cdots$ arranged in $n$ rows and $n$ columns so that the same Latin character is not duplicated in the same row or same column is called an $n \times n$ Latin Square. By making experimental arrangements using this kind of Latin square to eliminate nonuniformity in the row and column directions, you can test differences in significance among treatments $A, B, \cdots$ (Latin Square Method). However, in this case, interactions between row effects and column effects must not exist, and the general mean, row effect, column effect, treatment effect, and experimental error must be additively combined for the experimental data. An analysis of variance performed using a table obtained by combining two orthogonal Latin Squares (Greco-Latin square) in place of a Latin square is called the Greco-Latin Square Method. The Greco-Latin Square Method can be used when the number of factors is 4, the number of levels for each factor is the same and at least 4, and no interactions of the factors exist.

(6) **Cumulative Method**
For details, refer to the reference bibliography.

(7) **Balanced Incomplete Block design (BIBD)** In a block experiment, when each block contains a complete set of treatments, it is called a complete block design. If this is not the case and one set of treatments to be compared is incomplete and is not entered in the blocks, it is called an incomplete block design. In particular, a block design in which the number of repetitions of each trial is equal and the number of times two arbitrary trials appear in the same block is equal is called a balanced incomplete block design.

### 7.1.2 Reference Bibliography

(1) Fisher, R. A. , "The design of experiments", 7th ed. , Oliver and Boyd, Edinburgh (1966)

## 7.2 ONE-WAY LAYOUT

### 7.2.1 D41WR1, R41WR1
### One-Way Layout Analysis of Variance

(1) **Function**

Given one-way layout data $\{x_{ij}\}(i = 1, \cdots, n_j; j = 1, \cdots, m)$ consisting of $m$ levels so that the number of repetitions in each level is $n_j$, the D41WR1 or R41WR1 obtains the mean and variance of each level and the total mean over all levels and performs an analysis of variance.

The mean and variance of each level and the total mean over all levels for the one-way layout data $\{x_{ij}\}$ $(i = 1, \cdots, n_j; j = 1, \cdots, m)$ are defined by the following equations.

Mean of each level :

$$\bar{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} \qquad j = 1, \cdots, m$$

Variance of each level :

$$V_j = \frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{\alpha_j} \qquad j = 1, \cdots, m$$

Total mean :

$$\bar{x} = \frac{\sum_{j=1}^{m} \sum_{i=1}^{n_j} x_{ij}}{\sum_{j=1}^{m} n_j}$$

Here, $\alpha_j$ is $n_j$ when a sample variance is used, and $\alpha_j$ is $n_j - 1$ when an unbiased variance is used.

Also, the analysis of variance results are defined by the following equations.

Variation :

- Total variation

$$S_T = \sum_{j=1}^{m} \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2$$

- Between class variation

$$S_A = \sum_{j=1}^{m} (\bar{x}_j - \bar{x})^2$$

- Error variation

$$S_E = S_T - S_A$$

Degrees of freedom :

- Degrees of freedom of total variation

$$\phi_T = \sum_{j=1}^{m} n_j - 1$$

- Degrees of freedom of between class variation

$$\phi_A = m - 1$$

- Degrees of freedom of error variation

$$\phi_E = \sum_{j=1}^{m} (n_j - 1)$$

Unbiased variance :

- Unbiased variance of between class variation

$$V_A = \frac{S_A}{\phi_A}$$

- Unbiased variance of error variation

$$V_E = \frac{S_E}{\phi_E}$$

Variance ratio :

$$F_A = \frac{V_A}{V_E}$$

Contribution ratio :

- Contribution ratio of between class variation

$$P_A = \frac{S_A - \phi_A \cdot V_E}{S_T}$$

- Contribution ratio of error variation

$$P_E = 1 - P_A$$

(2) **Usage**

Double precision:

CALL D41WR1 (A, NA, M, N, NR, STAT, X1, V, ISW, IERR)

Single precision:

CALL R41WR1 (A, NA, M, N, NR, STAT, X1, V, ISW, IERR)

## (3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | NA,M | Input | Matrix in which observed values are stored ($x_{ij}$) (See Note (a)) |
| 2 | NA | I | 1 | Input | Adjustable dimension of array A |
| 3 | M | I | 1 | Input | Number of levels $m$ |
| 4 | N | I | M | Input | Number of repetitions in $j$th level $n_j$ (not used when NR$\geq$ 1) |
| 5 | NR | I | 1 | Input | Number of repetitions when the numbers of repetitions in each level are equal $n_1 = n_2 = \cdots = n_g$. When the numbers of repetitions in each level are not equal, set a value that is less than or equal to zero. |
| 6 | STAT | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | M,2 | Output | Mean and variance of each level (See Note (b)) |
| 7 | X1 | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Output | Total mean |
| 8 | V | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 11 | Output | Analysis of variance results<br>(For the storage method of the analysis of variance results, see Table 7−1 in Note (c).) |
| 9 | ISW | I | 1 | Input | Processing switch<br>0: Calculate the unbiased variance<br>1: Calculate the sample variance |
| 10 | IERR | I | 1 | Output | Error indicator |

## (4) **Restrictions**

(a) ISW = 0, 1

(b) NR$\leq$0 and NA $\geq$ N(i) $\quad$ (i $= 1, \cdots,$ M)
or NA$\geq$NR

(c) M $\geq$ 1

(d) NR$\leq$0 and N(i)$\geq$ 1 $\quad$ (i $= 1, \cdots,$ M)
or NR$\geq$1

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (a) was not satisfied. | Processing continues with ISW=0. |
| 1010 | The following five conditions were satisfied at the same time.<br>(a) ISW= 0<br>(b) M> 1<br>(c) NR< 1<br>(d) N(i) = 1 for some i (i = 1, $\cdots$, M).<br>(e) N(1)=N(2)= $\cdots$ =N(M)= 1 does not hold. | The absolute value maximum that can be represented is set for each STAT(i, 2) for which N(i)= 1, and the other STAT(i, 2) are calculated normally. |
| 1020 | One of the following conditions was satisfied.<br>(a) M=1 and NR< 1 and N(1)> 1<br>(b) M= 1 and ISW= 1<br>(c) M> 1 and NR= 1 and ISW= 1<br>(d) M> 1 and NR< 1 and ISW= 1 and N(1)=N(2)= $\cdots$ =N(M)= 1. | The absolute value maximum that can be represented is set for V(7) to V(11). |
| 1030 | One of the following conditions was satisfied.<br>(a) NR= 1 and ISW= 0<br>(b) NR< 1 and ISW= 0 and N(1)=N(2)= $\cdots$ =N(M)= 1 | The absolute value maximum that can be represented is set for all STAT(i, 2) (i= 1, $\cdots$,M) and for V(7) to V(11). |
| 3000 | Any of restriction (b), (c) or (d) were not satisfied. | Processing is aborted. |

(6) **Notes**

(a) The observed values ($x_{ij}$) are stored in array A as a real matrix (two-dimensional array type) data (See Appendix A).

(b) The mean and variance of each level are stored as follows in the array STAT

STAT($j$, 1) : Mean $\bar{x}_j$
STAT($j$, 2) : Variance $v_j$ , $j = 1, \cdots, \mathrm{M}$

(c) The analysis of variance results are stored as follows in the array V.

Table 7−1 Analysis of Variance Table

| Element | Variation | Degrees of freedom | Unbiased variance | Variance ratio | Contribution ratio |
|---|---|---|---|---|---|
| Total | V(1) | V(4) | | | |
| Between level | V(2) | V(5) | V(7) | V(9) | V(10) |
| Error | V(3) | V(6) | V(8) | | V(11) |

(d) Statistics obtained when calculations are performed using an unbiased estimate can be applied to a population for which sampling with replacement is performed from an infinite or finite population. Statistics obtained when calculations are performed using a sample variance can be applied to a population for which the population and sample match.

(7) **Example**

(a) Problem

For one-way layout data given by the matrix $X$ shown below, obtain the mean and variance of each level and the total mean over all levels and perform an analysis of variance.

$$X = \begin{bmatrix} 71 & 83 & 85 & 84 & 82 \\ 74 & 79 & 84 & 80 & 78 \\ 76 & 83 & 89 & 82 & 83 \\ 72 & 77 & 82 & 85 & 83 \end{bmatrix}$$

(b) Input data

One-way layout data $X$, NA=100, M=5, NR=4 and ISW=0.

(c) Main program

```
      PROGRAM B41WR1
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 100, M = 5 )
      DIMENSION A(NA,M),N(M),STAT(M,2),V(11)
!
      WRITE(6,6000)
      IERR =   0
      DO 100 I=1,M
         N(I) = 0
  100 CONTINUE
      READ(5,*) NR
      READ(5,*) ISW
      DO 110 I=1,NR
         READ(5,*) (A(I,J),J=1,M)
  110 CONTINUE
      WRITE(6,6010) ISW,M,NR,(N(I),I=1,M)
      DO 120 I=1,NR
         WRITE(6,6020) (A(I,J),J=1,M)
  120 CONTINUE
      CALL D41WR1(A,NA,M,N,NR,STAT,X1,V,ISW,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040)
      DO 130 L=1,M
         WRITE(6,6050) L,STAT(L,1),STAT(L,2)
  130 CONTINUE
      WRITE(6,6060) X1
      WRITE(6,6070) 'TOTAL',V(1),V(4)
      WRITE(6,6070) 'LEVEL',V(2),V(5),V(7),V(9),V(10)
      WRITE(6,6080) 'ERROR',V(3),V(6),V(8),V(11)
!
      STOP
 6000 FORMAT(  ' *** D41WR1 ***',/,&
             /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'ISW = ',I6,/,&
             /,7X,'M = ',I6,5X,'NR = ',I6,/,&
             /,7X,'NUMBER OF REPETITIONS IN EACH LEVEL',/,/,&
              7X,5(2X,I6),/,&
             /,7X,'OBSERVATION MATRIX',/)
 6020 FORMAT(  7X,5(2X,F11.2))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
             /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'VALUE OF EACH LEVEL',/,&
             /,10X,'LEVEL',6X,'MEAN',6X,'VARIANCE',/,&
              9X,35('-'))
 6050 FORMAT(  7X,I6,2X,F11.2,2X,D15.8)
 6060 FORMAT( /,7X,'MEAN OVER ALL LEVELS = ',D15.8,/,&
             /,7X,'ANALYSIS OF VARIANCE TABLE',/,&
             /,10X,'FACTOR',&
                  6X,'S.S.',9X,'D.F.',7X,'M.S.',8X,'V.R.',8X,'C.R.',/,&
              9X,69('-'))
 6070 FORMAT(  10X,A,2(2X,F11.2),3(1X,D11.4))
 6080 FORMAT(  10X,A,2(2X,F11.2),1X,D11.4,13X,D11.4)
      END
```

(d) Output results

```
*** D41WR1 ***

  ** INPUT **

      ISW =      0

      M =      5     NR =      4

      NUMBER OF REPETITIONS IN EACH LEVEL

             0       0       0       0       0
```

```
    OBSERVATION MATRIX

            71.00         83.00         85.00         84.00         82.00
            74.00         79.00         84.00         80.00         78.00
            76.00         83.00         89.00         82.00         83.00
            72.00         77.00         82.00         85.00         83.00

** OUTPUT **

    IERR =      0

    VALUE OF EACH LEVEL

       LEVEL      MEAN      VARIANCE
      ------------------------------------
         1        73.25   0.49166667D+01
         2        80.50   0.90000000D+01
         3        85.00   0.86666667D+01
         4        82.75   0.49166667D+01
         5        81.50   0.56666667D+01

    MEAN OVER ALL LEVELS =  0.80600000D+02

    ANALYSIS OF VARIANCE TABLE

       FACTOR      S.S.         D.F.       M.S.        V.R.        C.R.
      ---------------------------------------------------------------------
       TOTAL      414.80       19.00
       LEVEL      315.30        4.00  0.7882D+02  0.1188D+02  0.6962D+00
       ERROR       99.50       15.00  0.6633D+01              0.3038D+00
```

## 7.3 TWO-WAY LAYOUT

### 7.3.1 D42WRN, R42WRN
### Two-Way Layout Analysis of Variance

(1) **Function**

For factors A and B consisting of $m_a$ and $m_b$ levels respectively and two-way layout data $\{x_{ij}\}(i = 1, \cdots, m_a; j = 1, \cdots, m_b)$ for which there are no repetitions in the combination of each level, the D42WRN or R42WRN obtains the mean and variance of each level for each factor and the total mean over all levels and performs an analysis of variance.

The mean and variance of each level for each factor and the total mean over all levels are defined by the following equations.

Mean of each level of factor A :

$$\bar{x}_{i\cdot} = \frac{1}{m_b}\sum_{j=1}^{m_b} x_{ij}$$

Variance of each level of factor A :

$$V_{ai} = \frac{1}{\alpha_b}\sum_{j=1}^{m_b}(x_{ij} - \bar{x}_{i\cdot})^2$$

Mean of each level of factor B :

$$\bar{x}_{\cdot j} = \frac{1}{m_a}\sum_{i=1}^{m_a} x_{ij}$$

Variance of each level of factor B :

$$V_{bj} = \frac{1}{\alpha_a}\sum_{i=1}^{m_a}(x_{ij} - \bar{x}_{\cdot j})^2$$

Total mean :

$$\bar{x} = \frac{1}{m_a \cdot m_b}\sum_{i=1}^{m_a}\sum_{j=1}^{m_b} x_{ij}$$

Here, $\alpha_a = m_a$ and $\alpha_b = m_b$ when a sample variance is used, and $\alpha_a = m_a - 1$ and $\alpha_b = m_b - 1$ when an unbiased variance is used.

Also, the analysis of variance results are defined by the following equations.

Variation :

- Total variation

$$S_T = \sum_{i=1}^{m_a}\sum_{j=1}^{m_b}(x_{ij} - \bar{x})^2$$

- Variation of factor A

$$S_A = m_b\sum_{i=1}^{m_a}(\bar{x}_{i\cdot} - \bar{x})^2$$

- Variation of factor B

  $$S_B = m_a \sum_{j=1}^{m_b} (\bar{x}_{\cdot j} - \bar{x})^2$$

- Error variation

  $$S_E = S_T - (S_A + S_B)$$

Degrees of freedom :

- Degrees of freedom of total variation

  $$\phi_T = m_a \cdot m_b - 1$$

- Degrees of freedom of factor A variation

  $$\phi_A = m_a - 1$$

- Degrees of freedom of factor B variation

  $$\phi_B = m_b - 1$$

- Degrees of freedom of error variation

  $$\phi_E = (m_a - 1) \cdot (m_b - 1)$$

Unbiased variance :

- Unbiased variance of factor A variation

  $$V_A = \frac{S_A}{\phi_A}$$

- Unbiased variance of factor B variation

  $$V_B = \frac{S_B}{\phi_B}$$

- Unbiased variance of error variation

  $$V_E = \frac{S_E}{\phi_E}$$

Variance ratio :

- Variance ratio for unbiased variance of factor A variation

  $$F_A = \frac{V_A}{V_E}$$

- Variance ratio for unbiased variance of factor B variation

  $$F_B = \frac{V_B}{V_E}$$

Contribution ratio :

- Contribution ratio of factor A variation

  $$P_A = \frac{S_A - \phi_A \cdot V_E}{S_T}$$

- Contribution ratio of factor B variation

  $$P_B = \frac{S_B - \phi_B \cdot V_E}{S_T}$$

- Contribution ratio of error variation

$$P_E = 1 - P_A - P_B$$

(2) **Usage**

Double precision:
CALL D42WRN (A, NA, LA, LB, STATA, STATB, X1, V, ISW, IERR)
Single precision:
CALL R42WRN (A, NA, LA, LB, STATA, STATB, X1, V, ISW, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\begin{Bmatrix}\text{D}\\ \text{R}\end{Bmatrix}$ | NA,LB | Input | Matrix in which observed values are stored ($x_{ij}$) (See Note (a)) |
| 2 | NA | I | 1 | Input | Adjustable dimension of array A |
| 3 | LA | I | 1 | Input | Number of levels of factor A $m_a$ |
| 4 | LB | I | 1 | Input | Number of levels of factor B $m_b$ |
| 5 | STATA | $\begin{Bmatrix}\text{D}\\ \text{R}\end{Bmatrix}$ | LA,2 | Output | Mean and variance of each level of factor A (See Note (b)) |
| 6 | STATB | $\begin{Bmatrix}\text{D}\\ \text{R}\end{Bmatrix}$ | LB,2 | Output | Mean and variance of each level of factor B (See Note (b)) |
| 7 | X1 | $\begin{Bmatrix}\text{D}\\ \text{R}\end{Bmatrix}$ | 1 | Output | Total mean |
| 8 | V | $\begin{Bmatrix}\text{D}\\ \text{R}\end{Bmatrix}$ | 16 | Output | Analysis of variance table (For the storage method of the analysis of variance table, see Table 7−2 in Note (c).) |
| 9 | ISW | I | 1 | Input | Processing switch<br>0: Calculate the unbiased variance<br>1: Calculate the sample variance |
| 10 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW = 0, 1

(b) NA $\geq$ LA $\geq$ 1

(c) NB $\geq$ 1

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (a) was not satisfied. | Processing continues with ISW=0. |
| 1010 | ISW = 0 was specified when LA = 1 or LB = 1. | If LA = 1, the absolute value maximum that can be represented is set for the variance of each level of factor B. If LB = 1, the absolute value maximum that can be represented is set for the variance of each level of factor A. Also, the absolute value maximum that can be represented is set for V(9) to V(16). |
| 1020 | ISW = 1 was specified when LA = 1 or LB = 1. | The absolute value maximum that can be represented is set for V(9) to V(16). |
| 3000 | Any of restriction (b) or (c) were not satisfied. | Processing is aborted. |

(6) **Notes**

(a) The observed values ($x_{ij}$) are stored in array A as a real matrix (two-dimensional array type) data (See Appendix A).

(b) The mean and variance of each level of factors A and B are stored as follows in the arrays STATA and STATB.

STATA($i$, 1) : Mean of each level of factor A $\bar{x}_{i\cdot}$
STATA($i$, 2) : Variance of each level of factor A $V_{ai}$
STATB($j$, 1) : Mean of each level of factor B $\bar{x}_{\cdot j}$
STATB($j$, 2) : Variance of each level of factor B $V_{bj}$

, $i = 1, \cdots, \mathrm{LA}; j = 1, \cdots, \mathrm{LB}$

(c) The analysis of variance table elements are stored as follows in the array V.

Table 7−2 Analysis of Variance Table Storage Status

| Element | Variation | Degrees of freedom | Unbiased variance | Variance ratio | Contribution ratio |
|---|---|---|---|---|---|
| Total | V(1) | V(5) | | | |
| Factor A | V(2) | V(6) | V(9) | V(12) | V(14) |
| Factor B | V(3) | V(7) | V(10) | V(13) | V(15) |
| Error | V(4) | V(8) | V(11) | | V(16) |

(d) Statistics obtained when calculations are performed using an unbiased estimate can be applied to a population for which sampling with replacement is performed from an infinite or finite population. Statistics obtained when calculations are performed using a sample variance can be applied to a population for which the population and sample match.

(7) **Example**

(a) Problem

Given two-way layout data for which there are no repetitions and having factors A and B as shown in matrix $X$ below, obtain the mean and variance of each level of each factor and the total mean over all levels and perform an analysis of variance.

$$X = \begin{bmatrix} 1.26 & 1.21 & 1.19 \\ 1.29 & 1.23 & 1.23 \\ 1.38 & 1.27 & 1.22 \end{bmatrix}$$

(b) Input data

Two-way layout data $X$, NA=10, LA=3, LB=3 and ISW=0.

(c) Main program

```
      PROGRAM B42WRN
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 10, LA = 3, LB = 3 )
      DIMENSION A(NA,LB),STATA(LA,2),STATB(LB,2),V(16)
!
      WRITE(6,6000)
      IERR = 0
      READ(5,*) ISW
      DO 100 I = 1,LA
         READ(5,*) (A(I,J),J=1,LB)
  100 CONTINUE
      WRITE(6,6010) ISW,LA,LB
      DO 110 I=1,LA
         WRITE(6,6020) (A(I,J),J=1,LB)
  110 CONTINUE
      CALL D42WRN(A,NA,LA,LB,STATA,STATB,X1,V,ISW,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) 'A'
      DO 120 L=1,LA
         WRITE(6,6050) L,STATA(L,1),STATA(L,2)
  120 CONTINUE
      WRITE(6,6040) 'B'
      DO 130 L=1,LB
         WRITE(6,6050) L,STATB(L,1),STATB(L,2)
  130 CONTINUE
      WRITE(6,6060) X1
      WRITE(6,6070) 'TOTAL',V(1),V(5)
      WRITE(6,6070) '  A  ',V(2),V(6),V(9),V(12),V(14)
      WRITE(6,6070) '  B  ',V(3),V(7),V(10),V(13),V(15)
      WRITE(6,6080) 'ERROR',V(4),V(8),V(11),V(16)
!
      STOP
 6000 FORMAT(  ' *** D42WRN ***',/,&
             /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'ISW = ',I6,/,&
             /,7X,'LA = ',I6,5X,'LB = ',I6,/,&
             /,7X,'OBSERVATION MATRIX',/)
 6020 FORMAT(  7X,5(2X,F11.2))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
             /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'VALUE OF EACH LEVEL OF FACTOR ',A,/,&
             /,10X,'LEVEL',7X,'MEAN',6X,'VARIANCE',/,&
              9X,35('-'))
 6050 FORMAT(  7X,I6,2X,F11.2,2X,D15.8)
 6060 FORMAT( /,7X,'MEAN OVER ALL LEVELS = ',F11.2,/,&
             /,7X,'ANALYSIS OF VARIANCE TABLE',/,&
             /,10X,'FACTOR',&
                   8X,'S.S.',9X,'D.F.',6X,'M.S.',8X,'V.R.',8X,'C.R',/,&
              9X,69('-'))
 6070 FORMAT(  10X,A,2(2X,F11.4),3(1X,D11.4))
 6080 FORMAT(  10X,A,2(2X,F11.4),1X,D11.4,13X,D11.4)
      END
```

(d) Output results

```
*** D42WRN ***

  ** INPUT **

      ISW =      0

      LA =      3     LB =      3

      OBSERVATION MATRIX

               1.26         1.21         1.19
               1.29         1.23         1.23
               1.38         1.27         1.22

  ** OUTPUT **

      IERR =      0

      VALUE OF EACH LEVEL OF FACTOR A
```

| LEVEL | MEAN | VARIANCE |
|---|---|---|
| 1 | 1.22 | 0.13000000D-02 |
| 2 | 1.25 | 0.12000000D-02 |
| 3 | 1.29 | 0.67000000D-02 |

VALUE OF EACH LEVEL OF FACTOR B

| LEVEL | MEAN | VARIANCE |
|---|---|---|
| 1 | 1.31 | 0.39000000D-02 |
| 2 | 1.24 | 0.93333333D-03 |
| 3 | 1.21 | 0.43333333D-03 |

MEAN OVER ALL LEVELS = 1.25

ANALYSIS OF VARIANCE TABLE

| FACTOR | S.S. | D.F. | M.S. | V.R. | C.R |
|---|---|---|---|---|---|
| TOTAL | 0.0258 | 8.0000 | | | |
| A | 0.0074 | 2.0000 | 0.3700D-02 | 0.4723D+01 | 0.2261D+00 |
| B | 0.0153 | 2.0000 | 0.7633D-02 | 0.9745D+01 | 0.5310D+00 |
| ERROR | 0.0031 | 4.0000 | 0.7833D-03 | | 0.2429D+00 |

### 7.3.2 D42WRM, R42WRM
### Two-Way Layout Analysis of Variance (With Missing Values)

(1) **Function**

For factors A and B consisting of $m_a$ and $m_b$ levels respectively and two-way layout data $\{x_{ij}\}(i = 1, \cdots, m_a; j = 1, \cdots, m_b)$ for which there are no repetitions in the combinations of each level and no data has been obtained for combinations of $n_s$ levels, the D42WRM or R42WRM obtains the mean and variance of each level for each factor and the total mean over all levels and performs an analysis of variance. The data values of the combinations of levels for which no data has been obtained are called missing values, and estimates are substituted for the missing values in the calculations of each statistic.

Mean of each level of factor A:

$$\bar{x}_{i\cdot} = \frac{1}{m_b} \sum_{j=1}^{m_b} x_{ij}$$

Variance of each level of factor A:

$$V_{ai} = \frac{1}{\alpha_b} \sum_{j=1}^{m_b} (x_{ij} - \bar{x}_{i\cdot})^2$$

Mean of each level of factor B:

$$\bar{x}_{\cdot j} = \frac{1}{m_a} \sum_{i=1}^{m_a} x_{ij}$$

Variance of each level of factor B:

$$V_{bj} = \frac{1}{\alpha_a} \sum_{i=1}^{m_a} (x_{ij} - \bar{x}_{\cdot j})^2$$

Total mean:

$$\bar{x} = \frac{1}{m_a \cdot m_b} \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} x_{ij}$$

Here, ($\alpha_a = m_a$ and $\alpha_b = m_b$ when a sample variance is used, and $\alpha_a = m_a - 1$ and $\alpha_b = m_b - 1$ when an unbiased variance is used.

Also, the analysis of variance results are defined by the following equations.

Variation:

- Total variation

$$S_T = \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} (x_{ij} - \bar{x})^2$$

- Variation of factor A

$$S_A = m_b \sum_{i=1}^{m_a} (\bar{x}_{i\cdot} - \bar{x})^2$$

- Variation of factor B

$$S_B = m_a \sum_{j=1}^{m_b} (\bar{x}_{\cdot j} - \bar{x})^2$$

- Error variation

  $$S_E = S_T - (S_A + S_B)$$

Degrees of freedom:

- Degrees of freedom of total variation

  $$\phi_T = m_a \cdot m_b - n_s - 1$$

- Degrees of freedom of factor A variation

  $$\phi_A = m_a - 1$$

- Degrees of freedom of factor B variation

  $$\phi_B = m_b - 1$$

- Degrees of freedom of error variation

  $$\phi_E = (m_a - 1) \cdot (m_b - 1) - n_s$$

Unbiased variance:

- Unbiased variance of factor A variation

  $$V_A = \frac{S_A}{\phi_A}$$

- Unbiased variance of factor B variation

  $$V_B = \frac{S_B}{\phi_B}$$

- Unbiased variance of error variation

  $$V_E = \frac{S_E}{\phi_E}$$

Variance ratio:

- Variance ratio for unbiased variance of factor A variation

  $$F_A = \frac{V_A}{V_E}$$

- Variance ratio for unbiased variance of factor B variation

  $$F_B = \frac{V_B}{V_E}$$

Contribution ratio:

- Contribution ratio of factor A variation

  $$P_A = \frac{S_A - \phi_A \cdot V_E}{S_T}$$

- Contribution ratio of factor B variation

  $$P_B = \frac{S_B - \phi_B \cdot V_E}{S_T}$$

- Contribution ratio of factor error variation

  $$P_E = 1 - P_A - P_B$$

Missing value estimates:
The missing value estimates are determined so that the error variation $S_E$ is minimized. To obtain estimates that will minimize $S_E$, you should solve the following equation with the missing values $x_{st}$ $((s,t) \in S)$ as unknowns.

$$\frac{\partial S_E}{\partial x_{st}} = 0 \qquad ((s,t) \in S)$$

Here, $S$ is assumed to be the set of combinations of levels for which missing values occurred. Since $S_E$ is a quadratic equation, this equation is a set of $n_s$ simultaneous linear equations with the missing values as unknowns.
For example, for the following two-way layout data in which $x_{13}$ and $x_{22}$ are missing values,

$$X = \begin{bmatrix} 1.0 & 1.1 & x_{13} \\ 1.2 & x_{22} & 1.3 \\ 1.1 & 1.0 & 1.3 \end{bmatrix}$$

the error variation is as follows.

$$S_E = 1.78 - 1.4x_{13} - 1.4x_{22} + \frac{4}{9}x_{13}^2 + \frac{4}{9}x_{22}^2 + \frac{2}{9}x_{13}x_{22}$$

Differentiating this with respect to $x_{13}$ and $x_{22}$ produces the following simultaneous linear equations.

$$\frac{\partial S_E}{\partial x_{13}} = -1.4 + \frac{8}{9}x_{13} + \frac{2}{9}x_{22} = 0$$

$$\frac{\partial S_E}{\partial x_{22}} = -1.4 + \frac{2}{9}x_{13} + \frac{8}{9}x_{22} = 0$$

Solving these simultaneous linear equations yields the missing value estimates $x_{13} = 1.26$ and $x_{22} = 1.26$.

(2) **Usage**
Double precision:
CALL D42WRM (A, NA, LA, LB, IST, ISN, STATA, STATB, X1, V, ISW, IWK, WK, IERR)
Single precision:
CALL R42WRM (A, NA, LA, LB, IST, ISN, STATA, STATB, X1, V, ISW, IWK, WK, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: { INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer }
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | {D, R} | NA,LB | Input | Matrix in which observed values are stored ($x_{ij}$) (See Note (b)). |
| | | | | Output | Matrix in which observed values are stored ($x_{ij}$). However, estimates are stored in the elements corresponding to missing values. |
| 2 | NA | I | 1 | Input | Adjustable dimension of array A. |
| 3 | LA | I | 1 | Input | Number of levels of factor A $m_a$ |
| 4 | LB | I | 1 | Input | Number of levels of factor B $m_b$ |
| 5 | IST | I | ISN,2 | Input | Information about combinations of levels for which missing values occurred. (See Note (a)) |
| 6 | ISN | I | 1 | Input | Number of missing values $n_s$ |
| 7 | STATA | {D, R} | LA,2 | Output | Mean and variance of each level of factor A (See Note (c)) |
| 8 | STATB | {D, R} | LB,2 | Output | Mean and variance of each level of factor B (See Note (c)) |
| 9 | X1 | {D, R} | 1 | Output | Total mean |
| 10 | V | {D, R} | 16 | Output | Analysis of variance table<br>(For the storage method of the analysis of variance table, see Table 7−3.) |
| 11 | ISW | I | 1 | Input | Processing switch<br>0: Calculate the unbiased variance<br>1: Calculate the sample variance |
| 12 | IWK | I | See Contents | Work | Work area<br>**Size:**LA $\times$ LB $+$ ISN |
| 13 | WK | {D, R} | See Contents | Work | Work area<br>**Size:**ISN$^2 + 2 \times$ ISN $+$ LA $+$ LB $+ 1$ |
| 14 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $= 0,\ 1$

(b) NA $\geq$ LA $\geq 2$

(c) LB $\geq 2$

(d) $1 \leq$ ISN $<$ (LA $-$ 1) $\times$ (LB $-$ 1)

(e) $1 \leq \mathrm{IST}(\mathrm{i}, 1) \leq \mathrm{LA}\ (\mathrm{i} = 1, 2, \cdots, \mathrm{ISN})$

(f) $1 \leq \mathrm{IST}(\mathrm{i}, 2) \leq \mathrm{LB}\ (\mathrm{i} = 1, 2, \cdots, \mathrm{ISN})$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (a) was not satisfied. | Processing continues with ISW=0. |
| 2000 | Estimates of missing values could not be calculated. | The missing value data is replaced by the mean of the non-missing value data, and processing continues. |
| 3000 | Any of restrictions (b) to (f) was not satisfied. | Processing is aborted. |

(6) **Notes**

(a) The factor A level of the i-th missing value is stored in IST(i, 1) and the factor B level is stored in IST(i, 2).

(b) The observed values ($x_{ij}$) are stored in array A as a real matrix (two-dimensional array type) (See Appendix A).

(c) The mean and variance of each level of factors A and B are stored as follows in the arrays STATA and STATB.

$$\left.\begin{array}{lll} \text{STATA}(i,\ 1) & : & \text{Mean of each level of factor A } \bar{x}_{i\cdot} \\ \text{STATA}(i,\ 2) & : & \text{Variance of each level of factor A } V_{ai} \\ \text{STATB}(j,\ 1) & : & \text{Mean of each level of factor B } \bar{x}_{\cdot j} \\ \text{STATB}(j,\ 2) & : & \text{Variance of each level of factor B } V_{bj} \end{array}\right., \quad i = 1, \cdots, \text{LA}; j = 1, \cdots, \text{LB}$$

(d) The analysis of variance table elements are stored as follows in the array V.

Table 7−3 Analysis of Variance Table Storage Status

| Element | Variation | Degrees of freedom | Unbiased variance | Variance ratio | Contribution ratio |
|---|---|---|---|---|---|
| Total | V(1) | V(5) | | | |
| Factor A | V(2) | V(6) | V(9) | V(12) | V(14) |
| Factor B | V(3) | V(7) | V(10) | V(13) | V(15) |
| Error | V(4) | V(8) | V(11) | | V(16) |

(e) Statistics obtained when an unbiased variance is calculated can be applied to a population for which sampling with replacement is performed from an infinite or finite population. Statistics obtained when a sample variance is calculated can be applied to a population for which the population and sample match.

(7) **Example**

(a) Problem

Given two-way layout data for which there are no repetitions and having factors A and B as shown in matrix $X$ below, obtain the mean and variance of each level of each factor and the total mean over all levels and perform an analysis of variance. However, $*$ indicates that the corresponding data is a missing value.

$$X = \begin{bmatrix} 4 & 6 & 8 & 10 & 12 \\ 7 & 10 & * & 16 & * \\ * & 14 & 18 & 22 & 26 \\ 13 & 18 & 23 & 28 & 33 \end{bmatrix}$$

(b) Input data

Two-way layout data $X$, NA=4, LA=4, LB=5, ISN=3, ISW=0,

IST(1,1)=2, IST(1,2)=3, IST(2,1)=4, IST(2,2)=1 and IST(2,1)=2, IST(2,2)=5.

(c) Main program

```
      PROGRAM N42WRM
      IMPLICIT NONE
!
      INTEGER NN,MM,NS,NW,NIW
      PARAMETER(NN=4,MM=5,NS=3,NW=25,NIW=23)
      REAL(8) A(NN,MM),W1(NW),STATA(NN,2),STATB(MM,2)
      REAL(8) X1,V(16)
!     REAL(8) C0,C1
!     PARAMETER(C0=0.0D0,C1=1.0D0)
      INTEGER NA,LA,LB,ISN,IST(NS,2),IW1(NIW)
      INTEGER I,J,L,ISW,IERR
!     WK1 = 0.0D0
!
!     EXAMPLE OF D/R42WRM
!
      WRITE(6,5000)
      WRITE(6,5001)
      IERR = 0
      READ(5,*) ISW
      READ(5,*) NA
      READ(5,*) LA
      READ(5,*) LB
      READ(5,*) ISN
      WRITE(6,5100) ISW,NA,LA,LB,ISN
      WRITE(6,5200)
      DO 100 I = 1,ISN
         READ(5,*) IST(I,1),IST(I,2)
         WRITE(6,5250) IST(I,1),IST(I,2)
 100  CONTINUE
      WRITE(6,5260)
      DO 110 I=1,LA
         READ(5,*) (A(I,J),J=1,LB)
         WRITE(6,5270) (A(I,J),J=1,LB)
 110  CONTINUE
      CALL D42WRM&
      (A,NA,LA,LB,IST,ISN,STATA,STATB,X1,V,ISW,IW1,W1,IERR)
      WRITE(6,5300) IERR
      WRITE(6,5350)
      DO 130 I=1,ISN
         WRITE(6,5360) IST(I,1),IST(I,2),A(IST(I,1),IST(I,2))
 130  CONTINUE
      WRITE(6,5400)
      WRITE(6,5500) X1
      WRITE(6,5600)
      WRITE(6,7400) (STATA(L,1),L=1,LA)
      WRITE(6,5800)
      WRITE(6,7400) (STATA(L,2),L=1,LA)
      WRITE(6,5610)
      WRITE(6,7500) (STATB(L,1),L=1,LB)
      WRITE(6,5810)
      WRITE(6,7500) (STATB(L,2),L=1,LB)
      WRITE(6,5900)
      WRITE(6,6050)
      WRITE(6,6100) V(1),V(5)
      WRITE(6,6200) V(2),V(6),V(9),V(12),V(14)
      WRITE(6,6210) V(3),V(7),V(10),V(13),V(15)
      WRITE(6,6300) V(4),V(8),V(11),V(16)
 5000 FORMAT( /,' **** D42WRM ****')
 5001 FORMAT( /,' ** INPUT **',/)
 5100 FORMAT( /,7X,'ISW =',I6,/,&
              /,7X,'NA  = ',I6&
                  ,5X,'LA  = ',I6&
                  ,5X,'LB  = ',I6&
                  ,5X,'ISN = ',I6)
 5200 FORMAT( /,/,6X,'MISSED VALUES',/)
 5250 FORMAT( /,8X,'A(',I1,',',I1,')',/)
 5260 FORMAT( /,6X,'OBSERVATION MATRIX',/)
 5270 FORMAT( /,8X,5(D11.5,1X),/)
 5300 FORMAT( /,' ** OUTPUT **',/,&
              /,7X,'IERR = ',I6)
 5350 FORMAT( /,/,' ESTIMATED MISSED VALUES',/)
 5360 FORMAT( /,3X,'A(',I1,',',I1,') = ',D15.8,/)
 5400 FORMAT( /,/,' MEAN OVER ALL LEVELS',/,/)
 5500 FORMAT(  '           ',D11.5,' ')
 5600 FORMAT( /,/,' MEAN IN EACH LEVEL OF FACTOR A',/,/)
 5610 FORMAT( /,/,' MEAN IN EACH LEVEL OF FACTOR B',/,/)
 5800 FORMAT( /,/,' VARIANCE IN EACH LEVEL OF FACTOR A',/,/)
 5810 FORMAT( /,/,' VARIANCE IN EACH LEVEL OF FACTOR B',/,/)
 5900 FORMAT( /,/,' ANALYSIS-OF-VARIANCE TABLE',/,/)
 6050 FORMAT(  ' ',&
      ' FACTOR  ',4X,'S.S.',9X,'D.F.',9X,'U.V.',9X,'R.V.',9X,'R.C',/,/)
 6100 FORMAT(  ' ',&
      ' TOTAL   ',1X,D11.5,2X,D11.5)
 6200 FORMAT(  ' ',&
      '   A     ',5(2X,D11.5))
 6210 FORMAT(  ' ',&
```

```
     '   B    ',5(2X,D11.5))
6300 FORMAT(  ' ',&
     ' ERROR  ',3(2X,D11.5),15X,D11.5)
7300 FORMAT(  '          ',3(' ',D11.5,' '))
7400 FORMAT(  '          ',4(' ',D11.5,' '))
7500 FORMAT(  '          ',5(' ',D11.5,' '))
     STOP
     END
```

(d) Output results

```
**** D42WRM ****

** INPUT **

      ISW =     0

      NA  =      4     LA  =      4     LB  =      5     ISN =      3

     MISSED VALUES

      A(2,3)

      A(4,1)

      A(2,5)

     OBSERVATION MATRIX

      0.40000D+01 0.60000D+01 0.80000D+01 0.10000D+02 0.12000D+02

      0.70000D+01 0.10000D+02 0.00000D+00 0.16000D+02 0.00000D+00

      0.00000D+00 0.14000D+02 0.18000D+02 0.22000D+02 0.26000D+02

      0.13000D+02 0.18000D+02 0.23000D+02 0.28000D+02 0.33000D+02

** OUTPUT **
      IERR =      0

ESTIMATED MISSED VALUES

    A(2,3) =  0.14408805D+02

    A(4,1) =  0.14320755D+02

    A(2,5) =  0.21742138D+02

MEAN OVER ALL LEVELS

         0.15274D+02

MEAN IN EACH LEVEL OF FACTOR A

          0.80000D+01  0.13830D+02  0.16000D+02  0.23264D+02

VARIANCE IN EACH LEVEL OF FACTOR A

          0.10000D+02  0.32241D+02  0.10000D+03  0.56245D+02

MEAN IN EACH LEVEL OF FACTOR B

          0.63302D+01  0.12000D+02  0.15852D+02  0.19000D+02  0.23186D+02

VARIANCE IN EACH LEVEL OF FACTOR B

          0.36600D+02  0.26667D+02  0.39815D+02  0.60000D+02  0.77148D+02
```

ANALYSIS-OF-VARIANCE TABLE

| FACTOR | S.S. | D.F. | U.V. | R.V. | R.C |
|---|---|---|---|---|---|
| TOTAL | 0.13908D+04 | 0.16000D+02 | | | |
| A | 0.59683D+03 | 0.30000D+01 | 0.19894D+03 | 0.14455D+02 | 0.39945D+00 |
| B | 0.67008D+03 | 0.40000D+01 | 0.16752D+03 | 0.12172D+02 | 0.44222D+00 |
| ERROR | 0.12386D+03 | 0.90000D+01 | 0.13762D+02 | | 0.15833D+00 |

### 7.3.3 D42WR1, R42WR1 Two-Way Layout Analysis of Variance (Repetition Data)

(1) **Function**

For factors A and B consisting of $m_a$ and $m_b$ levels respectively and two-way layout data $\{x_{kij}\}(k = 1, \cdots, n_{ij}; i = 1, \cdots, m_a; j = 1, \cdots, m_b)$ for which the number of repetitions in the combination of each level is $n_{ij}$, the D42WR1 or R42WR1 obtains the mean and variance of each level for each factor and the total mean over all levels and performs an analysis of variance.

The mean of the repetitions in the combination of each level, the mean and variance of each level for each factor, and the total mean over all levels are defined by the following equations.

Mean of the repetitions in the combination of each level :

$$\bar{x}_{\cdot ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{kij}$$

Mean of each level of factor A :

$$\bar{x}_{\cdot i\cdot} = \frac{1}{\displaystyle\sum_{j=1}^{m_b} n_{ij}} \sum_{j=1}^{m_b} \sum_{k=1}^{n_{ij}} x_{kij}$$

Variance of each level of factor A :

$$V_{ai} = \frac{1}{\alpha_i} \sum_{j=1}^{m_b} \sum_{k=1}^{n_{ij}} (x_{kij} - \bar{x}_{\cdot i\cdot})^2$$

Mean of each level of factor B :

$$\bar{x}_{\cdot\cdot j} = \frac{1}{\displaystyle\sum_{i=1}^{m_a} n_{ij}} \sum_{i=1}^{m_a} \sum_{k=1}^{n_{ij}} x_{kij}$$

Variance of each level of factor B :

$$V_{bj} = \frac{1}{\beta_j} \sum_{i=1}^{m_a} \sum_{k=1}^{n_{ij}} (x_{kij} - \bar{x}_{\cdot\cdot j})^2$$

Total mean :

$$\bar{x} = \frac{1}{\displaystyle\sum_{i=1}^{m_a} \sum_{j=1}^{m_b} n_{ij}} \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} \sum_{k=1}^{n_{ij}} x_{kij}$$

Here, $alpha_i$ and $beta_j$ are given by

$$\alpha_i = \sum_{j=1}^{m_b} n_{ij}, \quad \beta_j = \sum_{i=1}^{m_a} n_{ij}$$

when a sample variance is used, and $alpha_i$ and $beta_j$ are given by

$$\alpha_i = \sum_{j=1}^{m_b} n_{ij} - 1, \quad \beta_j = \sum_{i=1}^{m_a} n_{ij} - 1$$

when an unbiased variance is used.
Also, the analysis of variance results are defined by the following equations.
Variation :

- Total variation

$$S_T = \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} \sum_{k=1}^{n_{ij}} (x_{kij} - \bar{x})^2$$

- Variation of factor A

$$S_A = \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} n_{ij} (\bar{x}_{\cdot i \cdot} - \bar{x})^2$$

- Variation of factor B

$$S_B = \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} n_{ij} (\bar{x}_{\cdot \cdot j} - \bar{x})^2$$

- Interaction variation

$$S_{A \times B} = S_{AB} - (S_A + S_B)$$

  Here,

$$S_{AB} = \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} n_{ij} (\bar{x}_{\cdot ij} - \bar{x})^2$$

- Error variation

$$S_E = S_T - S_{AB}$$

Degrees of freedom :

- Degrees of freedom of total variation

$$\phi_T = m_a \cdot m_b - 1$$

- Degrees of freedom of factor A variation

$$\phi_A = m_a - 1$$

- Degrees of freedom of factor B variation

$$\phi_B = m_b - 1$$

- Degrees of freedom of interaction variation

$$\phi_{A \times B} = (m_a - 1)(m_b - 1)$$

- Degrees of freedom of error variation

$$\phi_E = \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} n_{ij} - m_a \cdot m_b$$

Unbiased variance :

- Unbiased variance of factor A variation

$$V_A = \frac{S_A}{\phi_A}$$

- Unbiased variance of factor B variation

$$V_B = \frac{S_B}{\phi_B}$$

- Unbiased variance of interaction variation

$$V_{A\times B} = \frac{S_{A\times B}}{\phi_{A\times B}}$$

- Unbiased variance of error variation

$$V_E = \frac{S_E}{\phi_E}$$

Variance ratio :

- Variance ratio for unbiased variance of factor A variation

$$F_A = \frac{V_A}{V_E}$$

- Variance ratio for unbiased variance of factor B variation

$$F_B = \frac{V_B}{V_E}$$

- Variance ratio for unbiased variance of interaction variation

$$F_{A\times B} = \frac{V_{A\times B}}{V_E}$$

Contribution ratio :

- Contribution ratio of factor A variation

$$P_A = \frac{S_A - \phi_A \cdot V_E}{S_T}$$

- Contribution ratio of factor B variation

$$P_B = \frac{S_B - \phi_B \cdot V_E}{S_T}$$

- Contribution ratio of interaction variation

$$P_{A\times B} = \frac{S_{A\times B} - \phi_{A\times B} \cdot V_E}{S_T}$$

- Contribution ratio of error variation

$$P_E = 1 - P_A - P_B - P_{A\times B}$$

(2) **Usage**

Double precision:

CALL D42WR1 (A, NA, MA, LA, LB, N, NR, Y, STATA, STATB, X1, V, ISW, WK, IERR)

Single precision:

CALL R42WR1 (A, NA, MA, LA, LB, N, NR, Y, STATA, STATB, X1, V, ISW, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: { INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer }

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | {D / R} | See Contents | Input | Observed values ($x_{kij}$)<br>**Size**: NA, MA, LB |
| 2 | NA | I | 1 | Input | Adjustable dimension of first dimension of array A |
| 3 | MA | I | 1 | Input | Adjustable dimension of second dimension of array A |
| 4 | LA | I | 1 | Input | Number of levels of factor A $m_a$ |
| 5 | LB | I | 1 | Input | Number of levels of factor B $m_b$ |
| 6 | N | I | MA,LB | Input | Number of repetitions in the combination $(i, j)$ of each level $n_{ij}$<br>(not used when NR $\geq 1$) |
| 7 | NR | I | 1 | Input | Number of repetitions when the numbers of repetitions in the combination of each level are equal<br>$n_{11} = \cdots = n_{1m_b} = n_{21} = \cdots = n_{2m_b} = \cdots = n_{m_a1} = \cdots = n_{m_am_b}$<br>When the numbers of repetitions in the combination of each level are not equal, set a value that is less than or equal to zero. |
| 8 | Y | {D / R} | MA,LB | Output | Mean of the repetitions in the combination $(i, j)$ of each level $\bar{x}_{\cdot ij}$ |
| 9 | STATA | {D / R} | LA,2 | Output | Mean and variance of each level of factor A (See Note (a)) |
| 10 | STATB | {D / R} | LB,2 | Output | Mean and variance of each level of factor B (See Note (a)) |
| 11 | X1 | {D / R} | 1 | Output | Total mean $\bar{x}$ |
| 12 | V | {D / R} | 21 | Output | Analysis of variance table<br>(For the storage method of the analysis of variance table, see Table 7−4 in Note (b).) |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 13 | ISW | I | 1 | Input | Processing switch<br>0: Calculate the unbiased variance<br>1: Calculate the sample variance |
| 14 | WK | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | LA | Work | Work area |
| 15 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) ISW $= 0, 1$

(b) MA $\geq$ LA $\geq 1$

(c) LB $\geq 1$

(d) NR<1 and NA$\geq$N$(i,j) \geq 1$ $\quad (i = 1,\cdots,$LA ; $j$ =1, $\ldots$,LB)
or NA$\geq$NR$\geq 1$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (a) was not satisfied. | Processing continues with ISW=0. |
| 1010 | LA=1 or LB=1 was specified. | The absolute value maximum that can be represented is set for V(11) to V(21). For the values stored in STATA and STATB. (See Note (d)) |
| 3000 | Any of restriction (b) to (d) was not satisfied. | Processing is aborted. |

(6) **Notes**

(a) The mean and variance of each level of factors A and B are stored as follows in the arrays STATA and STATB.

STATA($i$, 1) : Mean of each level of factor A $\bar{x}_{\cdot i \cdot}$
STATA($i$, 2) : Variance of each level of factor A $V_{ai}$
STATB($j$, 1) : Mean of each level of factor B $\bar{x}_{\cdot \cdot j}$
STATB($j$, 2) : Variance of each level of factor B $V_{bj}$

, $\quad i = 1, \cdots,$ LA; $j = 1, \cdots,$ LB

(b) The analysis of variance table elements are stored as follows in the array V.

Table 7−4 Analysis of Variance Table Storage Status

| Factor | Variation | Degrees of freedom | Unbiased variance | Variance ratio | Contribution ratio |
|---|---|---|---|---|---|
| Total | V(1) | V(6) | | | |
| Factor A | V(2) | V(7) | V(11) | V(15) | V(18) |
| Factor B | V(3) | V(8) | V(12) | V(16) | V(19) |
| A×B | V(4) | V(9) | V(13) | V(17) | V(20) |
| Error | V(5) | V(10) | V(14) | | V(21) |

(c) Statistics obtained when calculations are performed using an unbiased estimate can be applied to a population for which sampling with replacement is performed from an infinite or finite population. Statistics obtained when calculations are performed using a sample variance can be applied to a population for which the population and sample match.

(d) If IERR=1010 occurred when ISW=0 was specified, the following processing is performed.

i. When NR=1 and LA=1 :
The absolute value maximum that can be represented is set for all STATB($j$, 2) ($j = 1, \cdots$,LB).

ii. When NR=1 and LB=1 :
The absolute value maximum that can be represented is set for all STATA($i$, 2) ($i = 1, \cdots$,LA).

iii. When N(1, $j$)=1 for a $j$ for which NR<1 and LA=1:
The absolute value maximum that can be represented is set for STATB($j$, 2).

iv. When N($i$, 1)=1 for an $i$ for which NR<1 and LB=1:
The absolute value maximum that can be represented is set for STATA($i$, 2).

v. When none of the conditions in 6(d)i. to 6(d)iv. occurs:
All STATA($i$, 2) ($i = 1, \cdots$,LA) and STATB($j$, 2)($j = 1, \cdots$,LB) are calculated normally.

(7) **Example**

(a) Problem
Given two-way layout data for which the number of repetitions is 2 and having factors A and B as shown in matrices $X_1$ and $X_2$ below, obtain the mean in the combinations of each level, mean and variance of each level of each factor, and the total mean over all levels and perform an analysis of variance.

$$X_1 = \begin{bmatrix} 7.9 & 9.8 & 13.2 & 13.1 \\ 10.3 & 14.6 & 15.9 & 9.5 \\ 9.1 & 12.1 & 11.1 & 7.0 \end{bmatrix}$$

$$X_2 = \begin{bmatrix} 8.7 & 9.3 & 14.0 & 12.0 \\ 11.0 & 14.0 & 14.6 & 8.5 \\ 8.6 & 12.9 & 10.0 & 8.2 \end{bmatrix}$$

(b) Input data
Two-way layout data $X_1$ and $X_2$,
NA=100, MA=5, LA=3, LB=4, NR=2 and ISW=0.

(c) Main program

```
      PROGRAM B42WR1
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 100, MA = 5, LA = 3, LB = 4 )
      DIMENSION A(NA,MA,LB),N(MA,LB),Y(MA,LB)
      DIMENSION STATA(LA,2),STATB(LB,2),V(21),WK(MA)
!
      WRITE(6,6000)
      IERR =    0
      READ(5,*) NR
      READ(5,*) ISW
      DO 100 K=1,NR
      DO 101 I=1,LA
         READ(5,*) (A(K,I,J),J=1,LB)
  101 CONTINUE
  100 CONTINUE
      DO 110 I = 1,LA
      DO 111 J = 1,LB
         N(I,J) = 0
  111 CONTINUE
  110 CONTINUE
      WRITE(6,6010) ISW,LA,LB,NR
      WRITE(6,6020)
      DO 120 I = 1,LA
         WRITE(6,6030) (N(I,J),J=1,LB)
  120 CONTINUE
      WRITE(6,6040)
      DO 130 K=1,NR
         WRITE(6,6050) K
         DO 140 I=1,LA
            WRITE(6,6060) (A(K,I,J),J=1,LB)
  140    CONTINUE
  130 CONTINUE
      CALL D42WR1&
      (A,NA,MA,LA,LB,N,NR,Y,STATA,STATB,X1,V,ISW,WK,IERR)
      WRITE(6,6070) IERR
      WRITE(6,6080)
      DO 150 I = 1,LA
         WRITE(6,6090) (Y(I,J),J=1,LB)
  150 CONTINUE
      WRITE(6,6100) 'A'
      DO 160 L=1,LA
         WRITE(6,6110) L,STATA(L,1),STATA(L,2)
  160 CONTINUE
      WRITE(6,6100) 'B'
      DO 170 L=1,LB
         WRITE(6,6110) L,STATB(L,1),STATB(L,2)
  170 CONTINUE
      WRITE(6,6120) X1
      WRITE(6,6130) 'TOTAL',V(1),V(6)
      WRITE(6,6130) '  A  ',V(2),V(7),V(11),V(15),V(18)
      WRITE(6,6130) '  B  ',V(3),V(8),V(12),V(16),V(19)
      WRITE(6,6130) ' AXB ',V(4),V(9),V(13),V(17),V(20)
      WRITE(6,6140) 'ERROR',V(5),V(10),V(14),V(21)
!
      STOP
 6000 FORMAT(   ' *** D42WR1 ***',/,&
             /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'ISW = ',I6,/,&
             /,7X,'LA = ',I6,5X,'LB = ',I6,/,&
             /,7X,'NR = ',I6)
 6020 FORMAT( /,7X,'NUMBER OF REPETITION N(I,J)',/)
 6030 FORMAT(  9X,5(2X,I6))
 6040 FORMAT( /,7X,'OBSERVATION MATRIX')
 6050 FORMAT( /,9X,'A(',I1,',I,J)',/)
 6060 FORMAT(  7X,5(2X,F11.2))
 6070 FORMAT( /,3X,'** OUTPUT **',/,&
             /,7X,'IERR = ',I6)
 6080 FORMAT( /,7X,'MEAN FOR REPETITION',/)
 6090 FORMAT(  7X,5(2X,F11.2))
 6100 FORMAT( /,7X,'VALUE OF EACH LEVEL OF FACTOR ',A,/,&
             /,10X,'LEVEL',7X,'MEAN',6X,'VARIANCE',/,&
              9X,35('-'))
 6110 FORMAT(  7X,I6,2X,F11.2,2X,D15.8)
 6120 FORMAT( /,7X,'MEAN OVER ALL LEVELS = ',D15.8,/,&
             /,7X,'ANALYSIS OF VARIANCE TABLE',/,&
             /,10X,'FACTOR',&
                 6X,'S.S.',9X,'D.F.',7X,'M.S.',8X,'V.R.',8X,'C.R.',/,&
              9X,69('-'))
 6130 FORMAT(  10X,A,2(2X,F11.2),3(1X,D11.4))
 6140 FORMAT(  10X,A,2(2X,F11.2),1X,D11.4,13X,D11.4)
 6150 FORMAT( /,7X,'VARIANCE OF EACH LEVEL OF FACTOR ',A,/)
      END
```

(d) Output results

```
*** D42WR1 ***
  ** INPUT **
      ISW =      0
      LA =      3     LB =      4
      NR =      2
      NUMBER OF REPETITION N(I,J)
               0       0       0       0
               0       0       0       0
               0       0       0       0
      OBSERVATION MATRIX
        A(1,I,J)
                7.90          9.80         13.20         13.10
               10.30         14.60         15.90          9.50
                9.10         12.10         11.10          7.00
        A(2,I,J)
                8.70          9.30         14.00         12.00
               11.00         14.00         14.60          8.50
                8.60         12.90         10.00          8.20
  ** OUTPUT **
      IERR =      0
      MEAN FOR REPETITION
                8.30          9.55         13.60         12.55
               10.65         14.30         15.25          9.00
                8.85         12.50         10.55          7.60
      VALUE OF EACH LEVEL OF FACTOR A
         LEVEL       MEAN      VARIANCE
        -----------------------------------
           1        11.00   0.54971429D+01
           2        12.30   0.77714286D+01
           3         9.88   0.41307143D+01
      VALUE OF EACH LEVEL OF FACTOR B
         LEVEL       MEAN      VARIANCE
        -----------------------------------
           1         9.27   0.13466667D+01
           2        12.12   0.47256667D+01
           3        13.13   0.49026667D+01
           4         9.72   0.55736667D+01
      MEAN OVER ALL LEVELS =  0.11058333D+02
      ANALYSIS OF VARIANCE TABLE
         FACTOR      S.S.         D.F.       M.S.        V.R.        C.R.
        ---------------------------------------------------------------------
         TOTAL      145.36       23.00
           A         23.56        2.00  0.1178D+02  0.2879D+02  0.1565D+00
           B         62.62        3.00  0.2087D+02  0.5101D+02  0.4223D+00
          AXB        54.27        6.00  0.9045D+01  0.2211D+02  0.3565D+00
         ERROR        4.91       12.00  0.4092D+00              0.6474D-01
```

## 7.4 MULTIPLE-WAY LAYOUT

### 7.4.1 D4MWRF, R4MWRF
### Multiple-Way Layout Analysis of Variance

(1) **Function**

For $m$ factors $A_1, A_2, \cdots, A_m$ (where $m$ is at most 6) consisting of $l_1, l_2, \cdots, l_m$ levels respectively and multi-way layout data $x_{kj_1j_2\cdots j_m}(k = 1, \cdots, n; i = 1, \cdots, m; j_i = 1, \cdots, l_i)$ for which the number of repetitions in the combination of each factor level is a fixed value $n$, the D4MWRF or R4MWRF performs an analysis of variance. At this time, specified interactions can be confounded. For the explanations below, the operations $\Sigma_i$ and $\Delta_i$ are defined as follows.

$$\Sigma_i \equiv \sum_{j_i=1}^{l_i}, \quad \Delta_i \equiv l_i - \sum_{j_i=1}^{l_i}$$

The mean of the repetitions in the combination of each level and the total mean over all levels are defined by the following equations.

Mean of the repetitions in the combination of each level :

$$\bar{x}_{\cdot j_1j_2\cdots j_m} = \frac{1}{n}\sum_{k=1}^{n} x_{kj_1j_2\cdots j_m}$$

Total mean :

$$\bar{x} = \frac{1}{n\prod_{i=1}^{m} l_i}\Sigma_1\Sigma_2\cdots\Sigma_m\sum_{k=1}^{n} x_{kj_1j_2\cdots j_m}$$

Also, the analysis of variance results are defined by the following equations.

Variation :

- Total variation

$$S_T = \Sigma_1\Sigma_2\cdots\Sigma_m\sum_{k=1}^{n}(x_{kj_1j_2\cdots j_m} - \bar{x})^2$$

- Variation of factor $A_i$

$$S_{A_i} = \frac{n}{l_i\prod_{j=1}^{m} l_j}\Sigma_i(\Sigma_1\Sigma_2\cdots\Sigma_{i-1}\Delta_i\Sigma_{i+1}\cdots\Sigma_m\bar{x}_{\cdot j_1j_2\cdots j_m})^2$$

- Variation of the interaction $A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}$ is of $s(s \le m)$ factors $A_{i_1}, A_{i_2}, \cdots, A_{i_s}$

$$\begin{aligned} S_{A_{i_1}\times A_{i_2}\times\cdots\times A_{i_s}} &= \frac{n}{(\prod_{k=1}^{s} l_{i_s})\cdot(\prod_{k=1}^{m} l_k)} \\ &\times \Sigma_{i_1}\Sigma_{i_2}\cdots\Sigma_{i_s}(\Sigma_1\cdots\Sigma_{i_1-1}\Delta_{i_1}\Sigma_{i_1+1}\cdots \\ &\quad \Sigma_{i_2-1}\Delta_{i_2}\Sigma_{i_2+1}\cdots\Sigma_{i_s-1}\Delta_{i_s}\Sigma_{i_s+1}\cdots\Sigma_m\bar{x}_{\cdot j_1j_2\cdots j_m})^2 \end{aligned}$$

- Error variation
  When there are no repetitions :

$$\begin{aligned} S_E &= S_T - \text{(Sum of variations of each factor)} \\ &- \text{(Sum of variations of interactions other than the highest order interaction} \\ &\quad [A_1 \times A_2 \times \cdots \times A_m]) \end{aligned}$$

  When there are repetitions :

$$S_E = S_T - \text{(Sum of variations of each factor)} - \text{(Sum of variations of interactions)}$$

Degrees of freedom :

- Degrees of freedom of total variation

$$\phi_T = n \prod_{i=1}^{m} l_i - 1$$

- Degrees of freedom of factor $A_i$ variation

$$\phi_A = l_i - 1$$

- Degrees of freedom of variation of the interaction $A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}$

$$\phi_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}} = \prod_{k=1}^{s} (l_{i_k} - 1)$$

- Degrees of freedom of error variation
  When there are no repetitions :

$$\begin{aligned} \phi_E &= \phi_T - \text{(Sum of degrees of freedom of variations of each factor)} \\ &- \text{(Sum of degrees of freedom of variations of interactions other than the highest} \\ &\quad \text{order interaction)} \end{aligned}$$

  When there are repetitions :

$$\begin{aligned} \phi_E &= \phi_T - \text{(Sum of degrees of freedom of variations of each factor)} \\ &- \text{(Sum of degrees of freedom of variations of interactions)} \end{aligned}$$

Unbiased variance :

- Unbiased variance of factor $A_i$ variation

$$V_{A_i} = \frac{S_{A_i}}{\phi_{A_i}}$$

- Unbiased variance of variation of the interaction $A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}$

$$V_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}} = \frac{S_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}}}{\phi_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}}}$$

  However, the unbiased variance of the variation of the highest order interaction can be defined only when there are repetitions.

- Unbiased variance of error variation

$$V_E = \frac{S_E}{\phi_E}$$

Variance ratio :

- Variance ratio for unbiased variance of factor $A_i$ variation

  $$F_{A_i} = \frac{V_{A_i}}{V_E}$$

- Variance ratio for unbiased variance of variation of the interaction $A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}$

  $$F_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}} = \frac{V_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}}}{V_E}$$

  However, the variance ratio of the variation of the highest order interaction can be defined only when there are repetitions.

Contribution ratio :

- Contribution ratio of factor $A_i$ variation

  $$P_{A_i} = \frac{S_{A_i} - \phi_{A_i} \cdot V_E}{S_T}$$

- Contribution ratio of variation of the interaction $A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}$

  $$P_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}} = \frac{S_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}} - \phi_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}} \cdot V_E}{S_T}$$

- Contribution ratio of error variation
  When there are no repetitions :

  $$\begin{aligned} P_E &= \phi_T - (\text{Sum of contribution ratios of variations of each factor}) \\ &- (\text{Sum of contribution ratios of variations of interactions other than the highest} \\ &\quad \text{order interaction}) \end{aligned}$$

  When there are repetitions :

  $$\begin{aligned} P_E &= \phi_T - (\text{Sum of contribution ratios of variations of each factor}) \\ &- (\text{Sum of contribution ratios of variations of interactions}) \end{aligned}$$

(2) **Usage**

Double precision:

CALL D4MWRF (A, NA, N, LT, M, IPT, IPN, Y, X1, V, WK, IERR)

Single precision:

CALL R4MWRF (A, NA, N, LT, M, IPT, IPN, Y, X1, V, WK, IERR)

## (3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\begin{Bmatrix}\text{D}\\ \text{R}\end{Bmatrix}$ | NA,M1 | Input | Observed values ($x_{k\beta}$) (See Note (a)) Here, $\text{M1} = \prod_{\text{i}=1}^{\text{M}} \text{LT(i)}$ |
| 2 | NA | I | 1 | Input | Adjustable dimension of array A |
| 3 | N | I | 1 | Input | Number of repetitions in the combination of each level $n$ |
| 4 | LT | I | M | Input | Number of levels of each factor $l_i$ |
| 5 | M | I | 1 | Input | Number of factors $m$ |
| 6 | IPT | I | M2 | Input | Factor numbers of factors to be confounded (See Note (b)) Here, when IPN>0, M2=IPN, and when IPN=0, the argument IPT should be a dummy argument. |
| 7 | IPN | I | 1 | Input | Number of factors to be confounded |
| 8 | Y | $\begin{Bmatrix}\text{D}\\ \text{R}\end{Bmatrix}$ | M1 | Output | Mean of repetitions in combination of each level $\bar{x}_{\cdot\beta}$. (See Note (a).) Here, $\text{M1} = \prod_{\text{i}=1}^{\text{M}} \text{LT(i)}$ |
| 9 | X1 | $\begin{Bmatrix}\text{D}\\ \text{R}\end{Bmatrix}$ | 1 | Output | Total mean $\bar{x}$ |
| 10 | V | $\begin{Bmatrix}\text{D}\\ \text{R}\end{Bmatrix}$ | $2^{\text{M}}+1, 5$ | Output | Analysis of variance table (For the storage method of the analysis of variance table, see Table 7−5 in note (c).) |
| 11 | WK | $\begin{Bmatrix}\text{D}\\ \text{R}\end{Bmatrix}$ | M3 | Work | Work area. Here, $\text{M3} = \prod_{\text{i}=1}^{\text{N}} (\text{LT(i)}+1)$ |
| 12 | IERR | I | 1 | Output | Error indicator |

## (4) **Restrictions**

(a) $\text{NA} \geq \text{N} \geq \begin{cases} 2 & (\text{M}=1) \\ 1 & (\text{M}>1) \end{cases}$

(b) $1 \leq \text{M} \leq 6$

(c) LT($i$)≥1 $(i = 1, \cdots, \text{M})$

(d) $0 \leq \text{IPN} \leq \begin{cases} 2^{\text{M}}-2 & (\text{N}=1) \\ 2^{\text{M}}-1 & (\text{N}>1) \end{cases}$

(e) $2 \leq \text{IPT}(i) \leq \begin{cases} 2^{\text{M}}-1 & (\text{N}=1) \\ 2^{\text{M}} & (\text{N}>1) \end{cases} \quad (i = 1, \cdots, \text{M})$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | LT($i$) was specified for some $i$ ($i = 1, 2, \cdots, \mathrm{M}$). | The absolute value maximum that can be represented is set for items other than the variation and degrees of freedom items of the analysis of variance table. |
| 3000 | Any of restriction (a) to (e) was not satisfied. | Processing is aborted. |

(6) **Notes**

(a) Multi-way layout data having $m$ factors $A_1, A_2, \cdots, A_m$ for which the number of repetitions in the combination of each factor level is a fixed value $n$ is represented using $m+1$ subscripts as follows.

$$x_{kj_1j_2\cdots j_m} \ (k = 1, \cdots, n; i = 1, \cdots, m; j_i = 1, \cdots, l_i)$$

However, in this subroutine, the multi-way layout data is represented using 2 subscripts as follows, and that data is stored in array A as a real matrix (two-dimensional array type) (See Appendix A).

$$x_{kj_1j_2\cdots j_m} \rightarrow x_{k\beta}$$

Here, $\beta$ is defined as follows.

$$\beta = j_1 + \sum_{i=2}^{m}(j_i \prod_{k=1}^{i-1} l_k)$$

Similarly, the mean of the repetitions in the combination of each level, which is given by

$$\bar{x}_{\cdot j_1j_2\cdots j_m} \ (i = 1, \cdots, m; j_i = 1, \cdots, l_i)$$

can be represented by one subscript as follows and stored in array Y as a one-dimensional vector.

$$\bar{x}_{\cdot j_1j_2\cdots j_m} \rightarrow \bar{x}_{\cdot\beta}$$

Here, $\beta$ is defined as follows.

$$\beta = j_1 + \sum_{i=2}^{m}(j_i \prod_{k=1}^{i-1} l_k)$$

(b) The factors that are subject to an analysis of variance are numbered as follows.

- One-way layout

| Factor number | Factor |
|---|---|
| 1 | Total |
| 2 | $A_1$ |
| 3 | Error |

- Two-way layout

| Factor number | Factor |
|---|---|
| 1 | Total |
| 2 | $A_1$ |
| 3 | $A_2$ |
| 4 | $A_1 \times A_2$ |
| 5 | Error |

- Three-way layout

| Factor number | Factor |
|---|---|
| 1 | Total |
| 2 | $A_1$ |
| 3 | $A_2$ |
| 4 | $A_1 \times A_2$ |
| 5 | $A_3$ |
| 6 | $A_1 \times A_3$ |
| 7 | $A_2 \times A_3$ |
| 8 | $A_1 \times A_2 \times A_3$ |
| 9 | Error |

- Four-way layout

| Factor number | Factor |
|---|---|
| 1 | Total |
| 2 | $A_1$ |
| 3 | $A_2$ |
| 4 | $A_1 \times A_2$ |
| 5 | $A_3$ |
| 6 | $A_1 \times A_3$ |
| 7 | $A_2 \times A_3$ |
| 8 | $A_1 \times A_2 \times A_3$ |
| 9 | $A_4$ |
| 10 | $A_1 \times A_4$ |
| 11 | $A_2 \times A_4$ |
| 12 | $A_1 \times A_2 \times A_4$ |
| 13 | $A_3 \times A_4$ |
| 14 | $A_1 \times A_3 \times A_4$ |
| 15 | $A_2 \times A_3 \times A_4$ |
| 16 | $A_1 \times A_2 \times A_3 \times A_4$ |
| 17 | Error |

- Five-way layout

| Factor number | Factor |
|---|---|
| 1 | Total |
| 2 | $A_1$ |
| 3 | $A_2$ |
| 4 | $A_1 \times A_2$ |
| 5 | $A_3$ |
| 6 | $A_1 \times A_3$ |
| 7 | $A_2 \times A_3$ |
| 8 | $A_1 \times A_2 \times A_3$ |
| 9 | $A_4$ |
| 10 | $A_1 \times A_4$ |
| 11 | $A_2 \times A_4$ |
| 12 | $A_1 \times A_2 \times A_4$ |
| 13 | $A_3 \times A_4$ |
| 14 | $A_1 \times A_3 \times A_4$ |
| 15 | $A_2 \times A_3 \times A_4$ |
| 16 | $A_1 \times A_2 \times A_3 \times A_4$ |
| 17 | $A_5$ |
| 18 | $A_1 \times A_5$ |
| 19 | $A_2 \times A_5$ |
| 20 | $A_1 \times A_2 \times A_5$ |
| 21 | $A_3 \times A_5$ |
| 22 | $A_1 \times A_3 \times A_5$ |
| 23 | $A_2 \times A_3 \times A_5$ |
| 24 | $A_1 \times A_2 \times A_3 \times A_5$ |
| 25 | $A_4 \times A_5$ |
| 26 | $A_1 \times A_4 \times A_5$ |
| 27 | $A_2 \times A_4 \times A_5$ |
| 28 | $A_1 \times A_2 \times A_4 \times A_5$ |
| 29 | $A_3 \times A_4 \times A_5$ |
| 30 | $A_1 \times A_3 \times A_4 \times A_5$ |
| 31 | $A_2 \times A_3 \times A_4 \times A_5$ |
| 32 | $A_1 \times A_2 \times A_3 \times A_4 \times A_5$ |
| 33 | Error |

- Six-way layout

| Factor number | Factor | Factor number | Factor |
|---|---|---|---|
| 1 | Total | 34 | $A_1 \times A_6$ |
| 2 | $A_1$ | 35 | $A_2 \times A_6$ |
| 3 | $A_2$ | 36 | $A_1 \times A_2 \times A_6$ |
| 4 | $A_1 \times A_2$ | 37 | $A_3 \times A_6$ |
| 5 | $A_3$ | 38 | $A_1 \times A_3 \times A_6$ |
| 6 | $A_1 \times A_3$ | 39 | $A_2 \times A_3 \times A_6$ |
| 7 | $A_2 \times A_3$ | 40 | $A_1 \times A_2 \times A_3 \times A_6$ |
| 8 | $A_1 \times A_2 \times A_3$ | 41 | $A_4 \times A_6$ |
| 9 | $A_4$ | 42 | $A_1 \times A_4 \times A_6$ |
| 10 | $A_1 \times A_4$ | 43 | $A_2 \times A_4 \times A_6$ |
| 11 | $A_2 \times A_4$ | 44 | $A_1 \times A_2 \times A_4 \times A_6$ |
| 12 | $A_1 \times A_2 \times A_4$ | 45 | $A_3 \times A_4 \times A_6$ |
| 13 | $A_3 \times A_4$ | 46 | $A_1 \times A_3 \times A_4 \times A_6$ |
| 14 | $A_1 \times A_3 \times A_4$ | 47 | $A_2 \times A_3 \times A_4 \times A_6$ |
| 15 | $A_2 \times A_3 \times A_4$ | 48 | $A_1 \times A_2 \times A_3 \times A_4 \times A_6$ |
| 16 | $A_1 \times A_2 \times A_3 \times A_4$ | 49 | $A_5 \times A_6$ |
| 17 | $A_5$ | 50 | $A_1 \times A_5 \times A_6$ |
| 18 | $A_1 \times A_5$ | 51 | $A_2 \times A_5 \times A_6$ |
| 19 | $A_2 \times A_5$ | 52 | $A_1 \times A_2 \times A_5 \times A_6$ |
| 20 | $A_1 \times A_2 \times A_5$ | 53 | $A_3 \times A_5 \times A_6$ |
| 21 | $A_3 \times A_5$ | 54 | $A_1 \times A_3 \times A_5 \times A_6$ |
| 22 | $A_1 \times A_3 \times A_5$ | 55 | $A_2 \times A_3 \times A_5 \times A_6$ |
| 23 | $A_2 \times A_3 \times A_5$ | 56 | $A_1 \times A_2 \times A_3 \times A_5 \times A_6$ |
| 24 | $A_1 \times A_2 \times A_3 \times A_5$ | 57 | $A_4 \times A_5 \times A_6$ |
| 25 | $A_4 \times A_5$ | 58 | $A_1 \times A_4 \times A_5 \times A_6$ |
| 26 | $A_1 \times A_4 \times A_5$ | 59 | $A_2 \times A_4 \times A_5 \times A_6$ |
| 27 | $A_2 \times A_4 \times A_5$ | 60 | $A_1 \times A_2 \times A_4 \times A_5 \times A_6$ |
| 28 | $A_1 \times A_2 \times A_4 \times A_5$ | 61 | $A_3 \times A_4 \times A_5 \times A_6$ |
| 29 | $A_3 \times A_4 \times A_5$ | 62 | $A_1 \times A_3 \times A_4 \times A_5 \times A_6$ |
| 30 | $A_1 \times A_3 \times A_4 \times A_5$ | 63 | $A_2 \times A_3 \times A_4 \times A_5 \times A_6$ |
| 31 | $A_2 \times A_3 \times A_4 \times A_5$ | 64 | $A_1 \times A_2 \times A_3 \times A_4 \times A_5 \times A_6$ |
| 32 | $A_1 \times A_2 \times A_3 \times A_4 \times A_5$ | 65 | Error |
| 33 | $A_6$ | | |

(c) The analysis of variance table elements are stored as follows in the array V.

Table 7−5 Analysis of Variance Table Storage Status

| Factor number | Variation | Degrees of freedom | Unbiased variance | Variance ratio | Contribution ratio |
|---|---|---|---|---|---|
| 1 | V(1, 1) | V(1, 2) | $*$ | $*$ | $*$ |
| 2 | V(2, 1) | V(2, 2) | V(2, 3) | V(2, 4) | V(2, 5) |
| 3 | V(3, 1) | V(3, 2) | V(3, 3) | V(3, 4) | V(3, 5) |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| N1 − 1 | V(N1 − 1, 1) | V(N1 − 1, 2) | V(N1 − 1, 3) | V(N1 − 1, 4) | V(N1 − 1, 5) |
| N1 | V(N1, 1) | V(N1, 2) | V(N1, 3) | $*$ | V(N1, 5) |

Here, $\mathrm{N1} = 2^{\mathrm{M}} + 1$. The absolute value maximum that can be represented is set for array elements V(1, 3), V(1, 4), V(1, 5), and V(N1, 4), which are elements for which there is no corresponding analysis of variance table item. Also, when there are no repetitions, 0.0 is set for the value of each item for the highest order interaction factor (factor number N1 − 1). In addition, for the error, 0.0 is set for the value of each item for the factors specified in array IPT, which are to be confounded.

(7) **Example**

(a) Problem

Perform an analysis of variance for three-way layout data for which the number of levels for each factor is 3 and the number of repetitions is 2. Assume that the transpose matrix $X^T$ of the observation matrix $X$ is given as follows.

$$\begin{bmatrix} 5.5 & 5.4 \\ 6.3 & 6.5 \\ 6.9 & 6.8 \\ 5.4 & 5.3 \\ 6.5 & 6.3 \\ 6.9 & 6.5 \\ 5.5 & 5.3 \\ 6.5 & 6.2 \\ 7.0 & 6.8 \\ 4.9 & 4.6 \\ 5.7 & 5.6 \\ 6.2 & 6.0 \\ 5.0 & 5.2 \\ 6.1 & 6.5 \\ 6.6 & 6.5 \\ 4.8 & 4.2 \\ 5.7 & 5.4 \\ 6.9 & 6.4 \\ 4.2 & 4.0 \\ 5.0 & 4.9 \\ 5.4 & 5.3 \\ 4.5 & 4.3 \\ 5.2 & 5.0 \\ 6.1 & 6.0 \\ 4.9 & 4.3 \\ 5.3 & 5.2 \\ 6.4 & 6.3 \end{bmatrix}$$

(b) Input data

Observation matrix $X$, NA=2, N=2, M=3, IPN=0, LT(1)=3, LT(2)=3 and LT(3)=3.

(c) Main program

```
      PROGRAM B4MWRF
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 2, M = 3, NIPT = 7 )
      PARAMETER( M1 = 27, M2 = 64, M3 = 9 )
      DIMENSION A(NA,M1),LT(M),IPT(NIPT),Y(M1),V(M3,5),WK(M2)
!
      WRITE(6,6000)
      IERR  = 0
      READ(5,*) N
      LT(1) = 3
      LT(2) = 3
      LT(3) = 3
      READ(5,*) IPN
      DO 100 I=1,N
         READ(5,*) (A(I,J),J=1,M1)
  100 CONTINUE
      WRITE(6,6010) N,M
      WRITE(6,6020) (LT(I),I=1,3)
      WRITE(6,6030) IPN
      DO 110 I=1,N
         WRITE(6,6040) I
         WRITE(6,6050) (A(I,J),J=1,M1)
  110 CONTINUE
      CALL D4MWRF(A,NA,N,LT,M,IPT,IPN,Y,X1,V,WK,IERR)
      WRITE(6,6060) IERR
      WRITE(6,6070)
      WRITE(6,6050) (Y(I),I=1,M1)
      WRITE(6,6080) X1
      WRITE(6,6090) (V(1,J),J=1,2)
      DO 120 I = 2,M3-1
         WRITE(6,6100) I,(V(I,J),J=1,5)
```

```
  120 CONTINUE
      WRITE(6,6110) (V(M3,J),J=1,3),V(M3,5)
!
      STOP
 6000 FORMAT(  ' *** D4MWRF ***',/,&
              /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'N = ',I6,5X,'M = ',I6,/,&
              /,7X,'NUMBER OF LEVEL OF EACH FACTOR',/)
 6020 FORMAT(  7X,5(2X,I6))
 6030 FORMAT( /,7X,'NUMBER OF POOLING SOURCE OF VARIANCES = ',I6,/,&
              /,7X,'OBSERVATION MATRIX')
 6040 FORMAT( /,9X,'A(',I1,',J),J=1,27',/)
 6050 FORMAT(      6(8X,5(1X,F11.2),/))
 6060 FORMAT( /,3X,'** OUTPUT **',/,&
              /,7X,'IERR = ',I6)
 6070 FORMAT( /,7X,'MEAN FOR REPETITION',/)
 6080 FORMAT( /,7X,'MEAN OVER ALL LEVELS = ',F11.2,/,&
              /,7X,'ANALYSIS OF VARIANCE TABLE',/,&
              /,10X,'FACTOR',&
                   7X,'S.S.',8X,'D.F.',5X,'M.S.',8X,'V.R.',8X,'C.R.',/,&
               9X,67('-'))
 6090 FORMAT(  10X,'TOTAL',2(1X,F11.2),3(1X,D11.4))
 6100 FORMAT(  7X,I6,2X,2(1X,F11.2),3(1X,D11.4))
 6110 FORMAT(  10X,'ERROR',2(1X,F11.2),1X,D11.4,13X,D11.4)
      END
```

(d) Output results

```
*** D4MWRF ***

  ** INPUT **

      N =      2     M =      3

      NUMBER OF LEVEL OF EACH FACTOR

             3       3       3

      NUMBER OF POOLING SOURCE OF VARIANCES =      0

      OBSERVATION MATRIX

        A(1,J),J=1,27

               5.50        6.30        6.90        5.40        6.50
               6.90        5.50        6.50        7.00        4.90
               5.70        6.20        5.00        6.10        6.60
               4.80        5.70        6.90        4.20        5.00
               5.40        4.50        5.20        6.10        4.90
               5.30        6.40

        A(2,J),J=1,27

               5.40        6.50        6.80        5.30        6.30
               6.50        5.30        6.20        6.80        4.60
               5.60        6.00        5.20        6.50        6.50
               4.20        5.40        6.40        4.00        4.90
               5.30        4.30        5.00        6.00        4.30
               5.20        6.30

  ** OUTPUT **

      IERR =      0

      MEAN FOR REPETITION

               5.45        6.40        6.85        5.35        6.40
               6.70        5.40        6.35        6.90        4.75
               5.65        6.10        5.10        6.30        6.55
               4.50        5.55        6.65        4.10        4.95
               5.35        4.40        5.10        6.05        4.60
               5.25        6.35

      MEAN OVER ALL LEVELS =        5.67

      ANALYSIS OF VARIANCE TABLE

         FACTOR       S.S.        D.F.     M.S.        V.R.        C.R.
        -------------------------------------------------------------------
         TOTAL       36.09       53.00
            2        21.59        2.00  0.1080D+02  0.2886D+03  0.5962D+00
            3         0.70        2.00  0.3513D+00  0.9391D+01  0.1739D-01
            4         0.51        4.00  0.1282D+00  0.3428D+01  0.1007D-01
            5        10.35        2.00  0.5176D+01  0.1384D+03  0.2847D+00
            6         0.31        4.00  0.7852D-01  0.2099D+01  0.4556D-02
            7         1.25        4.00  0.3130D+00  0.8366D+01  0.3054D-01
            8         0.36        8.00  0.4449D-01  0.1189D+01  0.1570D-02
         ERROR        1.01       27.00  0.3741D-01              0.5493D-01
```

### 7.4.2 D4MWRM, R4MWRM Multiple-Way Layout Analysis of Variance (With Missing Values)

(1) **Function**

For $m$ factors $A_1, A_2, \cdots, A_m$ (where m is at most 6) consisting of $l_1, l_2, \cdots, l_m$ levels respectively and multi-way layout data $x_{j_1 j_2 \cdots j_m} (i = 1, \cdots, m; j_i = 1, \cdots, l_i)$ for which there are no repetitions in the combinations of levels of each factor and no data has been obtained for combinations of $n_s$ combinations of levels, the D4MWRM or R4MWRM performs an analysis of variance. At this time, specified interactions can be confounded. The data values of the combinations of levels for which no data has been obtained are called missing values, and estimates are substituted for the missing values in the calculations of each statistic.

For the explanations below, the operations $\Sigma_i$ and $\Delta_i$ for factor $\mathrm{A}_i$ are defined as follows.

$$\Sigma_i \equiv \sum_{j_i=1}^{l_i}, \qquad \Delta_i \equiv l_i - \sum_{j_i=1}^{l_i}$$

The total mean over all levels is defined by the following equation.

Total mean:

$$\bar{x} = \frac{1}{\displaystyle\prod_{i=1}^{m} l_i} \Sigma_1 \Sigma_2 \cdots \Sigma_m \sum_{k=1}^{n} x_{k j_1 j_2 \cdots j_m}$$

Also, the analysis of variance results are defined by the following equations.

Variation:

- Total variation

$$S_T = \Sigma_1 \Sigma_2 \cdots \Sigma_m \sum_{k=1}^{n} (x_{k j_1 j_2 \cdots j_m} - \bar{x})^2$$

- Variation of factor $A_i$

$$S_{A_i} = \frac{1}{l_i \displaystyle\prod_{j=1}^{m} l_j} \Sigma_i (\Sigma_1 \Sigma_2 \cdots \Sigma_{i-1} \Delta_i \Sigma_{i+1} \cdots \Sigma_m x_{j_1 j_2 \cdots j_m})^2$$

- Variation of the interaction $A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}$ of $s(s < m)$ factors $A_{i_1}, A_{i_2}, \cdots, A_{i_s}$.

$$\begin{aligned} S_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}} &= \frac{n}{(\displaystyle\prod_{k=1}^{s} l_{i_s}) \cdot (\displaystyle\prod_{k=1}^{m} l_k)} \\ &\times \Sigma_{i_1} \Sigma_{i_2} \cdots \Sigma_{i_s} (\Sigma_1 \cdots \Sigma_{i_1 - 1} \Delta_{i_1} \Sigma_{i_1+1} \cdots \\ &\quad \Sigma_{i_2-1} \Delta_{i_2} \Sigma_{i_2+1} \cdots \Sigma_{i_s-1} \Delta_{i_s} \Sigma_{i_s+1} \cdots \Sigma_m x_{j_1 j_2 \cdots j_m})^2 \end{aligned}$$

  However, the variation of the highest order interaction $A_1 \times A_2 \times \cdots \times A_m$.

- Error variation

$$\begin{aligned} S_E &= S_T - (\text{Sum of variations of each factor}) \\ &- (\text{Sum of variations of interactions other than the highest order interaction} \\ &\quad [A_1 \times A_2 \times \cdots \times A_m]) \end{aligned}$$

Degrees of freedom:

- Degrees of freedom of total variation

$$\phi_T = n \prod_{i=1}^{m} l_i - n_s - 1$$

- Degrees of freedom of factor $A_i$ variation

$$\phi_A = l_i - 1$$

- Degrees of freedom of variation of the interaction $A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}$

$$\phi_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}} = \prod_{k=1}^{s} (l_{i_k} - 1)$$

- Degrees of freedom of error variation

$$\begin{aligned}\phi_E &= \phi_T - \text{(Sum of degrees of freedom of variations of each factor)} \\ &- \text{(Sum of degrees of freedom of variations of interactions other than the highest} \\ &\quad \text{order interaction -}n_s)\end{aligned}$$

Unbiased variance:

- Unbiased variance of factor $A_i$ variation

$$V_{A_i} = \frac{S_{A_i}}{\phi_{A_i}}$$

- Unbiased variance of variation of the interaction $A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}$

$$V_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}} = \frac{S_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}}}{\phi_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}}}$$

  However, the unbiased variance of the variation of the highest order interaction is not defined.

- Unbiased variance of error variation

$$V_E = \frac{S_E}{\phi_E}$$

Variance ratio:

- Variance ratio for unbiased variance of factor $A_i$ variation

$$F_{A_i} = \frac{V_{A_i}}{V_E}$$

- Variance ratio for unbiased variance of variation of the interaction $A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}$

$$F_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}} = \frac{V_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}}}{V_E}$$

  However, the variance ratio of the variation of the highest order interaction is not defined.

Contribution ratio:

- Contribution ratio of factor $A_i$ variation

$$P_{A_i} = \frac{S_{A_i} - \phi_{A_i} \cdot V_E}{S_T}$$

- Contribution ratio of variation of the interaction $A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}$

$$P_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}} = \frac{S_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}} - \phi_{A_{i_1} \times A_{i_2} \times \cdots \times A_{i_s}} \cdot V_E}{S_T}$$

However, the contribution ratio of the variation of the highest order interaction is not defined.

- Contribution ratio of error variation

$$\begin{aligned} P_E &= \phi_T - (\text{Sum of contribution ratios of variations of each factor}) \\ &- (\text{Sum of contribution ratios of variations of interactions other than the highest} \\ &\quad \text{order interaction}) \end{aligned}$$

Missing value estimates:
The missing value estimates are determined so that the error variation $S_E$ is minimized. To obtain estimates that will minimize $S_E$, you should solve the following equation with the missing values

$$x_{s_1 s_2 \ldots s_m} \ ((s_1, s_2, \ldots, s_m) \in S)$$

as unknowns.

$$\frac{\partial S_E}{\partial x_{st}} = 0 \qquad ((s,t) \in S)$$

Here, $S$ is assumed to be the set of combinations of levels for which missing values occurred. Since $S_E$ is a quadratic equation, this equation is a set of $n_s$ simultaneous linear equations with the missing values as unknowns.
For example, for the following three-way layout data in which there are three levels of each of the factors A, B, and C and $x_{132}$ and $x_{322}$ are missing values,

$$X_1 = \begin{bmatrix} 11 & 12 & 10 \\ 11 & 10 & 12 \\ 12 & x_{132} & 13 \end{bmatrix} \qquad \text{Factor A level 1 data}$$

$$X_2 = \begin{bmatrix} 11 & 11 & 12 \\ 13 & 14 & 11 \\ 9 & 10 & 12 \end{bmatrix} \qquad \text{Factor A level 2 data}$$

$$X_3 = \begin{bmatrix} 10 & 12 & 11 \\ 10 & x_{322} & 12 \\ 12 & 8 & 11 \end{bmatrix} \qquad \text{Factor A level 3 data}$$

the error variation is as follows.

$$\begin{aligned} S_E &= S_T - S_A - S_B - S_C - S_{A\times B} - S_{B\times C} - S_{C\times A} \\ &= 69714 - 5400x_{132} - 5724x_{322} + 216x_{132}^2 + 216x_{322} + 108x_{132}x_{322} \end{aligned}$$

Differentiating this with respect to $x_{132}$ and $x_{322}$ produces the following simultaneous linear equations.

$$\frac{\partial S_E}{\partial x_{132}} = -5400 + 432x_{132} + 108x_{322} = 0$$

$$\frac{\partial S_E}{\partial x_{322}} = -5724 + 108x_{132} + 432x_{322} = 0$$

Solving these simultaneous linear equations yields the missing value estimates $x_{132} = 9.8$ and $x_{322} = 10.8$.

(2) **Usage**

Double precision:

CALL D4MWRM (A, LT, M, IST, ISN, IPT, IPN, X1, V, IWK, WK, IERR)

Single precision:

CALL R4MWRM (A, LT, M, IST, ISN, IPT, IPN, X1, V, IWK, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: {INTEGER(4) as for 32bit Integer, INTEGER(8) as for 64bit Integer}
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/Output | Contents |
|---|---|---|---|---|---|
| 1 | A | {D, R} | M1 | Input | Observed values ($x_\beta$) . (See Note (a).) Here $\mathrm{M1} = \prod_{\mathrm{i=1}}^{\mathrm{M}} \mathrm{LT(i)}$ |
| | | | | Output | Observed values ($x_\beta$). However, estimates are stored in the elements corresponding to missing values. |
| 2 | LT | I | M | Input | Number of levels of each factor $l_i$ |
| 3 | M | I | 1 | Input | Number of factors $m$ |
| 4 | IST | I | ISN,M | Input | Information about combinations of levels for which missing values occurred (See Note (b)) |
| 5 | ISN | I | 1 | Input | Number of missing values $n_s$ |
| 6 | IPT | I | M2 | Input | Number of factors to be confounded (See Note (c)) However, when IPN $>$ 0, M2 = IPN, and when IPN = 0, the argument IPT should be a dummy argument. |
| 7 | IPN | I | 1 | Input | Number of factors to be confounded |
| 8 | X1 | {D, R} | 1 | Output | Total mean $\bar{x}$ |
| 9 | V | {D, R} | $(2^{\mathrm{M}}+1), 5$ | Output | Analysis of variance table (For the storage method of the analysis of variance table, see Table 7−6 in Note (c).) |
| 10 | IWK | I | M3 | Work | Work area. $\mathrm{M3} = \prod_{\mathrm{i=1}}^{\mathrm{M}} \mathrm{LT(i)} + \mathrm{ISN}$ |
| 11 | WK | {D, R} | M4 | Work | Work area. $\mathrm{M4} = \prod_{\mathrm{i=1}}^{\mathrm{M}} (\mathrm{LT(i)}+1) + \mathrm{ISN}^2 + 2 \times \mathrm{ISN} + 1$ |
| 12 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $1 \leq \mathrm{M} \leq 6$

(b) $\mathrm{LT}(i) \geq 1 \quad (i = 1, \cdots, \mathrm{M})$

(c) $0 \leq \mathrm{IPN} \leq 2^{\mathrm{M}} - 2$

(d) $2 \leq \mathrm{IPT}(i) \leq 2^{\mathrm{M}} - 1 \qquad (i = 1, \cdots, \mathrm{M})$

(e) $1 \leq \mathrm{ISN} < \prod_{i=1}^{\mathrm{M}}(\mathrm{LT(i)} - 1)$

(f) $1 \leq \mathrm{IST}(i, j) \leq \mathrm{LT}(j) \quad (i = 1, 2, \cdots, \mathrm{ISN}\ ;\ j = 1, 2, \cdots, \mathrm{M})$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 2000 | Estimates of missing values could not be calculated. | The missing value data is replaced by the mean of the non-missing value data, and processing continues. |
| 3000 | Any of restrictions (a) to (f) was not satisfied. | Processing is aborted. |

(6) **Notes**

(a) Multi-way layout data with no repetitions and having $m$ factors $A_1, A_2, \cdots, A_m$ is represented using $m$ subscripts as follows.

$x_{j_1 j_2 \cdots j_m}$ $(i = 1, \cdots, m; j_i = 1, \cdots, l_i)$

However, in this subroutine, the multi-way layout data is represented using one subscripts as follows, and that data is stored in array A as a real vector (one-dimensional array).

$x_{j_1 j_2 \cdots j_m} \rightarrow x_\beta$

Here, $\beta$ is defined as follows.

$$\beta = j_1 + \sum_{i=2}^{m}(j_i \prod_{k=1}^{i-1} l_k)$$

(b) The combinations of levels of the $i$-th missing value $(j_1, j_2, \cdots, j_m)$ are represented as a single integer $\beta$ according to the method described in Note (a), and that value is stored in $\mathrm{IST}(i)$).

(c) The factors that are subject to an analysis of variance are numbered as follows.

- One-way layout

| Factor number | Factor |
|---|---|
| 1 | Total |
| 2 | Error |

- Two-way layout

| Factor number | Factor |
|---|---|
| 1 | Total |
| 2 | $A_1$ |
| 3 | $A_2$ |
| 4 | Error |

- Three-way layout

| Factor number | Factor |
|---|---|
| 1 | Total |
| 2 | $A_1$ |
| 3 | $A_2$ |
| 4 | $A_1 \times A_2$ |
| 5 | $A_3$ |
| 6 | $A_1 \times A_3$ |
| 7 | $A_2 \times A_3$ |
| 8 | Error |

- Four-way layout

| Factor number | Factor |
|---|---|
| 1 | Total |
| 2 | $A_1$ |
| 3 | $A_2$ |
| 4 | $A_1 \times A_2$ |
| 5 | $A_3$ |
| 6 | $A_1 \times A_3$ |
| 7 | $A_2 \times A_3$ |
| 8 | $A_1 \times A_2 \times A_3$ |
| 9 | $A_4$ |
| 10 | $A_1 \times A_4$ |
| 11 | $A_2 \times A_4$ |
| 12 | $A_1 \times A_2 \times A_4$ |
| 13 | $A_3 \times A_4$ |
| 14 | $A_1 \times A_3 \times A_4$ |
| 15 | $A_2 \times A_3 \times A_4$ |
| 16 | Error |

- Five-way layout

| Factor number | Factor |
|---|---|
| 1 | Total |
| 2 | $A_1$ |
| 3 | $A_2$ |
| 4 | $A_1 \times A_2$ |
| 5 | $A_3$ |
| 6 | $A_1 \times A_3$ |
| 7 | $A_2 \times A_3$ |
| 8 | $A_1 \times A_2 \times A_3$ |
| 9 | $A_4$ |
| 10 | $A_1 \times A_4$ |
| 11 | $A_2 \times A_4$ |
| 12 | $A_1 \times A_2 \times A_4$ |
| 13 | $A_3 \times A_4$ |
| 14 | $A_1 \times A_3 \times A_4$ |
| 15 | $A_2 \times A_3 \times A_4$ |
| 16 | $A_1 \times A_2 \times A_3 \times A_4$ |
| 17 | $A_5$ |
| 18 | $A_1 \times A_5$ |
| 19 | $A_2 \times A_5$ |
| 20 | $A_1 \times A_2 \times A_5$ |
| 21 | $A_3 \times A_5$ |
| 22 | $A_1 \times A_3 \times A_5$ |
| 23 | $A_2 \times A_3 \times A_5$ |
| 24 | $A_1 \times A_2 \times A_3 \times A_5$ |
| 25 | $A_4 \times A_5$ |
| 26 | $A_1 \times A_4 \times A_5$ |
| 27 | $A_2 \times A_4 \times A_5$ |
| 28 | $A_1 \times A_2 \times A_4 \times A_5$ |
| 29 | $A_3 \times A_4 \times A_5$ |
| 30 | $A_1 \times A_3 \times A_4 \times A_5$ |
| 31 | $A_2 \times A_3 \times A_4 \times A_5$ |
| 32 | Error |

- Six-way layout

| Factor number | Factor | Factor number | Factor |
|---|---|---|---|
| 1 | Total | 33 | $A_1 \times A_6$ |
| 2 | $A_1$ | 34 | $A_1 \times A_6$ |
| 3 | $A_2$ | 35 | $A_2 \times A_6$ |
| 4 | $A_1 \times A_2$ | 36 | $A_1 \times A_2 \times A_6$ |
| 5 | $A_3$ | 37 | $A_3 \times A_6$ |
| 6 | $A_1 \times A_3$ | 38 | $A_1 \times A_3 \times A_6$ |
| 7 | $A_2 \times A_3$ | 39 | $A_2 \times A_3 \times A_6$ |
| 8 | $A_1 \times A_2 \times A_3$ | 40 | $A_1 \times A_2 \times A_3 \times A_6$ |
| 9 | $A_4$ | 41 | $A_4 \times A_6$ |
| 10 | $A_1 \times A_4$ | 42 | $A_1 \times A_4 \times A_6$ |
| 11 | $A_2 \times A_4$ | 43 | $A_2 \times A_4 \times A_6$ |
| 12 | $A_1 \times A_2 \times A_4$ | 44 | $A_1 \times A_2 \times A_4 \times A_6$ |
| 13 | $A_3 \times A_4$ | 45 | $A_3 \times A_4 \times A_6$ |
| 14 | $A_1 \times A_3 \times A_4$ | 46 | $A_1 \times A_3 \times A_4 \times A_6$ |
| 15 | $A_2 \times A_3 \times A_4$ | 47 | $A_2 \times A_3 \times A_4 \times A_6$ |
| 16 | $A_1 \times A_2 \times A_3 \times A_4$ | 48 | $A_1 \times A_2 \times A_3 A_4 \times A_6$ |
| 17 | $A_5$ | 49 | $A_5 \times A_6$ |
| 18 | $A_1 \times A_5$ | 50 | $A_1 \times A_5 \times A_6$ |
| 19 | $A_2 \times A_5$ | 51 | $A_2 \times A_5 \times A_6$ |
| 20 | $A_1 \times A_2 \times A_5$ | 52 | $A_1 \times A_2 \times A_5 \times A_6$ |
| 21 | $A_3 \times A_5$ | 53 | $A_3 \times A_5 \times A_6$ |
| 22 | $A_1 \times A_3 \times A_5$ | 54 | $A_1 \times A_3 \times A_5 \times A_6$ |
| 23 | $A_2 \times A_3 \times A_5$ | 55 | $A_2 \times A_3 \times A_5 \times A_6$ |
| 24 | $A_1 \times A_2 \times A_3 \times A_5$ | 56 | $A_1 \times A_2 \times A_3 \times A_5 \times A_6$ |
| 25 | $A_4 \times A_5$ | 57 | $A_4 \times A_5 \times A_6$ |
| 26 | $A_1 \times A_4 \times A_5$ | 58 | $A_1 \times A_4 \times A_5 \times A_6$ |
| 27 | $A_2 \times A_4 \times A_5$ | 59 | $A_2 \times A_4 \times A_5 \times A_6$ |
| 28 | $A_1 \times A_2 \times A_4 \times A_5$ | 60 | $A_1 \times A_2 \times A_4 \times A_5 \times A_6$ |
| 29 | $A_3 \times A_4 \times A_5$ | 61 | $A_3 \times A_4 \times A_5 \times A_6$ |
| 30 | $A_1 \times A_3 \times A_4 \times A_5$ | 62 | $A_1 \times A_3 \times A_4 \times A_5 \times A_6$ |
| 31 | $A_2 \times A_3 \times A_4 \times A_5$ | 63 | $A_2 \times A_3 \times A_4 \times A_5 \times A_6$ |
| 32 | $A_1 \times A_2 \times A_3 \times A_4 \times A_5$ | 64 | Error |

(d) The analysis of variance table elements are stored as follows in the array V.

Table 7−6 Analysis of Variance Table Storage Status

| Factor number | Variation | Degrees of freedom | Unbiased variance | Variance ratio | Contribution ratio |
|---|---|---|---|---|---|
| 1 | V(1, 1) | V(1, 2) | $*$ | $*$ | $*$ |
| 2 | V(2, 1) | V(2, 2) | V(2, 3) | V(2, 4) | V(2, 5) |
| 3 | V(3, 1) | V(3, 2) | V(3, 3) | V(3, 4) | V(3, 5) |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| N1 − 1 | V(N1 − 1, 1) | V(N1 − 1, 2) | V(N1 − 1, 3) | V(N1 − 1, 4) | V(N1 − 1, 5) |
| N1 | V(N1, 1) | V(N1, 2) | V(N1, 3) | $*$ | V(N1, 5) |

Here, N1 = $2^{\mathrm{M}}$. The absolute value maximum that can be represented is set for array elements V(1, 3), V(1, 4), V(1, 5) and V(N1, 4), which are elements for which there is no corresponding analysis of variance table item. In addition, for the error, 0.0 is set for the value of each item for the factors specified in array IPT, which are to be confounded.

(7) **Example**

(a) Problem

Perform an analysis of variance for three-way layout data having no repetitions and for which the number of levels for each factor is 3. However, the observation value vector $X = x_\beta$ is given as follows, where the $*$ portions are missing values.

$$\begin{bmatrix} 13 \\ 21 \\ * \\ 18 \\ 29 \\ 40 \\ 23 \\ 37 \\ 51 \\ 21 \\ * \\ 47 \\ 27 \\ 44 \\ 61 \\ * \\ 54 \\ 75 \\ 29 \\ 47 \\ 65 \\ 36 \\ 59 \\ 82 \\ 43 \\ 71 \\ 99 \end{bmatrix}$$

(b) Input data

Observation matrix $X$, NA=2, N=2, M=3, IPN=0, ISN=3, LT(1)=3, LT(2)=3, LT(3)=3, IST(1,1)=1, IST(1,2)=3, IST(1,3)=2, IST(2,1)=2, IST(2,2)=1, IST(2,3)=2, IST(3,1)=3, IST(3,2)=1 and IST(3,3)=1.

(c) Main program

```
      PROGRAM B4MWRM
!
      IMPLICIT NONE
      REAL(8) V(8,5),WK(80),A(27),X1
      INTEGER NE,IWK(30),IST(3,3),ISN,IPN,LT(7),IPT(15)
      INTEGER IERR,M,M1,I,J,IWK1
!
!     EXAMPLE OF D/R4MWRM
!
      WRITE(6,1000) '*** DFWTFF ***'
      WRITE(6,1000) ' ** INPUT ** '
      READ(5,*) M
      WRITE(6,6020) M
      M1 = 1
      DO 100 I=1,M
         READ(5,*) LT(I)
         WRITE(6,6030) I,LT(I)
         M1 = M1 * LT(I)
 100  CONTINUE
      READ(5,*) IPN
      WRITE(6,6040) IPN
      READ(5,*) ISN
      WRITE(6,6045) ISN
      DO 110 I=1,ISN
         DO 120 J=1,M
            READ(5,*) IST(I,J)
            WRITE(6,6300) I,J,IST(I,J)
 120     CONTINUE
 110  CONTINUE
      DO 130 I=1,M1
```

```
            READ(5,*) A(I)
            WRITE(6,5360) I,A(I)
  130    CONTINUE
!
         CALL D4MWRM(A,LT,M,IST,ISN,IPT,IPN,X1,V,IWK,WK,IERR)
         WRITE(6,6050) IERR
         WRITE(6,5350)
         DO 170 I=1,ISN
            IWK1 = IST(I,1)+(IST(I,2)-1)*LT(1)+(IST(I,3)-1)*LT(1)*LT(2)
            WRITE(6,5360) IWK1,A(IWK1)
  170    CONTINUE
         WRITE(6,6060)
         WRITE(6,6070) X1
         WRITE(6,6110)
         WRITE(6,6120)
         NE = 2**M
         WRITE(6,6130) (V(1,J),J=1,2)
         DO 220 I = 2,NE-1
            WRITE(6,6150) I,(V(I,J),J=1,5)
   220 CONTINUE
         WRITE(6,6170) (V(NE,J),J=1,3),V(NE,5)
  1000 FORMAT( /,3X,A16)
  5200 FORMAT( /,/,6X,'MISSED VALUES',/)
  5250 FORMAT( /,8X,'A(',I1,')',/)
  5350 FORMAT( /,/,' ESTIMATED MISSED VALUES',/)
  5360 FORMAT( /,7X,'A(',I2,') = ',D11.5,/)
  6020 FORMAT( /,7X,'M = ',I6)
  6040 FORMAT( /,7X,'IPN = ',I6)
  6045 FORMAT( /,7X,'ISN = ',I6)
  6030 FORMAT( /,7X,'LT(',I1,') = ',I6)
  6200 FORMAT( /,7X,'IPT(',I1,') = ',I6)
  6300 FORMAT( /,7X,'IST(',I1,',',I1,') = ',I6)
  6050 FORMAT( /,' ** OUTPUT **',/,&
            /,7X,'IERR = ',I6)
  6060 FORMAT( /,/,' MEAN OVER ALL LEVELS',/,/)
  6090 FORMAT( /,/,' MEAN FOR REPETITION',/,/)
  6070 FORMAT(  '    ',2X,'      ',D11.5,' ')
  6100 FORMAT(  ' Y(',I2,') =  ',D11.5,' ')
  6080 FORMAT(  '  STRICT (',D11.5,')',/)
  6110 FORMAT( /,/,' ANALYSIS-OF-VARIANCE TABLE',/,/)
  6120 FORMAT(  ' ',&
         ' FACTOR  ',4X,'S.S.',9X,'D.F.',9X,'U.V.',9X,'R.V.',9X,'R.C',/,/)
  6130 FORMAT(  ' ',&
         ' TOTAL  ',2(2X,D11.4))
  6150 FORMAT(  ' ',&
         2X,I3,3X,5(2X,D11.4))
  6170 FORMAT(  '  ERROR  ',3(2X,D11.4),15X,D11.4)
         STOP
         END
```

(d) Output results

```
 *** DFWTFF ***

   ** INPUT **

   M =      3

   LT(1) =      3

   LT(2) =      3

   LT(3) =      3

   IPN =      0

   ISN =      3

   IST(1,1) =      1

   IST(1,2) =      3

   IST(1,3) =      2

   IST(2,1) =      2

   IST(2,2) =      1

   IST(2,3) =      2

   IST(3,1) =      3

   IST(3,2) =      1

   IST(3,3) =      1

   A( 1) = 0.13000D+02


   A( 2) = 0.21000D+02


   A( 3) = 0.10000D+01
```

```
      A( 4) = 0.18000D+02

      A( 5) = 0.29000D+02

      A( 6) = 0.40000D+02

      A( 7) = 0.23000D+02

      A( 8) = 0.37000D+02

      A( 9) = 0.51000D+02

      A(10) = 0.21000D+02

      A(11) = 0.10000D+01

      A(12) = 0.47000D+02

      A(13) = 0.27000D+02

      A(14) = 0.44000D+02

      A(15) = 0.61000D+02

      A(16) = 0.10000D+01

      A(17) = 0.54000D+02

      A(18) = 0.75000D+02

      A(19) = 0.29000D+02

      A(20) = 0.47000D+02

      A(21) = 0.65000D+02

      A(22) = 0.36000D+02

      A(23) = 0.59000D+02

      A(24) = 0.82000D+02

      A(25) = 0.43000D+02

      A(26) = 0.71000D+02

      A(27) = 0.99000D+02

** OUTPUT **
      IERR =      0

ESTIMATED MISSED VALUES

      A(16) = 0.32250D+02

      A(11) = 0.35125D+02

      A( 3) = 0.25250D+02


MEAN OVER ALL LEVELS

         0.43875D+02
```

ANALYSIS-OF-VARIANCE TABLE

| FACTOR | S.S. | D.F. | U.V. | R.V. | R.C |
|---|---|---|---|---|---|
| TOTAL | 0.1185D+05 | 0.2300D+02 | | | |
| 2 | 0.5101D+04 | 0.2000D+01 | 0.2551D+04 | 0.3001D+04 | 0.4302D+00 |
| 3 | 0.1838D+04 | 0.2000D+01 | 0.9190D+03 | 0.1081D+04 | 0.1549D+00 |
| 4 | 0.2311D+03 | 0.4000D+01 | 0.5778D+02 | 0.6798D+02 | 0.1921D-01 |
| 5 | 0.4164D+04 | 0.2000D+01 | 0.2082D+04 | 0.2449D+04 | 0.3511D+00 |
| 6 | 0.4794D+03 | 0.4000D+01 | 0.1198D+03 | 0.1410D+03 | 0.4015D-01 |
| 7 | 0.3631D+02 | 0.4000D+01 | 0.9078D+01 | 0.1068D+02 | 0.2776D-02 |
| ERROR | 0.4250D+01 | 0.5000D+01 | 0.8500D+00 | | 0.1649D-02 |

# 7.5 CUMULATIVE METHOD

## 7.5.1 D4MU01, R4MU01
## Analysis of Variance Using the Cumulative Method

(1) **Function**

The D4MU01 or R4MU01 performs an analysis of variance using the cumulative method for data for which the number of repetitions is fixed and the number of factors is at most 3. The number of decision makers is assumed to be $r$ people, and the decision is assumed to be made in $d$ steps. Although the explanation presented below takes as an example the case when the number of factors is 3, one- and two-way layouts can also be handled by setting the number of levels of unused factors to 1.

Let the input data be represented as follows

$$x_{ijkl}^{(s)} \quad \left( \begin{array}{lcll} i & = & 1,\cdots,m_a & ;\ \text{Factor A} \\ j & = & 1,\cdots,m_b & ;\ \text{Factor B} \\ k & = & 1,\cdots,m_c & ;\ \text{Factor C} \\ l & = & 1,\cdots,r & ;\ \text{Decision maker R} \\ s & = & 1,\cdots,n & ;\ \text{Number of repetitions} \end{array} \right)$$

and let it take integer values from 1 to $d$. $m_a$, $m_b$ and $m_c$ are the number of levels in factors A, B, and C, respectively, and $r$ is the number of decision makers. In the explanations below, unless explicitly stated otherwise, summations by the $\sum$ symbol and their weights $W_m$ $(m = 1,\cdots,d-1)$ are defined as follows for subscripts such as $i$, $j$, $k$, $l$ and $s$.

$$\delta_{ijkl}^{(s)(m)} = \begin{cases} 1 & (\text{When } x_{ijkl}^{(s)} \leq m) \\ 0 & (\text{otherwise}) \end{cases}$$

$$P_m = \frac{\displaystyle\sum_i\sum_j\sum_k\sum_l\sum_s \delta_{ijkl}^{(s)(m)}}{m_a \cdot m_b \cdot m_c \cdot r \cdot n}$$

$$W_m = \frac{1}{P_m(1-P_m)}$$

The correction factor $CF$ is defined as follows.

$$CF = \frac{1}{m_a \cdot m_b \cdot m_c \cdot r \cdot n}\sum_{m=1}^{d-1}\left\{ W_m \sum_i\sum_j\sum_k\sum_l\sum_s \left(\delta_{ijkl}^{(s)(m)}\right)^2 \right\}$$

Also, the analysis of variance table consists of the following elements.

- Sum of squares

$$S_{A_i}^{(m)} = \sum_j\sum_k\sum_l\sum_s \delta_{ijkl}^{(s)(m)}$$

$$S_{B_j}^{(m)} = \sum_i\sum_k\sum_l\sum_s \delta_{ijkl}^{(s)(m)}$$

$$S_{C_k}^{(m)} = \sum_i\sum_j\sum_l\sum_s \delta_{ijkl}^{(s)(m)}$$

$$S_{R_l}^{(m)} = \sum_i\sum_j\sum_k\sum_s \delta_{ijkl}^{(s)(m)}$$

$$
\begin{aligned}
S_{AB_{ij}}^{(m)} &= \sum_k \sum_l \sum_s \delta_{ijkl}^{(s)(m)} \\
S_{AC_{ik}}^{(m)} &= \sum_j \sum_l \sum_s \delta_{ijkl}^{(s)(m)} \\
S_{BC_{jk}}^{(m)} &= \sum_i \sum_l \sum_s \delta_{ijkl}^{(s)(m)} \\
S_{AR_{il}}^{(m)} &= \sum_j \sum_k \sum_s \delta_{ijkl}^{(s)(m)} \\
S_{BR_{jl}}^{(m)} &= \sum_i \sum_k \sum_s \delta_{ijkl}^{(s)(m)} \\
S_{CR_{kl}}^{(m)} &= \sum_i \sum_j \sum_s \delta_{ijkl}^{(s)(m)} \\
S_{ABC_{ijk}}^{(m)} &= \sum_l \sum_s \delta_{ijkl}^{(s)(m)} \\
& (i = 1, \cdots, m_a; j = 1, \cdots, m_b; k = 1, \cdots, m_c) \\
& (l = 1, \cdots, r; m = 1, \cdots, d-1)
\end{aligned}
$$

$$
\begin{aligned}
S_T &= m_a \cdot m_b \cdot m_c \cdot r \cdot n \cdot (d-1) \\
S_A &= \frac{1}{m_b \cdot m_c \cdot r \cdot n} \sum_{m=1}^{d-1} \sum_i (S_{A_i}^{(m)})^2 W_m - CF \\
S_B &= \frac{1}{m_a \cdot m_c \cdot r \cdot n} \sum_{m=1}^{d-1} \sum_j (S_{B_j}^{(m)})^2 W_m - CF \\
S_C &= \frac{1}{m_a \cdot m_b \cdot r \cdot n} \sum_{m=1}^{d-1} \sum_k (S_{C_k}^{(m)})^2 W_m - CF \\
S_{AB} &= \frac{1}{m_c \cdot r \cdot n} \sum_{m=1}^{d-1} \sum_i \sum_j (S_{AB_{ij}}^{(m)})^2 W_m - CF - S_A - S_B \\
S_{BC} &= \frac{1}{m_a \cdot r \cdot n} \sum_{m=1}^{d-1} \sum_j \sum_k (S_{AB_{ij}}^{(m)})^2 W_m - CF - S_B - S_C \\
S_{AC} &= \frac{1}{m_b \cdot r \cdot n} \sum_{m=1}^{d-1} \sum_i \sum_k (S_{AB_{ij}}^{(m)})^2 W_m - CF - S_A - S_C \\
S_R &= \frac{1}{m_a \cdot m_b \cdot m_c \cdot n} \sum_{m=1}^{d-1} \sum_l (S_{R_l}^{(m)})^2 W_m - CF \\
S_{AR} &= \frac{1}{m_b \cdot m_c \cdot n} \sum_{m=1}^{d-1} \sum_i \sum_l (S_{AR_{il}}^{(m)})^2 W_m - CF - S_A - S_R \\
S_{BR} &= \frac{1}{m_a \cdot m_c \cdot n} \sum_{m=1}^{d-1} \sum_j \sum_l (S_{BR_{jl}}^{(m)})^2 W_m - CF - S_B - S_R \\
S_{CR} &= \frac{1}{m_a \cdot m_b \cdot n} \sum_{m=1}^{d-1} \sum_k \sum_l (S_{CR_{kl}}^{(m)})^2 W_m - CF - S_C - S_R \\
S_{ABC} &= \frac{1}{\cdot r \cdot n} \sum_{m=1}^{d-1} \sum_i \sum_j \sum_k (S_{ABC_{ijk}}^{(m)})^2 W_m - CF - S_A - S_B - S_C - S_{AB} \\
& -S_{AC} - S_{BC}
\end{aligned}
$$

$$\begin{aligned} S_E &= S_T - S_A - S_B - S_C - S_{AB} - S_{AC} - S_{BC} - S_R - S_{AR} - S_{BR} \\ &\quad -S_{CR} - S_{ABC} \end{aligned}$$

- Degrees of freedom

$$\begin{aligned} \phi_T &= (m_a \cdot m_b \cdot m_c \cdot r \cdot n - 1)(d-1) \\ \phi_A &= (m_a - 1)(d-1) \\ \phi_B &= (m_b - 1)(d-1) \\ \phi_C &= (m_c - 1)(d-1) \\ \phi_{AB} &= (m_a - 1)(m_b - 1)(d-1) \\ \phi_{BC} &= (m_b - 1)(m_c - 1)(d-1) \\ \phi_{AC} &= (m_a - 1)(m_c - 1)(d-1) \\ \phi_R &= (r-1)(d-1) \\ \phi_{AR} &= (m_a - 1)(r-1)(d-1) \\ \phi_{BR} &= (m_b - 1)(r-1)(d-1) \\ \phi_{CR} &= (m_c - 1)(r-1)(d-1) \\ \phi_{ABC} &= (m_a - 1)(m_b - 1)(m_c - 1)(d-1) \\ \phi_e &= \phi_T - \phi_A - \phi_B - \phi_C - \phi_{AB} - \phi_{BC} - \phi_{AC} - \phi_R - \phi_{AR} - \phi_{BR} \\ &\quad -\phi_{CR} - \phi_{ABC} \end{aligned}$$

- Unbiased variance

$$\begin{aligned} V_A &= \frac{S_A}{\phi_A} \\ V_B &= \frac{S_B}{\phi_B} \\ V_C &= \frac{S_C}{\phi_C} \\ V_{AB} &= \frac{S_{AB}}{\phi_{AB}} \\ V_{BC} &= \frac{S_{BC}}{\phi_{BC}} \\ V_{AC} &= \frac{S_{AC}}{\phi_{AC}} \\ V_R &= \frac{S_R}{\phi_R} \\ V_{AR} &= \frac{S_{AR}}{\phi_{AR}} \\ V_{BR} &= \frac{S_{BR}}{\phi_{BR}} \\ V_{CR} &= \frac{S_{CR}}{\phi_{CR}} \\ V_{ABC} &= \frac{S_{ABC}}{\phi_{ABC}} \\ V_E &= \frac{S_E}{\phi_E} \end{aligned}$$

- Variance ratio

$$
\begin{aligned}
F_A &= \frac{V_A}{V_E} \\
F_B &= \frac{V_B}{V_E} \\
F_C &= \frac{V_C}{V_E} \\
F_{AB} &= \frac{V_{AB}}{V_E} \\
F_{BC} &= \frac{V_{BC}}{V_E} \\
F_{AC} &= \frac{V_{AC}}{V_E} \\
F_R &= \frac{V_R}{V_E} \\
F_{AR} &= \frac{V_{AR}}{V_E} \\
F_{BR} &= \frac{V_{BR}}{V_E} \\
F_{CR} &= \frac{V_{CR}}{V_E} \\
F_{ABC} &= \frac{V_{ABC}}{V_E}
\end{aligned}
$$

- Contribution ratio

$$
\begin{aligned}
\rho_A &= \frac{S_A - \phi_A \cdot V_E}{S_T} \\
\rho_B &= \frac{S_B - \phi_B \cdot V_E}{S_T} \\
\rho_C &= \frac{S_C - \phi_C \cdot V_E}{S_T} \\
\rho_{AB} &= \frac{S_{AB} - \phi_{AB} \cdot V_E}{S_T} \\
\rho_{BC} &= \frac{S_{BC} - \phi_{BC} \cdot V_E}{S_T} \\
\rho_{AC} &= \frac{S_{AC} - \phi_{AC} \cdot V_E}{S_T} \\
\rho_R &= \frac{S_R - \phi_R \cdot V_E}{S_T} \\
\rho_{AR} &= \frac{S_{AR} - \phi_{AR} \cdot V_E}{S_T} \\
\rho_{BR} &= \frac{S_{BR} - \phi_{BR} \cdot V_E}{S_T} \\
\rho_{CR} &= \frac{S_{CR} - \phi_{CR} \cdot V_E}{S_T} \\
\rho_{ABC} &= \frac{S_{ABC} - \phi_{ABC} \cdot V_E}{S_T}
\end{aligned}
$$

The contribution ratio is calculated only when a significance of 5% or 1% is obtained by an $F$ test. Otherwise, 0.0 is set.

The density frequency is defined as follows.

$$
\begin{aligned}
\delta_{ijkl}^{'(s)(m)} &= \begin{cases} 1 & \text{When } x_{ijkl}^{(m)} = m \\ 0 & \text{otherwise} \end{cases} \\
T_{\cdot\cdot\cdot\cdot} &= \sum_i \sum_j \sum_k \sum_l \sum_{m=1}^{n} \delta_{ijkl}^{'(s)(m)} \\
T_{i\cdot\cdot\cdot} &= \sum_j \sum_k \sum_l \sum_{m=1}^{n} \delta_{ijkl}^{'(s)(m)} \\
T_{\cdot j\cdot\cdot} &= \sum_i \sum_k \sum_l \sum_{m=1}^{n} \delta_{ijkl}^{'(s)(m)} \\
T_{\cdot\cdot k\cdot} &= \sum_i \sum_j \sum_l \sum_{m=1}^{n} \delta_{ijkl}^{'(s)(m)} \\
T_{ij\cdot\cdot} &= \sum_k \sum_l \sum_{m=1}^{n} \delta_{ijkl}^{'(s)(m)} \\
T_{\cdot jk\cdot} &= \sum_i \sum_l \sum_{m=1}^{n} \delta_{ijkl}^{'(s)(m)} \\
T_{i\cdot k\cdot} &= \sum_i \sum_k \sum_{m=1}^{n} \delta_{ijkl}^{'(s)(m)} \\
T_{\cdot\cdot\cdot l} &= \sum_i \sum_j \sum_k \sum_{m=1}^{n} \delta_{ijkl}^{'(s)(m)} \\
T_{i\cdot\cdot l} &= \sum_j \sum_k \sum_{m=1}^{n} \delta_{ijkl}^{'(s)(m)} \\
T_{\cdot j\cdot l} &= \sum_i \sum_k \sum_{m=1}^{n} \delta_{ijkl}^{'(s)(m)} \\
T_{\cdot\cdot kl} &= \sum_i \sum_j \sum_{m=1}^{n} \delta_{ijkl}^{'(s)(m)} \\
T_{ij\cdot l} &= \sum_k \sum_{m=1}^{n} \delta_{ijkl}^{'(s)(m)} \\
T_{\cdot jkl} &= \sum_i \sum_{m=1}^{n} \delta_{ijkl}^{'(s)(m)} \\
T_{i\cdot kl} &= \sum_j \sum_{m=1}^{n} \delta_{ijkl}^{'(s)(m)} \\
T_{ijk\cdot} &= \sum_l \sum_{m=1}^{n} \delta_{ijkl}^{'(s)(m)} \\
T_{ijkl} &= \sum_{m=1}^{n} \delta_{ijkl}^{'(s)(m)}
\end{aligned}
$$

The predicted frequency of each level of each factor is defined as follows.

$$
\begin{aligned}
\bar{A}_i^{(m)} &= \frac{S_{A_i}^{(m)}}{\sum_{m=1}^{m_a} \bar{A}_i^{(m)}} \\
\bar{B}_j^{(m)} &= \frac{S_{B_j}^{(m)}}{\sum_{m=1}^{m_b} \bar{B}_j^{(m)}} \\
\bar{C}_k^{(m)} &= \frac{S_{C_k}^{(m)}}{\sum_{m=1}^{m_c} \bar{C}_k^{(m)}} \\
\bar{T}^{(m)} &= \frac{\sum_{ijkls} \delta_{ijkl}^{(s)(m)}}{m_a \cdot m_b \cdot m_c \cdot r \cdot n}
\end{aligned}
$$

The cumulative predicted frequency $\hat{\mu}_{ijk}^{(m)}$ is defined as follows.

$$
\begin{aligned}
\Omega_{ijk}^{(m)} &= \frac{\left(\frac{1}{\bar{A}_i^{(m)}} - 1\right)\left(\frac{1}{\bar{B}_j^{(m)}} - 1\right)\left(\frac{1}{\bar{C}_k^{(m)}} - 1\right)}{\left(\frac{1}{\bar{T}^{(m)}} - 1\right)} \\
\hat{\mu}_{ijk}^{(m)} &= \frac{1}{1 + \Omega_{ijk}^{(m)}} \qquad (m = 1, \cdots, d-1) \\
\hat{\mu}_{ijk}^{(d)} &= 1
\end{aligned}
$$

The predicted frequency $\hat{\alpha}_{ijk}^{(m)}$ is defined as follows.

$$
\begin{aligned}
\hat{\alpha}_{ijk}^{(1)} &= \hat{\mu}_{ijk}^{(1)} \\
\hat{\alpha}_{ijk}^{(m)} &= \hat{\mu}_{ijk}^{(m)} - \hat{\mu}_{ijk}^{(m-1)} \qquad (m = 2, \cdots, d)
\end{aligned}
$$

(2) **Usage**

Double precision:

CALL D4MU01 (IA, ID, V, IX, NX, NTC, NT, F, TX, OM, MA, AM, AL, MT, P, G, IERR)

Single precision:

CALL R4MU01 (IA, ID, V, IX, NX, NTC, NT, F, TX, OM, MA, AM, AL, MT, P, G, IERR)

### (3) **Arguments**

D:Double precision real　Z:Double precision complex　I: { INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer }
R:Single precision real　C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | IA | I | M1 | Input | Input data $x_{ijkl}^{(s)}$ (See Note (a)) Here, $\mathrm{M1} = \prod_{\mathrm{i}=1}^{5} \mathrm{ID(i)}$ |
| 2 | ID | I | 6 | Input | Number of levels of each factor, number of decision makers, number of repetitions, and number of steps (See Note (b)) |
| 3 | V | {D, R} | M2 | Output | Analysis of variance table (For the storage method of the analysis of variance table, see Tables 7−7, 7−8 and 7−9 in note (c).) Here, M2 is 28 when the number of factors is 1, 46 when the number of factors is 2, and 74 when the number of factors is 3. |
| 4 | IX | I | NX, ID(6) | Output | Density frequency table (See Note (e)) |
| 5 | NX | I | 1 | Input | Adjustable dimension of arrays IX and NT |
| 6 | NTC | I | 1 | Output | Size of density frequency table |
| 7 | NT | I | NX, 4 | Output | Frequency density table numbers (See Note (e)) |
| 8 | F | {D, R} | 3, 9, ID(6) | Output | Frequency in each step of each level $\bar{A}_i^{(m)}$, $\bar{B}_j^{(m)}$, $\bar{C}_k^{(m)}$ |
| 9 | TX | {D, R} | ID(6) | Output | $\bar{T}^{(m)}$ |
| 10 | OM | {D, R} | MA, ID(6) | Output | $\Omega_{ijk}^{(m)}$ (See Note (f)) |
| 11 | MA | I | 1 | Input | Adjustable dimension of arrays OM, AM, AL and MT |
| 12 | AM | {D, R} | MA, ID(6) | Output | Predicted cumulative frequency $\hat{\mu}_{ijk}^{(m)}$ (See Note (f)) |
| 13 | AL | {D, R} | MA, ID(6) | Output | Predicted frequency $\hat{\alpha}_{ijk}^{(m)}$(See Note (f)) |
| 14 | MT | I | MA, 3 | Output | $\Omega_{ijk}^{(m)}$, $\hat{\mu}_{ijk}^{(m)}$ or $\hat{\alpha}_{ijk}^{(m)}$ numbers (See Note (f)) |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 15 | P | $\left\{\begin{matrix}\mathrm{D}\\\mathrm{R}\end{matrix}\right\}$ | ID(6) | Output | $P_m$ |
| 16 | G | $\left\{\begin{matrix}\mathrm{D}\\\mathrm{R}\end{matrix}\right\}$ | ID(6) | Output | Weight $W_m$ |
| 17 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $2 \leq \mathrm{ID}(1) \leq 9$

(b) $1 \leq \mathrm{ID(i)} \leq 9 \quad (\mathrm{i} = 2, 3)$

(c) $1 \leq \mathrm{ID}(4) \leq 100$

(d) $1 \leq \mathrm{ID}(5) \leq 5$

(e) $1 \leq \mathrm{ID}(6) \leq 11$

(f) $\mathrm{NX} \geq \ (\mathrm{ID}(4) + 1) \times \prod_{\mathrm{i}=1,2,3,\mathrm{ID(i)}\neq 1} (\mathrm{ID(i)} + 1)$

(g) $\mathrm{MA} \geq \prod_{\mathrm{i}=1}^{3} \mathrm{ID(i)}$

(h) If the number of factors of the input data is 2, ID(3) = 1. If the number of factors of the input data is 1, ID(2) = ID(3) = 1.

(i) $1 \leq \mathrm{IA(i)} \leq \mathrm{ID}(6)$
$(\mathrm{i} = 1, 2, \cdots, \prod_{\mathrm{j}=1}^{5} \mathrm{ID(j)})$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Any of restriction (a) to (g) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (h) was not satisfied. | |
| 3020 | Restriction (i) was not satisfied. | |

(6) **Notes**

(a) The input data $x_{ijkl}^{(s)}$ is stored in array IA as follows.

$$\begin{aligned} \mathrm{IA}(\alpha) &= x_{ijkl}^{(s)} \\ \alpha &= i + (j-1) \times \mathrm{ID}(1) \\ &\quad +(k-1) \times \mathrm{ID}(1) \times \mathrm{ID}(2) \\ &\quad +(l-1) \times \mathrm{ID}(1) \times \mathrm{ID}(2) \times \mathrm{ID}(3) \\ &\quad +(s-1) \times \mathrm{ID}(1) \times \mathrm{ID}(2) \times \mathrm{ID}(3) \times \mathrm{ID}(4) \end{aligned}$$

That is, the input data is stored so that the subscripts vary in the order of $i$, $j$, $k$ and $l$.

(b) Information related to the input data $x_{ijkl}^{(s)}$ is stored in array ID as follows.

ID(1) : Number of levels of factor A $m_a$
ID(2) : Number of levels of factor B $m_b$
ID(3) : Number of levels of factor C $m_c$
ID(4) : Number of decision makers $r$
ID(5) : Number of repetitions $n$
ID(6) : Number of steps $d$

If the number of factors is 2, the number of levels of factor C should be set to 1. If the number of factors is 1, the number of levels of factor B and the number of levels of factor C should be set to 1.

(c) The analysis of variance table elements are stored as follows in the array V.

- Number of factors = 1

Table 7−7 Analysis of Variance Table Storage Status (Number of Factors = 1)

| Factor | Sum of squares | Degrees of freedom | Unbiased variance | Variance ratio | Contribution ratio | $F$ test result |
|---|---|---|---|---|---|---|
| T | V(1) | V(7) | | | | |
| A | V(2) | V(8) | V(13) | V(18) | V(21) | V(26) |
| R | V(3) | V(9) | V(14) | V(19) | V(22) | V(27) |
| AR | V(4) | V(10) | V(15) | V(20) | V(23) | V(28) |
| E | V(5) | V(11) | V(16) | | V(24) | |
| (E) | V(6) | V(12) | V(17) | | V(25) | |

- Number of factors = 2
- Number of factors = 3

(d) The value of an $F$ test result term of array V is set to 1.0 if the test result is a 1% significance, and it is set to 5.0 if the test result is a 5% significance. Otherwise, it is set to 0.0.

(e) The method of storing the density frequency table in array IX is explained below by taking as an example the case when the number of factors is 3. First, the following subscripts are provided corresponding to the subscripts $i$, $j$, $k$ and $l$ of density frequency $T_{ijkl}^{(m)}$.

$$\begin{aligned} i' &= 0, 1, 2, \cdots, m_a \\ j' &= 0, 1, 2, \cdots, m_b \\ k' &= 0, 1, 2, \cdots, m_c \\ l' &= 0, 1, 2, \cdots, r \end{aligned}$$

Table 7−8 Analysis of Variance Table Storage Status (Number of Factors = 2)

| Factor | Sum of squares | Degrees of freedom | Unbiased variance | Variance ratio | Contribution ratio | $F$ test result |
|---|---|---|---|---|---|---|
| T | V(1) | V(10) | | | | |
| A | V(2) | V(11) | V(19) | V(27) | V(33) | V(41) |
| B | V(3) | V(12) | V(20) | V(28) | V(34) | V(42) |
| R | V(4) | V(13) | V(21) | V(29) | V(35) | V(43) |
| AB | V(5) | V(14) | V(22) | V(30) | V(36) | V(44) |
| AR | V(6) | V(15) | V(23) | V(31) | V(37) | V(45) |
| BR | V(7) | V(16) | V(24) | V(32) | V(38) | V(46) |
| E | V(8) | V(17) | V(25) | | V(39) | |
| (E) | V(9) | V(18) | V(26) | | V(40) | |

Table 7−9 Analysis of Variance Table Storage Status (Number of Factors = 3)

| Factor | Sum of squares | Degrees of freedom | Unbiased variance | Variance ratio | Contribution ratio | $F$ test result |
|---|---|---|---|---|---|---|
| T | V(1) | V(13) | | | | |
| A | V(2) | V(14) | V(27) | V(40) | V(51) | V(64) |
| B | V(3) | V(15) | V(28) | V(41) | V(52) | V(65) |
| C | V(4) | V(16) | V(29) | V(42) | V(53) | V(66) |
| R | V(5) | V(17) | V(30) | V(43) | V(54) | V(67) |
| AB | V(5) | V(18) | V(31) | V(44) | V(55) | V(68) |
| AC | V(6) | V(19) | V(32) | V(45) | V(56) | V(69) |
| AR | V(7) | V(20) | V(33) | V(46) | V(57) | V(70) |
| BC | V(8) | V(21) | V(34) | V(47) | V(58) | V(71) |
| BR | V(8) | V(22) | V(35) | V(48) | V(59) | V(72) |
| CR | V(9) | V(23) | V(36) | V(49) | V(60) | V(73) |
| ABC | V(10) | V(24) | V(37) | V(50) | V(61) | V(74) |
| E | V(11) | V(25) | V(38) | | V(62) | |
| (E) | V(12) | V(26) | V(39) | | V(63) | |

Then, $D^{(m)}_{i'j'k'l'}$ is defined as follows.
When none of $i'$, $j'$, $k'$ and $l'$ is 0 :

$$D^{(m)}_{i'j'k'l'} = T^{(m)}_{i'j'k'l'}$$

When any of $i'$, $j'$, $k'$ and $l'$ is 0 :
$D_{i'j'k'l'}$ is defined as the summation of $T^{(m)}_{i'j'k'l'}$ for the subscript or subscripts that are 0. For example, when $j' = 0$ and $i'$, $k'$ and $l'$ are not 0, $D^{(m)}_{i'j'k'l'}$ is defined as follows.

$$D^{(m)}_{i'j'k'l'} = \sum_{j} T^{(m)}_{i'jk'l'} = T^{(m)}_{i'\cdot k'l'}$$

Or, when $i' = k' = 0$ and $j'$ and $l'$ are not 0, $D^{(m)}_{i'j'k'l'}$ is defined as follows.

$$D^{(m)}_{i'j'k'l'} = \sum_{i}\sum_{k} T^{(m)}_{ij'kl'} = T^{(m)}_{\cdot j'\cdot l'}$$

The density frequency table and these numbers are stored as follows in arrays IX and NT.

$\text{IX}(\beta, m) = D^{(m)}_{i'j'k'l'}$

$\text{NT}(\beta, 1) = i'$

$\text{NT}(\beta, 2) = j'$

$\text{NT}(\beta, 3) = k'$

$\text{NT}(\beta, 4) = l'$

$\beta = l' + k' \times \text{ID}(4) + j' \times \text{ID}(3) \times \text{ID}(4) + i' \times \text{ID}(2) \times \text{ID}(3) \times \text{ID}(4)$

That is, the subscripts of $T_{i'j'k'l'}$ vary in the order of $l'$, $k'$, $j'$ and $i'$ according to the variation of array subscript $\beta$. However, if the number of factors is 2, the values of $\text{NT}(\beta, 3)$ are set to 0, and if the number of factors is 1, the values of $\text{NT}(\beta, 2)$ and $\text{NT}(\beta, 3)$ are set to 0.

(f) $\Omega^{(m)}_{ijk}$, $\hat{\mu}^{(m)}_{ijk}$, and $\hat{\alpha}^{(m)}_{ijk}$ and the numbers corresponding to them are stored as follows in arrays OM, AM, AL and MT.

$\text{OM}(\beta, m) = \Omega^{(m)}_{ijk}$

$\text{AM}(\beta, m) = \hat{\mu}^{(m)}_{ijk}$

$\text{AL}(\beta, m) = \hat{\alpha}^{(m)}_{ijk}$

$\text{MT}(\beta, 1) = i$

$\text{MT}(\beta, 2) = j$

$\text{MT}(\beta, 3) = k$

$\beta = k + j \times \text{ID}(3) + i \times \text{ID}(2) \times \text{ID}(3)$

That is, the subscripts of $\Omega^{(m)}_{ijk}$, $\hat{\mu}^{(m)}_{ijk}$, and $\hat{\alpha}^{(m)}_{ijk}$ vary in the order of $k$, $j$ and $i$ according to the variation of array subscript $\beta$. However, if the number of factors is 2, the values of $\text{MT}(\beta, 3)$ are set to 0, and if the number of factors is 1, the values of $\text{MT}(\beta, 2)$ and $\text{MT}(\beta, 3)$ are set to 0.

(7) **Example**

(a) Problem
Perform an analysis of variance according to the cumulative method for the following observation data for which the number of factors is 2.

$$\begin{aligned} X \;=\; \{&2,3,3,1,2,3,2,3,2,1,2,2,2,3,2,1,1,3,2,3,\\ &3,1,3,3,2,3,2,1,3,2,2,3,3,1,2,3,3,2,3,2,\\ &3,2,3,2,3,1,2,1,3,1,3,1,2,3,1,2,3,1,3,3,\\ &2,3,3,2,3,2,1,3,2,2,2,3\}\end{aligned}$$

Assume that the number of levels of factor A is 3, the number of levels of factor B is 2, the number of decision makers is 6, the number of repetitions is 2 and that the decisions are made in 3 steps.

(b) Input data
Observed data $X$, NX=84, MA=6,
ID(1)=3, ID(2)=2, ID(3)=1, ID(4)=6, ID(5)=2 and ID(6)=3.

(c) Main program

```
      PROGRAM B4MU01
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NX = 84, MA = 6 )
      DIMENSION IA(72),ID(6),V(46),IX(NX,3),NT(NX,4),F(3,9,3),TX(3)
      DIMENSION OM(MA,3),AM(MA,3),AL(MA,3),MT(MA,3),P(3),G(3)
!
      WRITE(6,6000)
      IERR =  0
      READ(5,*) (IA(I),I=1,72)
      READ(5,*) (ID(I),I=1,6)
      WRITE(6,6010) (IA(I),I=1,72)
      WRITE(6,6020) (ID(I),I=1,6)
      CALL D4MU01(IA,ID,V,IX,NX,NTC,NT,F,TX,OM,MA,AM,AL,MT,P,G,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040)
      JD = IA(1)
      DO 100 I=2,72
         JD = MAX(JD,IA(I))
  100 CONTINUE
      DO 110 I=1,JD-1
         WRITE(6,6050) P(I),G(I)
  110 CONTINUE
      WRITE(6,6060)
      DO 120 I=1,9
         IF(I.EQ.1) THEN
            WRITE(6,6070) I,V(I),V(9+I)
         ELSE IF(I.LE.7) THEN
            WRITE(6,6070)&
            I,V(I),V(9+I),V(17+I),V(25+I),V(31+I),V(39+I)
         ELSE
            WRITE(6,6080)&
            I,V(I),V(9+I),V(17+I),V(31+I)
         ENDIF
  120 CONTINUE
      WRITE(6,6090)
      DO 130 I=1,NTC
         WRITE(6,6100) (NT(I,J),J=1,4),(IX(I,J),J=1,JD)
  130 CONTINUE
      WRITE(6,6110) 'A'
      DO 140 I=1,ID(1)
         WRITE(6,6120) (F(1,I,J),J=1,JD)
  140 CONTINUE
      WRITE(6,6110) 'B'
      DO 150 I=1,ID(2)
         WRITE(6,6120) (F(2,I,J),J=1,JD)
  150 CONTINUE
      WRITE(6,6110) 'TX'
      WRITE(6,6120) (TX(J),J=1,JD-1)
      WRITE(6,6130) 'OMEGA'
      DO 160 I=1,6
         WRITE(6,6140) (MT(I,J),J=1,3),(OM(I,J),J=1,JD-1)
  160 CONTINUE
      WRITE(6,6130) 'MU'
      DO 170 I=1,6
         WRITE(6,6140) (MT(I,J),J=1,3),(AM(I,J),J=1,JD-1)
  170 CONTINUE
      WRITE(6,6130) 'ALPHA'
      DO 180 I=1,6
         WRITE(6,6140) (MT(I,J),J=1,3),(AL(I,J),J=1,JD-1)
  180 CONTINUE
!
      STOP
 6000 FORMAT(  ' *** D4MU01 ***',/,&
             /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'OBSERVATIONS',/,/,&
                 8(6X,10I6,/))
 6020 FORMAT( /,7X,'NUMBER OF A          = ',I6,/,&
             /,7X,'NUMBER OF B          = ',I6,/,&
             /,7X,'NUMBER OF C          = ',I6,/,&
             /,7X,'NUMBER OF PERSONS    = ',I6,/,&
             /,7X,'NUMBER OF ITERATIONS = ',I6,/,&
             /,7X,'NUMBER OF STEPS      = ',I6)
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
             /,7X,'IERR = ',I6)
 6040 FORMAT( /,17X,'P',9X,'WEIGHT',/,&
              7X,27('-'))
 6050 FORMAT(  7X,5(2X,F11.4))
 6060 FORMAT( /,7X,'ANALYSIS OF VARIANCE TABLE',/,&
             /,10X,'FACTOR',&
                 4X,'S.S.',7X,'D.F.',6X,'M.S.',8X,'V.R.',8X,'C.R.',&
                 6X,'R.F.',/,&
              9X,71('-'))
 6070 FORMAT(  10X,I2,2X,2(F11.4),3(1X,D11.4),1X,F6.2)
 6080 FORMAT(  10X,I2,2X,2(F11.4),1X,D11.4,13X,D11.4,1X,F6.2)
 6090 FORMAT( /,7X,'DENSITY FREQUENCIES',/,&
             /,10X,'ABCR',/,&
              9X,30('-'))
 6100 FORMAT(  10X,4I1,3(2X,I6))
 6110 FORMAT( /,7X,'FREQUENCIES AT N STEP IN ',A,' LEVEL',/)
 6120 FORMAT(  9X,5(2X,D15.8))
 6130 FORMAT( /,7X,A,/,&
```

```
                /,10X,'ABC',/,&
                 9X,39('-'))
 6140 FORMAT(    10X,3I1,2(2X,D15.8))
      END
```

(d) Output results

```
*** D4MU01 ***

  ** INPUT **

      OBSERVATIONS

          2     3     3     1     2     3     2     3     2     1
          2     2     2     3     2     1     1     3     2     3
          3     1     3     3     2     3     2     1     3     2
          2     3     3     1     2     3     3     2     3     2
          3     2     3     2     3     1     2     1     3     1
          3     1     2     3     1     2     3     1     3     3
          2     3     3     2     3     2     1     3     2     2
          2     3

      NUMBER OF A          =      3

      NUMBER OF B          =      2

      NUMBER OF C          =      1

      NUMBER OF PERSONS    =      6

      NUMBER OF ITERATIONS =      2

      NUMBER OF STEPS      =      3

  ** OUTPUT **

      IERR =      0

                P         WEIGHT
      ---------------------------
             0.1944       6.3842
             0.5694       4.0787

      ANALYSIS OF VARIANCE TABLE

         FACTOR    S.S.        D.F.      M.S.         V.R.         C.R.      R.F.
        -------------------------------------------------------------------------
          1      144.0000   142.0000
          2       30.1867     4.0000  0.7547D+01  0.1083D+02  0.1892D+00   1.00
          3        8.4506     2.0000  0.4225D+01  0.6062D+01  0.4845D-01   1.00
          4        6.9038    10.0000  0.6904D+00  0.9905D+00  0.0000D+00   0.00
          5        8.0788     4.0000  0.2020D+01  0.2898D+01  0.3563D-01   5.00
          6       20.4133    20.0000  0.1021D+01  0.1464D+01  0.0000D+00   0.00
          7        5.8398    10.0000  0.5840D+00  0.8378D+00  0.0000D+00   0.00
          8       64.1269    92.0000  0.6970D+00              0.0000D+00
          9       97.2838   132.0000  0.7370D+00              0.7268D+00

      DENSITY FREQUENCIES

         ABCR
        ------------------------------
         0000       14       27       31
         1000       11       10        3
         2000        2        9       13
         3000        1        8       15
         0100        3       14       19
         0200       11       13       12
         0001        1        5        6
         0002        3        6        3
         0003        4        3        5
         0004        3        2        7
         0005        1        6        5
         0006        2        5        5
         1100        2        7        3
         1200        9        3        0
         2100        1        3        8
         2200        1        6        5
         3100        0        4        8
         3200        1        4        7
         1001        1        2        1
         1002        2        1        1
         1003        2        1        1
         1004        3        1        0
         1005        1        3        0
         1006        2        2        0
         2001        0        2        2
         2002        0        3        1
         2003        2        1        1
         2004        0        1        3
         2005        0        0        4
         2006        0        2        2
         3001        0        1        3
         3002        1        2        1
         3003        0        1        3
```

```
   3004       0       0       4
   3005       0       3       1
   3006       0       1       3
   0101       0       2       4
   0102       0       3       3
   0103       1       2       3
   0104       1       2       3
   0105       0       3       3
   0106       1       2       3
   0201       1       3       2
   0202       3       3       0
   0203       3       1       2
   0204       2       0       4
   0205       1       3       2
   0206       1       3       2
   1101       0       1       1
   1102       0       1       1
   1103       0       1       1
   1104       1       1       0
   1105       0       2       0
   1106       1       1       0
   1201       1       1       0
   1202       2       0       0
   1203       2       0       0
   1204       2       0       0
   1205       1       1       0
   1206       1       1       0
   2101       0       1       1
   2102       0       1       1
   2103       1       0       1
   2104       0       1       1
   2105       0       0       2
   2106       0       0       2
   2201       0       1       1
   2202       0       2       0
   2203       1       1       0
   2204       0       0       2
   2205       0       0       2
   2206       0       2       0
   3101       0       0       2
   3102       0       1       1
   3103       0       1       1
   3104       0       0       2
   3105       0       1       1
   3106       0       1       1
   3201       0       1       1
   3202       1       1       0
   3203       0       0       2
   3204       0       0       2
   3205       0       2       0
   3206       0       0       2

FREQUENCIES AT N STEP IN A LEVEL

     0.45833333D+00   0.41666667D+00   0.12500000D+00
     0.83333333D-01   0.37500000D+00   0.54166667D+00
     0.41666667D-01   0.33333333D+00   0.62500000D+00

FREQUENCIES AT N STEP IN B LEVEL

     0.83333333D-01   0.38888889D+00   0.52777778D+00
     0.30555556D+00   0.36111111D+00   0.33333333D+00

FREQUENCIES AT N STEP IN TX LEVEL

     0.19444444D+00   0.56944444D+00

OMEGA

   ABC
  --------------------------------------
   110   0.31379310D+01   0.21116834D+00
   120   0.64833286D+00   0.94470046D-01
   210   0.29206897D+02   0.17469381D+01
   220   0.60344828D+01   0.78152493D+00
   310   0.61068966D+02   0.24636306D+01
   320   0.12617555D+02   0.11021505D+01

MU

   ABC
  --------------------------------------
   110   0.24166667D+00   0.82564906D+00
   120   0.60667358D+00   0.91368421D+00
   210   0.33105023D-01   0.36404170D+00
   220   0.14215686D+00   0.56131687D+00
   310   0.16111111D-01   0.28871439D+00
   320   0.73434622D-01   0.47570332D+00

ALPHA

   ABC
  --------------------------------------
   110   0.24166667D+00   0.58398239D+00
   120   0.60667358D+00   0.30701063D+00
```

```
210    0.33105023D-01    0.33093668D+00
220    0.14215686D+00    0.41916001D+00
310    0.16111111D-01    0.27260328D+00
320    0.73434622D-01    0.40226870D+00
```

# 7.6 RANDOMIZED BLOCK DESIGN

## 7.6.1 D4RB01, R4RB01
## Analysis of Variance Using Randomized Block Design

(1) **Function**

The D4RB01 or R4RB01 performs an analysis of variance using randomized block design.

The analysis of variance results for observed values $\{x_{ij}\}$, $(i = 1, \cdots, n;\ j = 1, \cdots, t)$ obtained using $n$ blocks and $t$ trials are defined as follows.

Total sum :

$$T = \sum_{i=1}^{n} \sum_{j=1}^{t} x_{ij}$$

Total sum for each block :

$$T_{i\cdot} = \sum_{j=1}^{t} x_{ij}$$

Total sum for each trial :

$$T_{\cdot j} = \sum_{i=1}^{n} x_{ij}$$

Variation :

- Total variation

$$S = \sum_{i=1}^{n} \sum_{j=1}^{t} \left( x_{ij} - \frac{T}{n \cdot t} \right)^2$$

- Mean variation

$$S_c = \frac{T^2}{n \cdot t}$$

- Inter-block variation

$$S_p = \frac{1}{t} \sum_{i=1}^{n} \left( T_{i\cdot} - \frac{T}{n} \right)^2$$

- Inter-trial variation

$$S_r = \frac{1}{n} \sum_{j=1}^{t} \left( T_{\cdot j} - \frac{T}{t} \right)^2$$

- Error variation

$$S_e = S - (S_p + S_r)$$

Degrees of freedom :

$$\phi = n \cdot t - 1, \quad \phi_c = 1, \quad \phi_p = n - 1, \quad \phi_r = t - 1$$

$$\phi_e = (n - 1)(t - 1)$$

Unbiased variance :

$$V_p = \frac{S_p}{\phi_p}, \quad V_r = \frac{S_r}{\phi_r}, \quad V_e = \frac{S_e}{\phi_e}$$

Variance ratio :

$$F_p = \frac{V_p}{V_e}, \quad F_r = \frac{V_r}{V_e}$$

(2) **Usage**

Double precision:
CALL D4RB01 (A, NA, NB, NT, V, IERR)
Single precision:
CALL R4RB01 (A, NA, NB, NT, V, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: $\left\{ \begin{array}{l} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{array} \right\}$

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | NA,NT | Input | Matrix in which observed values are stored ($x_{ij}$) |
| 2 | NA | I | 1 | Input | Adjustable dimension of array A |
| 3 | NB | I | 1 | Input | Number of blocks $n$ |
| 4 | NT | I | 1 | Input | Number of trials $t$ |
| 5 | V | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 15 | Output | Analysis of variance results (See Table 7−10 in note (a)) |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) NA≥NB≥1
(b) NT≥1

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | NB=1 or NT=1 was specified. | The absolute value maximum that can be represented is set for the unbiased variance and variance ratio. |
| 3000 | Restriction (a), (b) was not satisfied. | Processing is aborted. |

(6) **Notes**

(a) The analysis of variance results are stored as follows in the array V.

Table 7−10 Analysis of Variance Table

| Factor | Variation | Degrees of freedom | Unbiased variance | Variance ratio |
|---|---|---|---|---|
| Total | V(1) | V(6) | | |
| Mean | V(2) | V(7) | | |
| Inter-block | V(3) | V(8) | V(11) | V(14) |
| Inter-trial | V(4) | V(9) | V(12) | V(15) |
| Error | V(5) | V(10) | V(13) | |

(7) **Example**

(a) Problem

When the observed values are given by matrix $X$ shown below, perform an analysis of variance using randomized block design.

$$X = \begin{bmatrix} 42 & 28 & 19 & 7 & 4 \\ 15 & 14 & -19 & -4 & -6 \\ -8 & 30 & -17 & 9 & -31 \end{bmatrix}$$

(b) Input data

Observation matrix $X$, NA=100, NB=3 and NT=5.

(c) Main program

```
      PROGRAM B4RB01
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 100, NT = 5 )
      DIMENSION A(NA,NT),V(15)
!
      WRITE(6,6000)
      IERR = 0
      READ(5,*) NB
      DO 100 I=1,NB
         READ(5,*) (A(I,J),J=1,NT)
  100 CONTINUE
      WRITE(6,6010) NB,NT
      DO 110 I=1,NB
         WRITE(6,6020) (A(I,J),J=1,NT)
  110 CONTINUE
      CALL D4RB01(A,NA,NB,NT,V,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040)
      WRITE(6,6050) 'TOTAL    ',V(1),V(6)
      WRITE(6,6050) 'MEAN     ',V(2),V(7)
      WRITE(6,6050) 'BLOCK    ',V(3),V(8),V(11),V(14)
      WRITE(6,6050) 'TREATMENT',V(4),V(9),V(12),V(15)
      WRITE(6,6050) 'ERROR    ',V(5),V(10),V(13)
!
      STOP
 6000 FORMAT(  ' *** D4RB01 ***',/,&
             /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'NB = ',I6,5X,'NT = ',I6,/,&
             /,7X,'OBSERVATION MATRIX',/)
 6020 FORMAT(  7X,5(2X,F11.2))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
             /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'ANALYSIS OF VARIANCE TABLE',/,&
             /,10X,'FACTOR',&
                  10X,'S.S.',9X,'D.F.',8X,'M.S.',13X,'V.R.',/,&
              9X,69('-'))
 6050 FORMAT(  10X,A,2(1X,F11.2),2(2X,D15.8))
      END
```

(d) Output results

```
*** D4RB01 ***

  ** INPUT **
```

```
    NB =      3     NT =      5

    OBSERVATION MATRIX

            42.00         28.00         19.00          7.00          4.00
            15.00         14.00        -19.00         -4.00         -6.00
            -8.00         30.00        -17.00          9.00        -31.00

** OUTPUT **

    IERR =      0

    ANALYSIS OF VARIANCE TABLE

       FACTOR          S.S.          D.F.        M.S.              V.R.
      ----------------------------------------------------------------------
       TOTAL         5643.73       14.00
       MEAN           459.27        1.00
       BLOCK         1598.53        2.00   0.79926667D+03   0.43533043D+01
       TREATMENT     2576.40        4.00   0.64410000D+03   0.35081699D+01
       ERROR         1468.80        8.00   0.18360000D+03
```

## 7.7 GRECO-LATIN SQUARE METHOD

### 7.7.1 D4GL01, R4GL01 Analysis of Variance Using the Greco-Latin Square Method

(1) **Function**

The D4GL01 or R4GL01 performs an analysis of variance using the Greco-Latin Square method.

The analysis of variance results for the two orthogonal Latin squares $MT$ and $MG$ and observed values $\{x_{ij(kl)}\}$ ($i$: Levels of factor P; $j$: Levels of factor Q; $k$ and $l$: Constitute Greco-Latin square) are defined as follows.

Total sum, row sum, and column sum :

$$T = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij(kl)}, \quad T_{i\cdot} = \sum_{j=1}^{n} x_{ij(kl)}, \quad T_{\cdot j} = \sum_{i=1}^{n} x_{ij(kl)}$$

$$T_k = \sum_{k\in MT} x_{ij(kl)}, \quad T_l = \sum_{l\in MG} x_{ij(kl)}$$

Variation:

- Total variation
$$S_s = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(x_{ij(kl)} - \frac{T}{n^2}\right)^2$$
- Inter-row variation
$$S_p = \frac{1}{n}\sum_{i=1}^{n}\left(T_{i\cdot} - \frac{T}{n}\right)^2$$
- Inter-column variation
$$S_q = \frac{1}{n}\sum_{j=1}^{n}\left(T_{\cdot j} - \frac{T}{n}\right)^2$$
- Inter-R variation
$$S_r = \frac{1}{n}\sum_{k=1}^{n}\left(T_k - \frac{T}{n}\right)^2$$
- Inter-A variation
$$S_a = \frac{1}{n}\sum_{l=1}^{n}\left(T_l - \frac{T}{n}\right)^2$$
- Error variation
$$S_e = S_s - (S_p + S_q + S_r + S_a)$$

Degrees of freedom :

$$\phi_s = n^2 - 1, \quad \phi_p = \phi_q = \phi_r = \phi_a = n - 1, \quad \phi_e = (n-1)(n-3)$$

Unbiased variance :

$$V_p = \frac{S_p}{\phi_p}, \quad V_q = \frac{S_q}{\phi_q}, \quad V_r = \frac{S_r}{\phi_r}, \quad V_a = \frac{S_a}{\phi_a}, \quad V_e = \frac{S_e}{\phi_e}$$

Variance ratio :

$$F_p = \frac{V_p}{V_e}, \quad F_q = \frac{V_q}{V_e}, \quad F_r = \frac{V_r}{V_e}, \quad F_a = \frac{V_a}{V_e}$$

(2) **Usage**

Double precision:

CALL D4GL01 (A, NA, N, MT, MG, V, IWK, WK, IERR)

Single precision:

CALL R4GL01 (A, NA, N, MT, MG, V, IWK, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: {INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer}

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | {D / R} | NA,N | Input | Matrix in which observed values are stored ($x_{ij(kl)}$) |
| 2 | NA | I | 1 | Input | Adjustable dimension of arrays A, MT and MG |
| 3 | N | I | 1 | Input | Number of trials $n$ |
| 4 | MT | I | NA,N | Input | Latin square $MT$ |
| 5 | MG | I | NA,N | Input | Latin square $MG$ |
| 6 | V | {D / R} | 19 | Output | Analysis of variance results (See Table 7−11 in note (a)) |
| 7 | IWK | I | N×N | Work | Work area |
| 8 | WK | {D / R} | N | Work | Work area |
| 9 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) NA≥N≥3

(b) MT and MG are Latin squares.

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | N=3 was specified. | The absolute value maximum that can be represented is set for the unbiased variance and variance ratio. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |

(6) **Notes**

(a) The analysis of variance results are stored as follows in the array V.

Table 7−11 Analysis of Variance Table

| Factor | Variation | Degrees of freedom | Unbiased variance | Variance ratio |
|---|---|---|---|---|
| P (row) | V(1) | V(6) | V(11) | V(16) |
| Q (column) | V(2) | V(7) | V(12) | V(17) |
| R | V(3) | V(8) | V(13) | V(18) |
| A | V(4) | V(9) | V(14) | V(19) |
| Error | V(5) | V(10) | V(15) | |

(b) The factors are allocated so that factors P and Q are in a two-dimensional arrangement, and factors R and A are associated with Latin squares $MT$ and $MG$ for using the various levels.

Table 7−12 Experiment Allocation for Greco-Latin Square Method

| | $Q_1$ | $Q_2$ | $Q_3$ | $Q_4$ |
|---|---|---|---|---|
| $P_1$ | $R_2A_1$ | $R_4A_3$ | $R_1A_2$ | $R_3A_4$ |
| $P_2$ | $R_4A_2$ | $R_2A_4$ | $R_3A_1$ | $R_1A_3$ |
| $P_3$ | $R_1A_4$ | $R_3A_2$ | $R_2A_3$ | $R_4A_1$ |
| $P_4$ | $R_3A_3$ | $R_1A_1$ | $R_4A_4$ | $R_2A_2$ |

(7) **Example**

(a) Problem
When the observed values $X$ and two Latin squares $MT$ and $MG$ are given as shown below, perform an analysis of variance using the Greco-Latin square method.

$$X = \begin{bmatrix} 8.6 & 11.0 & 17.2 & 18.2 \\ 13.5 & 13.4 & 20.3 & 19.0 \\ 14.8 & 20.5 & 16.9 & 18.7 \\ 18.7 & 17.2 & 20.7 & 22.8 \end{bmatrix}$$

$$MT = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 3 & 4 & 1 & 2 \\ 4 & 3 & 2 & 1 \end{bmatrix}$$

$$MG = \begin{bmatrix} 1 & 3 & 4 & 2 \\ 2 & 4 & 3 & 1 \\ 3 & 1 & 2 & 4 \\ 4 & 2 & 1 & 3 \end{bmatrix}$$

(b) Input data
Observation matrix $X$, Latin squares $MT$ and $MG$, NA=100 and N=4.

(c) Main program

```
      PROGRAM B4GL01
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 100, N = 4 )
      DIMENSION A(NA,N),MT(NA,N),MG(NA,N),V(19)
      DIMENSION IWK(N*N),WK(N)
```

```
!
      WRITE(6,6000)
      IERR =    0
      DO 100 I=1,N
         READ(5,*) (MT(I,J),J=1,N)
  100 CONTINUE
      DO 110 I=1,N
         READ(5,*) (MG(I,J),J=1,N)
  110 CONTINUE
      DO 120 I=1,N
         READ(5,*) (A(I,J),J=1,N)
  120 CONTINUE
      WRITE(6,6010) N
      WRITE(6,6020)
      DO 130 I=1,N
         WRITE(6,6030) (MT(I,J),J=1,N)
  130 CONTINUE
      WRITE(6,6040)
      DO 140 I=1,N
         WRITE(6,6030) (MG(I,J),J=1,N)
  140 CONTINUE
      WRITE(6,6050)
      DO 150 I=1,N
         WRITE(6,6060) (A(I,J),J=1,N)
  150 CONTINUE
      CALL D4GL01(A,NA,N,MT,MG,V,IWK,WK,IERR)
      WRITE(6,6070) IERR
      WRITE(6,6080)
      WRITE(6,6090) 'ROW         ',V(1),V(6),V(11),V(16)
      WRITE(6,6090) 'COLUMN      ',V(2),V(7),V(12),V(17)
      WRITE(6,6090) 'TREATMENTS R',V(3),V(8),V(13),V(18)
      WRITE(6,6090) 'TREATMENTS A',V(4),V(9),V(14),V(19)
      WRITE(6,6090) 'ERROR       ',V(5),V(10),V(15)
!
      STOP
 6000 FORMAT(  ' *** D4GL01 ***',/,&
              /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'N = ',I6)
 6020 FORMAT( /,7X,'LATIN SQUARE MT',/)
 6030 FORMAT(  7X,5(2X,I6))
 6040 FORMAT( /,7X,'LATIN SQUARE MG',/)
 6050 FORMAT( /,7X,'OBSERVATION MATRIX',/)
 6060 FORMAT(  7X,5(2X,F11.2))
 6070 FORMAT( /,3X,'** OUTPUT **',/,&
              /,7X,'IERR = ',I6)
 6080 FORMAT( /,7X,'ANALYSIS OF VARIANCE TABLE',/,&
              /,10X,'FACTOR',&
                  12X,'S.S.',8X,'D.F.',8X,'M.S.',13X,'V.R.',/,&
               9X,70('-'))
 6090 FORMAT(  10X,A,2(F11.2),2(2X,D15.8))
      END
```

(d) Output results

```
*** D4GL01 ***

  ** INPUT **

      N =      4

      LATIN SQUARE MT

             1       2       3       4
             2       1       4       3
             3       4       1       2
             4       3       2       1

      LATIN SQUARE MG

             1       3       4       2
             2       4       3       1
             3       1       2       4
             4       2       1       3

      OBSERVATION MATRIX

               8.60        11.00        17.20        18.20
              13.50        13.40        20.30        19.00
              14.80        20.50        16.90        18.70
              18.70        17.20        20.70        22.80

  ** OUTPUT **

      IERR =      0

      ANALYSIS OF VARIANCE TABLE

         FACTOR            S.S.        D.F.        M.S.             V.R.
        ----------------------------------------------------------------------
         ROW               77.64        3.00  0.25878958D+02  0.58874354D+01
         COLUMN            88.35        3.00  0.29450625D+02  0.66999858D+01
         TREATMENTS R      37.64        3.00  0.12547292D+02  0.28544955D+01
         TREATMENTS A       1.56        3.00  0.51895833D+00  0.11806247D+00
         ERROR             13.19        3.00  0.43956250D+01
```

## 7.8 BALANCED INCOMPLETE BLOCK DESIGN

### 7.8.1 D4BI01, R4BI01
### Analysis of Variance Using Balanced Incomplete Block Design

(1) **Function**

The D4BI01 or R4BI01 performs an analysis for balanced incomplete block design data. An incomplete block design indicates the case in which one set of trials (treatments) to be compared is incomplete and is not entered in the blocks. In particular, a block design in which the number of repetitions of each trial is equal and the number of times two arbitrary trials appear in the same block is equal is called a balanced incomplete block design.

The following example shows a situation in which the number of blocks is 4, number of trials is 4, and number of trials per block is 3.

| Blocks | Trials | | | | $T_{i\cdot}$ |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| $A_1$ | $x_{11}$ | | $x_{13}$ | $x_{14}$ | $T_{1\cdot}$ |
| $A_2$ | | $x_{22}$ | $x_{23}$ | $x_{24}$ | $T_{2\cdot}$ |
| $A_3$ | $x_{31}$ | $x_{32}$ | $x_{33}$ | | $T_{3\cdot}$ |
| $A_4$ | $x_{41}$ | $x_{42}$ | | $x_{44}$ | $T_{4\cdot}$ |
| $T_{\cdot j}$ | $T_{\cdot 1}$ | $T_{\cdot 2}$ | $T_{\cdot 3}$ | $T_{\cdot 4}$ | $T$ |

For this example, the number of times two arbitrary trials appear in the same block at the same time is 2. Given the data $\{x_{ij}\}$, $(i = 1, \cdots, n; j = 1, \cdots, t)$, which is allocated using a balanced incomplete block design for which the number of blocks is $n$, number of trials is $t$, and number of trials per block is $m$, the analysis of variance results are defined as follows.

Total sum :

$$T = \sum_{i=1}^{n} \sum_{j=1}^{t} n_{ij} x_{ij}$$

Total sum for each block :

$$T_{i\cdot} = \sum_{j=1}^{t} n_{ij} x_{ij}$$

Total sum for each trial :

$$T_{\cdot j} = \sum_{i=1}^{n} n_{ij} x_{ij}$$

Total sum for each block-adjusted trial :

$$Q_{\cdot j} \quad = \quad m \cdot T_{\cdot j} - \sum_{i=1}^{n} (n_{ij} \cdot T_{i\cdot})$$

$(n_{ij} : (i, j)$ element of incidence matrix $N)$

Variation :

- Total variation

$$S = \sum_{i=1}^{n} \sum_{j=1}^{t} n_{ij} x_{ij}^2 - CF$$

  Here, $CF$ is defined as follows.

$$CF = \frac{T^2}{n \cdot m} = \frac{1}{n \cdot m} \cdot \left( \sum_{i=1}^{n} \sum_{j=1}^{t} n_{ij} x_{ij} \right)^2$$

- Inter-block variation

$$S_B = \frac{1}{m} \sum_{i=1}^{n} T_{i\cdot}^2 - CF$$

- Inter-trial variation (block adjusted)

$$S_T = \frac{t-1}{n \cdot m^2 \cdot (m-1)} \sum_{j=1}^{t} Q_{\cdot j}^2$$

- Error variation

$$S_E = S - (S_B + S_T)$$

Degrees of freedom :

$$\begin{aligned} \phi &= n \cdot m - 1 \\ \phi_B &= n - 1 \\ \phi_T &= t - 1 \\ \phi_E &= n \cdot m - n - t + 1 \end{aligned}$$

Unbiased variance :

$$\begin{aligned} V_T &= \frac{S_T}{\phi_T} \\ V_B &= \frac{S_B}{\phi_B} \\ V_E &= \frac{S_E}{\phi_E} \end{aligned}$$

Variance ratio :

$$F_T = \frac{V_T}{V_B}$$

(2) **Usage**

Double precision:

CALL D4BI01 (A, NA, NB, NT, M, N, V, W1, IERR)

Single precision:

CALL R4BI01 (A, NA, NB, NT, M, N, V, W1, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | NA,NT | Input | Matrix in which observed values are stored ($x_{ij}$) |
| 2 | NA | I | 1 | Input | Adjustable dimension of arrays A and N |
| 3 | NB | I | 1 | Input | Total number of blocks $n$ |
| 4 | NT | I | 1 | Input | Number of trials $t$ |
| 5 | M | I | 1 | Input | Number of trials per block $m$ |
| 6 | N | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | NA,NT | Input | Incidence matrix ($n_{ij}$). When trial $j$ is performed in the $i$-th block, $n_{ij} = 1$, otherwise $n_{ij} = 0$. |
| 7 | V | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 11 | Output | Analysis of variance results (For the method of storing the analysis of variance results, see Table 7−13 in note (b)) |
| 8 | W1 | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | NB | Work | Work area |
| 9 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) NA $\geq$ NB $\geq 2$

(b) NT $\geq$ M $\geq 2$

(c) $\dfrac{\text{NB} \cdot \text{M}}{\text{NT}}$ is an integer.

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) or (b) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (c) was not satisfied. | |

(6) **Notes**

(a) The analysis of variance results are stored as follows in the array V.

Table 7−13 Analysis of Variance Table

| Element | Variation | Degrees of freedom | Unbiased variance | Variance ratio |
|---|---|---|---|---|
| Sum of squares of observed values | V(1) | V(5) | | |
| Inter-block | V(2) | V(6) | V(9) | V(12) |
| Inter-trial (block adjusted) | V(3) | V(7) | V(10) | |
| Error | V(4) | V(8) | V(11) | |

(7) **Example**

(a) Problem

Perform an analysis of variance for the data allocated using a balanced incomplete block design given by the matrix $A$ shown below. The asterisks ($*$) represent portions for which no corresponding observed values exist.

$$A = \begin{bmatrix} 2 & * & 4 & 0 \\ * & 32 & 13 & 23 \\ 20 & 14 & 31 & * \\ 7 & 3 & * & 11 \end{bmatrix}$$

The incidence matrix $N$ is given as follows.

$$N = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

(b) Input dataData allocated by a balanced incomplete block design $A$, incidence matrix $N$, NA=4, NB=4, NT=4 and M=3.

(c) Main program

```
      PROGRAM B4BI01
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 5,NT = 4)
      DIMENSION A(NA,NT),N(NA,NT),W1(NA),V(12)
!
      WRITE(6,6000)
      IERR = 0
      READ(5,*) NB,M
      DO 100 I = 1,NB
         READ(5,*) (A(I,J),J=1,NT)
  100 CONTINUE
      DO 110 I = 1,NB
         READ(5,*) (N(I,J),J=1,NT)
  110 CONTINUE
      WRITE(6,6010) NA,NB,NT,M
      WRITE(6,6020)
      DO 120 I=1,NB
         WRITE(6,6030) (A(I,J),J=1,NT)
  120 CONTINUE
      WRITE(6,6040)
      DO 130 I=1,NB
         WRITE(6,6050) (N(I,J),J=1,NT)
  130 CONTINUE
      CALL D4BI01(A,NA,NB,NT,M,N,V,W1,IERR)
      WRITE(6,6060) IERR
      WRITE(6,6070)
      WRITE(6,6080) 'TOTAL',V(1),V(5)
      WRITE(6,6090) ' SB  ',V(2),V(6),V(9),V(12)
      WRITE(6,6100) ' ST  ',V(3),V(7),V(10)
      WRITE(6,6100) 'ERROR',V(4),V(8),V(11)
!
      STOP
 6000 FORMAT(  ' *** D4BI01 ***',/,&
             /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'NA = ',I6,5X,'NB = ',I6,/,&
             /,7X,'NT = ',I6,5X,'M  = ',I6)
 6020 FORMAT( /,3X,'A',/)
 6030 FORMAT(  7X,4(2X,F11.2),/)
 6040 FORMAT( /,3X,'N',/)
 6050 FORMAT(  7X,4(2X,I6,3X),/)
 6060 FORMAT( /,3X,'** OUTPUT **',/,&
             /,7X,'IERR = ',I6)
 6070 FORMAT( /,7X,'ANALYSIS OF VARIANCE TABLE',/,&
             /,10X,'FACTOR',&
                 8X,'S.S.',9X,'D.F.',6X,'M.S.',8X,'V.R.',/,&
              9X,69('-'))
 6080 FORMAT(  10X,A,2(2X,F11.4))
 6090 FORMAT(  10X,A,2(2X,F11.4),2(1X,D11.4))
 6100 FORMAT(  10X,A,2(2X,F11.4),1X,D11.4)
      END
```

(d) Output results

```
*** D4BI01 ***
  ** INPUT **
      NA =      5     NB =      4
      NT =      4     M  =      3
  A
             2.00         0.00         4.00         0.00
             0.00        32.00        13.00        23.00
            20.00        14.00        31.00         0.00
             7.00         3.00         0.00        11.00

  N
           1          0          1          1
           0          1          1          1
           1          1          1          0
           1          1          0          1

  ** OUTPUT **
      IERR =      0
      ANALYSIS OF VARIANCE TABLE
         FACTOR        S.S.         D.F.      M.S.        V.R.
        ---------------------------------------------------------------------
         TOTAL    1344.6667      11.0000
          SB       975.3333       3.0000  0.3251D+03  0.6323D-02
          ST         6.1667       3.0000  0.2056D+01
         ERROR     363.1667       5.0000  0.7263D+02
```

# Chapter 8

# NONPARAMETRIC TESTS

## 8.1 INTRODUCTION

In traditional statistical tests, a normal distribution often is assumed as the population distribution. It is also a prerequisite that the observed values are continuous quantities. A test of this kind for which the population distribution is fixed and the observed values are assumed to be continuous quantities is called a **parametric test**. In contrast, a test for which the population distribution is not fixed or for which the data subject to the test is represented by sequences or frequencies, not by observed values that are complete quantities, is called a **nonparametric test**.
This library provides the following functions for performing nonparametric tests.

- Test of Goodness of Fit
- $\chi^2$ Test ($2 \times 2$ Contingency Table)
- $\chi^2$ Test ($m \times n$ Contingency Table)
- Median Test
- Sign Test
- Wilcoxon Test
- Mann-Whitney's U Test
- Spearman's Rank Correlation Test

### 8.1.1 Explanation

(1) **Test of Goodness of Fit**

Given observed frequencies, this test tests whether or not their distribution is equal to a theoretical distribution postulated by the researcher. The goodness-of-fit test is performed by obtaining the $\chi^2$ value for the test for the theoretical probabilities $p_i,\ (i = 1, \cdots, n)$ and observed frequencies $f_i,\ (i = 1, \cdots, n)$ of $n$ classes as defined below:

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_i - e_i)^2}{e_i}$$

and comparing this value with the critical value of a $\chi^2$ distribution with $n - 1$ degrees of freedom.

(2) $\chi^2$ **Test** ($2 \times 2$ **Contingency Table)**

This is a test for testing the independence of two factors by using a $2 \times 2$ contingency table.

| | $A$ | $\bar{A}$ | Sum |
|---|---|---|---|
| $B$ | $a_{11}$ | $a_{12}$ | $a_{11} + a_{12}$ |
| $\bar{B}$ | $a_{21}$ | $a_{22}$ | $a_{21} + a_{22}$ |
| Sum | $a_{11} + a_{21}$ | $a_{12} + a_{22}$ | $n$ |

In the $2 \times 2$ contingency table shown above, $A$ means that the subject has the characteristic being observed by factor $A$, and $\bar{A}$ means that the subject does not have the characteristic. $B$ and $\bar{B}$ have similar meanings for factor $B$. Also, $a_{ij}$ represents the actual measured frequency of the corresponding characteristics. The $\chi^2$ value, which is the test quantity, for the above contingency table is given as follows.

$$\chi^2 = \frac{n\left(|a_{11}a_{22} - a_{12}a_{21}| - \dfrac{n}{2}\right)^2}{(a_{11} + a_{12})(a_{21} + a_{22})(a_{11} + a_{21})(a_{12} + a_{22})}$$

Here, $n = a_{11} + a_{12} + a_{21} + a_{22}$. This equation contains Yates' correction for continuity term $-\frac{n}{2}$. The $\chi^2$ test is performed by comparing this $\chi^2$ value with the critical value of a $\chi^2$ distribution with 1 degree of freedom.

(3) $\chi^2$ **Test** ($m \times n$ **Contingency Table)**

This is a test for testing the independence of two factors by using an $m \times n$ contingency table.

| | $B_1$ | $B_2$ | $\cdots$ | $B_j$ | $\cdots$ | $B_n$ | Sum |
|---|---|---|---|---|---|---|---|
| $A_1$ | $a_{11}$ | $a_{12}$ | | $\vdots$ | | $a_{1n}$ | $a_{1\cdot}$ |
| $A_2$ | $a_{21}$ | $a_{22}$ | | $\vdots$ | | $a_{2n}$ | $a_{2\cdot}$ |
| $\vdots$ | | | | $\vdots$ | | | $\vdots$ |
| $A_i$ | $\cdots$ | $\cdots$ | $\cdots$ | $a_{ij}$ | $\cdots$ | $\cdots$ | $a_{i\cdot}$ |
| $\vdots$ | | | | $\vdots$ | | | $\vdots$ |
| $A_m$ | $a_{m1}$ | $a_{m2}$ | | $\vdots$ | | $a_{mn}$ | $a_{m\cdot}$ |
| Sum | $a_{\cdot 1}$ | $a_{\cdot 2}$ | $\cdots$ | $a_{\cdot j}$ | $\cdots$ | $a_{\cdot n}$ | S |

In the $m \times n$ contingency table shown above, $A_i$ means that the subject has the $i$th level of the characteristic being observed by factor $A$, and each level is assumed to be independent. The meanings are similar for $B_j$ and factor $B$. Also, $a_{ij}$ represents the actual measured frequency of the subject having the characteristics for both the $i$th level of factor $A$ and the $j$th level of factor $B$. The $\chi^2$ value for the above contingency table is given as follows.

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(a_{ij} - e_{ij})^2}{e_{ij}}$$

The $\chi^2$ test is performed by obtaining this $\chi^2$ value and comparing it with the critical value of a $\chi^2$ distribution with $(m-1)(n-1)$ degrees of freedom. Here, $e_{ij}$ is defined according to the following values.
Row sum :

$$a_{i\cdot} = \sum_{j=1}^{n} a_{ij}$$

Column sum :

$$a_{\cdot j} = \sum_{i=1}^{m} a_{ij}$$

Total sum :

$$S = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}$$

Expected value :

$$e_{ij} = \frac{a_{i\cdot} \cdot a_{\cdot j}}{S}$$

(4) **Median Test**
The median test tests the hypothesis that the median values of two groups are equal. The test is performed by obtaining the $\chi^2$ value and comparing it with the critical value of a $\chi^2$ distribution with one degree of freedom.

(5) **Sign Test**
When the observed values of two samples are individually associated and their relative sizes can be judged (including determining that they are equal), the sign test tests the hypothesis that it is equally probable that the judgements "it is larger" $(+)$ and "it is smaller" $(-)$ will appear. The critical value of a standard normal distribution is used for the test.

(6) **Wilcoxon Test**
The Wilcoxon test is a more precise test than the sign test. The critical value of a standard normal distribution is used for the test.

(7) **Mann-Whitney's U Test**
When two independent groups of samples are given, Mann-Whitney's U Test tests whether or not the distributions of the populations to which those samples belong are equal. The critical value of a standard normal distribution is used for the test.

(8) **Spearman's Rank Correlation Test**
When $n$ pairs of observed values $(x_i, y_i)$, $(i = 1, \cdots, n)$ are given, if $x_i$ and $y_i$ are each independently

ranked and the rankings are represented by $R(x_i)$ and $R(y_i)$, the Spearman's rank correlation coefficient $r_s$ is defined as follows.

$$r_s = 1 - \frac{6\displaystyle\sum_{i=1}^{n}(R(x_i) - R(y_i))^2}{n^3 - n}$$

Spearman's rank correlation coefficient is an index representing the strength of the association of $R(x_i)$ and $R(y_i)$. This index can be used to test the hypothesis that there is no correlation between $x_i$ and $y_i$.

### 8.1.2 Reference Bibliography


(1) Gibbons, J. D. , "Nonparametric Statistical Inference", McGraw-Hill (1971)

(2) Lehmann, E. L. , "Nonparametrics: Statistical methods based on ranks", Holden-Day, San Francisco (1975)

## 8.2 TESTS USING $\chi^2$ DISTRIBUTION

### 8.2.1 D5CHEF, R5CHEF
### Test of Goodness of Fit

(1) **Function**

The D5CHEF or R5CHEF tests the goodness of fit to the expected frequencies (theoretical frequencies) of given observed frequencies. The values for the theoretical probabilities $p_i,\ (i = 1, \cdots, n)$ and observed frequencies $f_i,\ (i = 1, \cdots, n)$ of $n$ classes are defined by the equations below.

Total of frequencies of all classes :

$$S = \sum_{i=1}^{n} f_i$$

Expected frequency (theoretical frequency) of $i$th class :

$$e_i = S \cdot p_i$$

$\chi^2$ value for test :

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_i - e_i)^2}{e_i}$$

Degrees of freedom :

$$\phi = n - 1$$

(2) **Usage**

Double precision:

CALL D5CHEF (P, N, F, IDF, CHI, IERR)

Single precision:

CALL R5CHEF (P, N, F, IDF, CHI, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\\text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | P | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | N | Input | Theoretical probabilities $\{p_i\}$ |
| 2 | N | I | 1 | Input | Number of classes of expected frequencies $n$ |
| 3 | F | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | N | Input | Observed frequencies $\{f_i\}$ |
| 4 | IDF | I | 1 | Output | Degrees of freedom $\phi$ |
| 5 | CHI | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | 1 | Output | $\chi^2$ value |
| 6 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) N≥2

(b) 0<P(i)<1 (i=1, $\cdots$, N)

(c) $\sum_{\mathrm{i=1}}^{\mathrm{N}}\mathrm{P(i)}\leq 1$

(d) F(i)≥ 0 (i=1, $\cdots$, N)

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |

(6) **Notes**

(a) For the given theoretical probabilities P(i), $(\mathrm{i} = 1, \cdots, \mathrm{N})$, the error is reduced by replacing P(N) by the following:

$$\mathrm{P(N)} = 1 - \sum_{\mathrm{i=1}}^{\mathrm{N-1}} \mathrm{P(i)}$$

(7) **Example**

(a) Problem
When the number of classes is 7 and the theoretical probability $p_i$ and observed frequency $f_i$ of each class are given as shown below, perform a test of the goodness of fit to the expected frequencies (theoretical frequencies) of the observed frequencies.

| Class | $p_i$ | $f_i$ |
|---|---|---|
| 1 | 0.1 | 10 |
| 2 | 0.2 | 20 |
| 3 | 0.3 | 30 |
| 4 | 0.1 | 10 |
| 5 | 0.1 | 10 |
| 6 | 0.12 | 12 |
| 7 | 0.05 | 5 |

(b) Input data

Theoretical probabilities $p_i$, observed frequencies $f_i$ and N = 7.

(c) Main program

```
      PROGRAM B5CHEF
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( N = 7 )
      DIMENSION P(N),F(N)
!
      WRITE(6,6000)
      IERR = 0
      DO 100 I=1,N
         READ(5,*) P(I)
  100 CONTINUE
      DO 110 I=1,N
         READ(5,*) F(I)
  110 CONTINUE
      WRITE(6,6010) N
      DO 120 I=1,N
         WRITE(6,6020) P(I),F(I)
  120 CONTINUE
      CALL D5CHEF(P,N,F,IDF,CHI,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) IDF
      WRITE(6,6050) CHI
!
      STOP
 6000 FORMAT(   ' *** D5CHEF ***',/,&
             /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'N = ',I6,/,&
             /,8X,'PROBABILITY VALUE',4X,'FREQUENCY',/,&
              7X,32('-'))
 6020 FORMAT(  7X,F11.2,7X,F11.2)
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
             /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'DEGREE OF FREEDOM = ',5X,I6)
 6050 FORMAT( /,7X,'CHI SQUARE VALUE  = ',D15.8)
      END
```

(d) Output results

```
*** D5CHEF ***

  ** INPUT **

      N =      7

       PROBABILITY VALUE    FREQUENCY
      --------------------------------
             0.10             10.00
             0.20             20.00
             0.30             30.00
             0.10             10.00
             0.10             10.00
             0.12             12.00
             0.05              5.00

  ** OUTPUT **

      IERR =      0

      DEGREE OF FREEDOM =           6

      CHI SQUARE VALUE  =  0.10670103D+01
```

### 8.2.2 D5CHTT, R5CHTT
### $\chi^2$ Test ($2 \times 2$ Contingency Table)

(1) **Function**

The D5CHTT or R5CHTT obtains the $\chi^2$ value, which is the test quantity including the Yates' correction for continuity term, by using a 2×2 contingency table.

| | $A$ | $\bar{A}$ | Sum |
|---|---|---|---|
| $B$ | $a_{11}$ | $a_{12}$ | $a_{11} + a_{12}$ |
| $\bar{B}$ | $a_{21}$ | $a_{22}$ | $a_{21} + a_{22}$ |
| Sum | $a_{11} + a_{21}$ | $a_{12} + a_{22}$ | $n$ |

In the $2 \times 2$ contingency table shown above, $A$ means that the subject has the characteristic being observed by factor $A$, and $\bar{A}$ means that the subject does not have the characteristic. $B$ and $\bar{B}$ have similar meanings for factor $B$. Also, $a_{ij}$ represents the actual measured frequency of the corresponding characteristics. The $\chi^2$ value, which is the test quantity, for the above contingency table is given as follows.

$$\chi^2 = \frac{n\left(|a_{11}a_{22} - a_{12}a_{21}| - \dfrac{n}{2}\right)^2}{(a_{11}+a_{12})(a_{21}+a_{22})(a_{11}+a_{21})(a_{12}+a_{22})}$$

Here, $n = a_{11} + a_{12} + a_{21} + a_{22}$.

The $\chi^2$ test tests the independence of two factors by comparing this $\chi^2$ value with the critical value of a $\chi^2$ distribution with 1 degree of freedom.

(2) **Usage**

Double precision:

CALL D5CHTT (F, CHI, IERR)

Single precision:

CALL R5CHTT (F, CHI, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: {INTEGER(4) as for 32bit Integer, INTEGER(8) as for 64bit Integer}

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | F | {D, R} | 2,2 | Input | Observed frequencies constituting the contingency table $a_{ij}$ |
| 2 | CHI | {D, R} | 1 | Output | $\chi^2$ value |
| 3 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) F(i, j)≥0 (i=1, 2; j=1, 2)

## (5) Error indicator

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | All observed frequencies were zero. | The absolute value maximum that can be represented is set for the $\chi^2$ value. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |

## (6) Notes

(a) The observed frequencies of the contingency table are stored in array F as the following real matrix (two-dimensional array type) (See Appendix A).

$$\begin{bmatrix} a_{11} & a_{21} \\ a_{21} & a_{22} \end{bmatrix}$$

## (7) Example

(a) Problem
Obtain the $\chi^2$ value, which is the test quantity including the Yates' correction for continuity term, by using the following 2×2 contingency table.

| | $A$ | $\bar{A}$ | Sum |
|---|---|---|---|
| $B$ | 6 | 0 | 6 |
| $\bar{B}$ | 1 | 3 | 4 |
| Sum | 7 | 3 | 10 |

(b) Input data

$$\mathrm{F}(1,1) = 6.0,\ \mathrm{F}(2,1) = 1.0,\ \mathrm{F}(1,2) = 0.0 \text{ and } \mathrm{F}(2,2) = 3.0.$$

(c) Main program

```
      PROGRAM B5CHTT
!
      IMPLICIT REAL(8)(A-H,O-Z)
      DIMENSION F(2,2)
!
      WRITE(6,6000)
      IERR   =   0
      DO 100 I=1,2
         READ(5,*) (F(I,J),J=1,2)
  100 CONTINUE
      WRITE(6,6010)
      DO 110 I=1,2
         WRITE(6,6020) (F(I,J),J=1,2)
  110 CONTINUE
      CALL D5CHTT(F,CHI,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) CHI
!
      STOP
 6000 FORMAT(  ' *** D5CHTT ***',/,&
             /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'TWO-BY-TWO CONTINGENCY TABLE',/)
 6020 FORMAT(  7X,5(2X,F11.2))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
             /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'CHI SQUARE VALUE = ',D15.8)
      END
```

(d) Output results

```
*** D5CHTT ***

  ** INPUT **

      TWO-BY-TWO CONTINGENCY TABLE
```

```
            6.00          0.00
            1.00          3.00

** OUTPUT **

    IERR =       0

    CHI SQUARE VALUE =  0.33531746D+01
```

### 8.2.3 D5CHMN, R5CHMN $\chi^2$ Test ($m \times n$ Contingency Table)

(1) **Function**

The D5CHMN or R5CHMN obtains the $\chi^2$ value, which is the test quantity, by using an $m \times n$ contingency table.

| | $B_1$ | $B_2$ | $\cdots$ | $B_j$ | $\cdots$ | $B_n$ | Sum |
|---|---|---|---|---|---|---|---|
| $A_1$ | $a_{11}$ | $a_{12}$ | | $\vdots$ | | $a_{1n}$ | $a_{1\cdot}$ |
| $A_2$ | $a_{21}$ | $a_{22}$ | | $\vdots$ | | $a_{2n}$ | $a_{2\cdot}$ |
| $\vdots$ | | | | $\vdots$ | | | $\vdots$ |
| $A_i$ | $\cdots$ | $\cdots$ | $\cdots$ | $a_{ij}$ | $\cdots$ | $\cdots$ | $a_{i\cdot}$ |
| $\vdots$ | | | | $\vdots$ | | | $\vdots$ |
| $A_m$ | $a_{m1}$ | $a_{m2}$ | | $\vdots$ | | $a_{mn}$ | $a_{m\cdot}$ |
| Sum | $a_{\cdot 1}$ | $a_{\cdot 2}$ | $\cdots$ | $a_{\cdot j}$ | $\cdots$ | $a_{\cdot n}$ | S |

In the $m \times n$ contingency table shown above, $A_i$ means that the subject has the $i$th level of the characteristic being observed by factor $A$, and each level is assumed to be independent. The meanings are similar for $B_j$ and $B$. Also, $a_{ij}$ represents the actual measured frequency of the subject having the characteristics for both the $i$th level of factor $A$ and the $j$th level of factor $B$. The $\chi^2$ value, which is the test quantity, for the above contingency table is given as follows.

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(a_{ij} - e_{ij})^2}{e_{ij}}$$

Here, the expected values $e_{ij}$ are defined as follows.
Expected value :

$$e_{ij} = \frac{a_{i\cdot} \cdot a_{\cdot j}}{S}$$

Row sum :

$$a_{i\cdot} = \sum_{j=1}^{n} a_{ij}$$

Column sum :

$$a_{\cdot j} = \sum_{i=1}^{m} a_{ij}$$

Total sum :

$$S = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}$$

The $\chi^2$ test tests the independence of two factors by comparing this $\chi^2$ value with the critical value of a $\chi^2$ distribution with $(m-1)(n-1)$ degrees of freedom.

(2) **Usage**

Double precision:

CALL D5CHMN (A, NA, M, N, IDF, CHI, WK, IERR)

Single precision:

CALL R5CHMN (A, NA, M, N, IDF, CHI, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: {INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer}
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | {D / R} | NA,N | Input | Observed frequencies constituting the contingency table ($a_{ij}$) |
| 2 | NA | I | 1 | Input | Adjustable dimension of array A |
| 3 | M | I | 1 | Input | Number of rows of contingency table $m$ |
| 4 | N | I | 1 | Input | Number of columns of contingency table $n$ |
| 5 | IDF | I | 1 | Output | Degrees of freedom $\phi$ |
| 6 | CHI | {D / R} | 1 | Output | $\chi^2$ value |
| 7 | WK | {D / R} | MAX(M, N) | Work | Work area |
| 8 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) NA≥M≥2

(b) N≥2

(c) A(i, j)≥0 (i=1, ⋯, M; j=1, ⋯, N)

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | The expected value in one or more cells was less than or equal to 1.0. | CHI and IDF are calculated. |
| 3000 | Restriction (a), (b) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (c) was not satisfied. | |
| 4000 | All observation frequencies were 0.0 or the expected value was less than or equal to 0.0. | |

(6) **Notes**

(a) The observed frequencies of the contingency table are stored in array F as an $m \times n$ real matrix (two-dimensional array type) defined as $A = (a_{i,j})$ (See Appendix A).

(7) **Example**

(a) Problem
Obtain the $\chi^2$ value by using the following 3×4 matrix, which corresponds to a 3×4 contingency table of observed frequencies.

$$\begin{bmatrix} 34 & 50 & 24 & 12 \\ 22 & 65 & 115 & 24 \\ 12 & 41 & 68 & 20 \end{bmatrix}$$

(b) Input data

Array corresponding to observed frequencies A, NA = 5, M = 3 and N = 4.

(c) Main program

```
      PROGRAM B5CHMN
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( NA = 5, N = 4 )
      DIMENSION A(NA,N),WK(NA)
!
      WRITE(6,6000)
      IERR = 0
      READ(5,*) M
      DO 100 I=1,M
         READ(5,*) (A(I,J),J=1,N)
  100 CONTINUE
      WRITE(6,6010) M,N
      DO 110 I=1,M
         WRITE(6,6020) (A(I,J),J=1,N)
  110 CONTINUE
      CALL D5CHMN(A,NA,M,N,IDF,CHI,WK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) IDF
      WRITE(6,6050) CHI
!
      STOP
 6000 FORMAT(   ' *** D5CHMN ***',/,&
              /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'M = ',I6,5X,'N = ',I6,/,&
              /,7X,'CONTINGENCY TABLE',/)
 6020 FORMAT(  7X,5(2X,F11.2))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
              /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'DEGREE OF FREEDOM = ',5X,I6)
 6050 FORMAT( /,7X,'CHI SQUARE VALUE  = ',D15.8)
      END
```

(d) Output results

```
*** D5CHMN ***

  ** INPUT **

     M =      3     N =      4

     CONTINGENCY TABLE

             34.00        50.00        24.00        12.00
             22.00        65.00       115.00        24.00
             12.00        41.00        68.00        20.00

  ** OUTPUT **

     IERR =      0

     DEGREE OF FREEDOM =           6

     CHI SQUARE VALUE  =  0.48649182D+02
```

### 8.2.4 D5CHMD, R5CHMD Median Test

(1) **Function**

The D5CHMD or R5CHMD obtains the $\chi^2$ value, which is the test quantity, by using the median test for two independent samples.

The value for the observation values $x_i,\ (i = 1, 2, \cdots, n)$ and $y_j,\ (j = 1, 2, \cdots, m)$ of two samples is defined by the following equation.

$\chi^2$ value :

$$\chi^2 = \frac{(n+m)\left(\mid a \cdot d - b \cdot c \mid - \dfrac{n+m}{2}\right)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Here, $a$, $b$, $c$, and $d$ indicate the following values.

$a$: Among $x_i$, number of observed values that are larger than the median value of the $(n+m)$ observed values

$b$: Among $x_i$, number of observed values that are smaller than the median value of the $(n+m)$ observed values

$c$: Among $y_j$, number of observed values that are larger than the median value of the $(n+m)$ observed values

$d$: Among $y_j$, number of observed values that are smaller than the median value of the $(n+m)$ observed values

However, for observed values that are equal to the median value of the $(n+m)$ observed values, frequencies of 0.5 each are added to both the number that are larger and the number that are smaller.

The median test tests the hypothesis that the median values of two groups are equal by comparing this $\chi^2$ value with the critical value of a $\chi^2$ distribution with one degree of freedom.

(2) **Usage**

Double precision:

CALL D5CHMD (A, N, B, M, CHI, WK, IERR)

Single precision:

CALL R5CHMD (A, N, B, M, CHI, WK, IERR)

(3) **Arguments**

| D:Double precision real | Z:Double precision complex | I: INTEGER(4) as for 32bit Integer |
| --- | --- | --- |
| R:Single precision real | C:Single precision complex | INTEGER(8) as for 64bit Integer |

| No. | Argument | Type | Size | Input/ Output | Contents |
| --- | --- | --- | --- | --- | --- |
| 1 | A | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | N | Input | Observed values of sample A $\{x_i\}$ |
| 2 | N | I | 1 | Input | Number of observed values of sample A $n$ |
| 3 | B | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | M | Input | Observed values of sample B $\{y_i\}$ |
| 4 | M | I | 1 | Input | Number of observed values of sample B $m$ |
| 5 | CHI | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | 1 | Output | $\chi^2$ value |
| 6 | WK | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | N+M | Work | Work area |
| 7 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) N≥2

(b) M≥2

(5) **Error indicator**

| IERR value | Meaning | Processing |
| --- | --- | --- |
| 0 | Normal termination. | |
| 1000 | Sample A and sample B are mutually prime. | 0.0 is set for CHI. |
| 3000 | Restriction (a), (b) was not satisfied. | Processing is aborted. |

(6) **Notes**
None

(7) **Example**

(a) Problem
Given the following observed values for two independent samples,

$$\{x_i\} = \{160, 160, 140, 190\}$$

and

$$\{y_i\} = \{117, 145, 147, 120, 150, 120\}$$

obtain the $\chi^2$ value by using the median test.

(b) Input data

Observed values $\{x_i\}$, $\text{N} = 4$, observed values $\{y_i\}$ and $\text{M} = 6$.

(c) Main program

```
      PROGRAM B5CHMD
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( N = 4, M = 6 )
      DIMENSION A(N),B(M),WK(N+M)
!
      WRITE(6,6000)
      IERR = 0
      DO 100 I=1,N
         READ(5,*) A(I)
  100 CONTINUE
      DO 110 I=1,M
         READ(5,*) B(I)
  110 CONTINUE
      WRITE(6,6010) N,M
      WRITE(6,6020) 'A',(A(I),I=1,N)
      WRITE(6,6020) 'B',(B(I),I=1,M)
      CALL D5CHMD(A,N,B,M,CHI,WK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) CHI
!
      STOP
 6000 FORMAT(  ' *** D5CHMD ***',/,&
             /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'N = ',I6,5X,'M = ',I6)
 6020 FORMAT( /,7X,'OBSERVATIONS ',A,/,/,&
                  2(6X,5(2X,F11.2),/))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
             /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'CHI SQUARE VALUE = ',D15.8)
      END
```

(d) Output results

```
*** D5CHMD ***

  ** INPUT **

      N =      4     M =      6

      OBSERVATIONS A

            160.00       160.00       140.00       190.00

      OBSERVATIONS B

            117.00       145.00       147.00       120.00       150.00
            120.00

  ** OUTPUT **

      IERR =      0

      CHI SQUARE VALUE =  0.41666667D+00
```

## 8.3 TESTS USING OTHER DISTRIBUTION

### 8.3.1 D5TESG, R5TESG
### Sign Test

(1) **Function**

The D5TESG or R5TESG performs a sign test when the observed values of two samples are individually associated.

When $(X, Y)$ is assumed to be a pair of random variables and $n$ pairs of observed values $(x_i, y_i)$, $(i = 1, 2, \cdots, n)$ are given as the actual values for them, this subroutine tests the null hypothesis $\mathrm{H}_0$ :"$P_r(X > Y) = P_r(X < Y) = 0.5$" against the alternative hypothesis $\mathrm{H}_1$: "$P_r(X > Y) > 0.5$ (or $< 0.5$)."

The value for the $n$ pairs of observed values $(x_i, y_i)$, $(i = 1, 2, \cdots, n)$ is defined by the equations shown below.

The following kinds of values $a$, $b$, $x$, and $m$ are obtained from the observed values.

$a$ : Number of pairs of observed values for which $x_i > y_i$.

$b$ : Number of pairs of observed values for which $x_i < y_i$.

$x = \min(a, b)$

$m = a + b$

Probability :

- For $m \leq 25$

$$P = \frac{1}{2^m} \sum_{i=0}^{x} \begin{pmatrix} m \\ i \end{pmatrix}$$

- For $m > 25$

$$P = \int_{-\infty}^{Z} \frac{1}{\sqrt{2 \cdot \pi}} e^{-\frac{x^2}{2}} dx$$

Here, $Z$ is defined as follows.

$$Z = \frac{x + 0.5 - \dfrac{m}{2}}{\dfrac{1}{2}\sqrt{m}}$$

($Z$ obeys a standard normal distribution $N(0, 1)$.)

(2) **Usage**

Double precision:

CALL D5TESG (A, N, B, IZR, ISN, P, IERR)

Single precision:

CALL R5TESG (A, N, B, IZR, ISN, P, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | N | Input | Observed values of sample A $\{x_i\}$ |
| 2 | N | I | 1 | Input | Number of observed values in each of the two samples $n$ |
| 3 | B | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | N | Input | Observed values of sample B $\{y_i\}$ |
| 4 | IZR | I | 1 | Output | Number of times the difference of the corresponding observed values of samples A and B is not zero $m$ |
| 5 | ISN | I | 1 | Output | Smaller number of appearances of the signs of the difference of the corresponding observed values (number of appearances of $+$ or $-$) $x$ |
| 6 | P | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Output | Probability (one-tailed test) $P$ |
| 7 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) N≥1

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Samples A and B are equal (IZR=0). | ISN=0 and P=1 are set. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |

(6) **Notes**

(a) The number of observed values in the two samples must be equal.

(b) When this subroutine is used, the test is a one-tailed test. However, when this procedure is used for a two-tailed test, if the significance level is $\alpha$, the null hypothesis is rejected when the following relationship holds.

$$P \leq \frac{\alpha}{2}$$

(7) **Example**

(a) Problem

Perform a sign test for the following set of observed values of two samples.

$$\begin{aligned}\{(x_i, y_i)\} &= \{(4,2),(4,3),(5,3),(5,3),(3,3),(2,3),(5,3),(3,3),(1,2),\\ &\quad (5,3),(5,2),(5,2),(4,5),(5,2),(5,5),(5,3),(5,1)\}\end{aligned}$$

(b) Input data

Set of observed values $\{(x_i, y_i)\}$ and N=17.

(c) Main program

```
      PROGRAM B5TESG
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( N = 17 )
      DIMENSION A(N),B(N)
!
      WRITE(6,6000)
      IERR = 0
      DO 100 I=1,N
         READ(5,*) A(I),B(I)
  100 CONTINUE
      WRITE(6,6010) N
      DO 110 I=1,N
         WRITE(6,6020) I,A(I),B(I)
  110 CONTINUE
      CALL D5TESG(A,N,B,IZR,ISN,P,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) IZR
      WRITE(6,6050) ISN
      WRITE(6,6060) P
!
      STOP
 6000 FORMAT(   ' *** D5TESG ***',/,&
              /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'N = ',I6,/,&
              /,7X,'OBSERVATIONS',/,&
              /,11X,'NO.',9X,'A',12X,'B',/,&
               9X,31('-'))
 6020 FORMAT(  7X,I6,2(2X,F11.2))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
              /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'NUMBER OF PAIRS = ',I6)
 6050 FORMAT( /,7X,'NUMBER OF SIGNS = ',I6)
 6060 FORMAT( /,7X,'PROBABILITY = ',D15.8,2X,&
                   'SIGNIFICANT AT 0.05 LEVEL')
      END
```

(d) Output results

```
*** D5TESG ***

  ** INPUT **

      N =     17

      OBSERVATIONS

          NO.         A            B
        -------------------------------
           1        4.00         2.00
           2        4.00         3.00
           3        5.00         3.00
           4        5.00         3.00
           5        3.00         3.00
           6        2.00         3.00
           7        5.00         3.00
           8        3.00         3.00
           9        1.00         2.00
          10        5.00         3.00
          11        5.00         2.00
          12        5.00         2.00
          13        4.00         5.00
          14        5.00         2.00
          15        5.00         5.00
          16        5.00         3.00
          17        5.00         1.00

  ** OUTPUT **

      IERR =      0

      NUMBER OF PAIRS =     14

      NUMBER OF SIGNS =      3

      PROBABILITY =  0.28686523D-01  SIGNIFICANT AT 0.05 LEVEL
```

### 8.3.2 D5TEWL, R5TEWL
### Wilcoxon Test

(1) **Function**

The D5TEWL or R5TEWL performs a Wilcoxon test when the observed values of two samples are individually associated.

When $(X, Y)$ is assumed to be a pair of random variables and $n$ pairs of observed values $(x_i, y_i)$, $(i = 1, 2, \cdots, n)$ are given as the actual values for them, this subroutine tests the null hypothesis $\mathrm{H}_0$: "$P_r(X > Y) = P_r(X < Y) = 0.5$" against the alternative hypothesis $\mathrm{H}_1$: "$P_r(X > Y) > 0.5$ (or $< 0.5$)."

The value for the $n$ pairs of observed values $(x_i, y_i)$, $(i = 1, 2, \cdots, n)$ is defined by the equations shown below.

Difference of corresponding observed value of two samples :

$$d_i = x_i - y_i, \quad (i = 1, \cdots, n)$$

The following kinds of values $m$, $R_i$, $T_P$, and $T_N$ are obtained from the observed values.

$m$ : Number of $d_i$ that are not zero.

$R_i$ : For nonzero $d_i$, ranks assigned to their absolute values. The mean rank is assigned for equal ranks.

$T_P$ : Sum of ranks assigned to positive $d_i$

$T_N$ : Sum of ranks assigned to negative $d_i$

Test statistic :

$$T = \min(T_P, T_N)$$

Expected value of $T$ :

$$E[T] = \frac{m(m+1)}{4}$$

Variance of $T$:

$$V[T] = \frac{m(m+1)(2 \cdot m+1)}{24}$$

Probability :

$$P = \int_{-\infty}^{Z} \frac{1}{\sqrt{2 \cdot \pi}} e^{-\frac{x^2}{2}} dx$$

Here, $Z$ is defined as follows.

$$Z = \frac{T - E[T]}{\sqrt{V[T]}}$$

(The $Z$ for which $T$ has been normalized obeys a standard normal distribution $N(0, 1)$.)

(2) **Usage**

Double precision:

CALL D5TEWL (A, N, B, IZR, T, Z, P, IWK, WK, IERR)

Single precision:

CALL R5TEWL (A, N, B, IZR, T, Z, P, IWK, WK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: {INTEGER(4) as for 32bit Integer, INTEGER(8) as for 64bit Integer}
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | {D, R} | N | Input | Observed values of sample A $\{x_i\}$ |
| 2 | N | I | 1 | Input | Number of observed values in each of the two samples $n$ |
| 3 | B | {D, R} | N | Input | Observed values of sample B $\{y_i\}$ |
| 4 | IZR | I | 1 | Output | Number of times the difference of the corresponding observed values of samples A and B is not zero $m$ |
| 5 | T | {D, R} | 1 | Output | Test statistic $T$ |
| 6 | Z | {D, R} | 1 | Output | Value for computing the significance of T according to a normal distribution (value for which T has been normalized) $Z$ |
| 7 | P | {D, R} | 1 | Output | Probability (one-tailed test) $P$ |
| 8 | IWK | I | 3×N | Work | Work area |
| 9 | WK | {D, R} | 2×N | Work | Work area |
| 10 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) N≥2

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Samples A and B are equal (IZR=0). | T=0 and P=0 are set, and the absolute value maximum that can be represented is set for Z. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |

(6) **Notes**

(a) The number of observed values in the two samples must be equal.

(b) When this subroutine is used, the test is a one-tailed test. However, when this procedure is used for a two-tailed test, if the significance level is $\alpha$, the null hypothesis is rejected when the following relationship holds.

$$P \leq \frac{\alpha}{2}$$

(7) **Example**

(a) Problem

Perform a Wilcoxon test for the following set of observed values of two samples.

$$\begin{aligned}\{(x_i, y_i)\} &= \{(28,36),(96,24),(37,47),(34,73),(85,15),(56,34),(67,80),\\ &\quad (56,82),(19,78),(27,41),(61,56),(76,32),(13,31),(43,41)\}\end{aligned}$$

(b) Input data

Set of observed values $\{(x_i, y_i)\}$ and N=14.

(c) Main program

```
      PROGRAM B5TEWL
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( N = 14 )
      DIMENSION A(N),B(N),IWK(3*N),WK(2*N)
!
      WRITE(6,6000)
      IERR = 0
      DO 100 I=1,N
         READ(5,*) A(I),B(I)
  100 CONTINUE
      WRITE(6,6010) N
      DO 110 I=1,N
         WRITE(6,6020) I,A(I),B(I)
  110 CONTINUE
      CALL D5TEWL(A,N,B,IZR,T,Z,P,IWK,WK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) IZR
      WRITE(6,6050) T
      WRITE(6,6060) Z
      WRITE(6,6070) P
!
      STOP
 6000 FORMAT(  ' *** D5TEWL ***',/,&
              /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'N = ',I6,/,&
              /,7X,'OBSERVATIONS',/,&
              /,11X,'NO.',9X,'A',12X,'B',/,&
               9X,31('-'))
 6020 FORMAT(  7X,I6,2(2X,F11.2))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
              /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'NUMBER OF PAIRS   = ',5X,I6)
 6050 FORMAT( /,7X,'STATISTICAL VALUE = ',D15.8)
 6060 FORMAT( /,7X,'NORMALIZED VALUE  = ',D15.8)
 6070 FORMAT( /,7X,'PROBABILITY = ',D15.8,2X,&
                    'NOT SIGNIFICANT AT 0.05 LEVEL')
      END
```

(d) Output results

```
*** D5TEWL ***

  ** INPUT **

      N =      14

      OBSERVATIONS

          NO.          A            B
        -------------------------------
           1        28.00        36.00
           2        96.00        24.00
           3        37.00        47.00
           4        34.00        73.00
           5        85.00        15.00
           6        56.00        34.00
```

```
         7        67.00        80.00
         8        56.00        82.00
         9        19.00        78.00
        10        27.00        41.00
        11        61.00        56.00
        12        76.00        32.00
        13        13.00        31.00
        14        43.00        41.00

** OUTPUT **

    IERR =      0

    NUMBER OF PAIRS   =          14

    STATISTICAL VALUE =  0.49000000D+02

    NORMALIZED VALUE  = -0.21971769D+00

    PROBABILITY =  0.41304551D+00  NOT SIGNIFICANT AT 0.05 LEVEL
```

### 8.3.3 D5TEMH, R5TEMH Mann-Whitney's U Test

(1) **Function**

The D5TEMH or R5TEMH performs Mann-Whitney's U test for two independent samples.
Given $n$ observed values $x_i,\ (i = 1, \cdots, n)$ and $m$ observed values $y_j,\ (j = 1, \cdots, m)$, which were independently taken from two populations $\prod_1$ and $\prod_2$ having continuous distribution functions $F_1(x)$ and $F_2(x)$, this subroutine tests the null hypothesis $\mathrm{H}_0$: "$F_1(x) = F_2(x)$" against the alternative hypothesis $\mathrm{H}_1$: "$F_1(x) > F_2(x)$" or $\mathrm{H}_1$: "$F_1(x) < F_2(x)$." The value for the $n$ observed values $x_i,\ (i = 1, \cdots, n)$ and $m$ observed values $y_i,\ (j = 1, \cdots, m)$ (where $n \le m$) is defined by the equations shown below.
The following kinds of values $R_i$ and $T$ are obtained from the observed values.

$R_i$: Ranks assigned to $(n + m)$ observed values obtained by combining $x_i$ and $y_i$. The mean rank is assigned for equal ranks.

$T$: Sum of ranks assigned to $x_i$

Test statistic :

$$U = \min(U_1, U_2)$$

Here, $U_1$ and $U_2$ are defined as follows.

$$U_1 = n \cdot m + \frac{n(n+1)}{2} - T$$

$$U_2 = n \cdot m - U_1$$

Expected value of $U$ :

$$E[U] = \frac{n \cdot m}{2}$$

Variance of $U$:

- No equal ranks exist

$$V[U] = \frac{n \cdot m(n + m + 1)}{12}$$

- Equal ranks exist

$$V[U] = \frac{n \cdot m}{12(n+m)(n+m-1)}((n+m)^3 - (n+m) - \sum S)$$

Here, $S$ is defined as follows.

$$S = \sum(t^3 - t)$$

$t$ : Number of values having equal ranks according to assigned ranks

Probability :

$$P = \int_{-\infty}^{Z} \frac{1}{\sqrt{2 \cdot \pi}} e^{-\frac{x^2}{2}} dx$$

Here, $Z$ is defined as follows.

$$Z = \frac{U - E[U]}{\sqrt{V[U]}}$$

(The $Z$ for which $U$ has been normalized obeys a standard normal distribution $N(0, 1)$.)

(2) **Usage**

Double precision:

CALL D5TEMH (A, N, M, R, U, Z, P, IWK, IERR)

Single precision:

CALL R5TEMH (A, N, M, R, U, Z, P, IWK, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | N+M | Input | Of the two independent samples, enter the observed values of the one having fewer observed values first, followed by the observed values of the one having more observed values. |
| 2 | N | I | 1 | Input | Number of observed values in the sample having fewer observed values $n$ |
| 3 | M | I | 1 | Input | Number of observed values in the sample having more observed values $m$ |
| 4 | R | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | N+M | Output | Ranks assigned to observed values obtained by combining the two samples $\{R_i\}$ |
| 5 | U | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Output | Test statistic $U$ |
| 6 | Z | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Output | Value for computing the significance of U according to a normal distribution (value for which U has been normalized) $Z$ |
| 7 | P | $\left\{\begin{array}{c}D\\R\end{array}\right\}$ | 1 | Output | Probability (one-tailed test) $P$ |
| 8 | IWK | I | See Contents | Work | Work area<br>**Size:** 3×(N+M) |
| 9 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) N≥1

(b) M≥20

(c) N≤M

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Any of restriction (a) to (c) were not satisfied. | Processing is aborted. |

(6) **Notes**

(a) When this subroutine is used, the test is a one-tailed test. However, when this procedure is used for a two-tailed test, if the significance level is $\alpha$, the null hypothesis is rejected when the following relationship holds.

$$P \leq \frac{\alpha}{2}$$

(7) **Example**

(a) Problem
Perform Mann-Whitney's U test for the two independent samples given by the following observed values.

$$\{x_i\} = \{13, 12, 12, 10, 10, 10, 10, 9, 8, 8, 7, 7, 7, 7, 7, 6\}$$

and

$$\{y_i\} = \{17, 16, 15, 15, 15, 14, 14, 14, 13, 13, 13, 12, 12, 12, 12, 11, 11, 10, 10, 10, 8, 8, 6\}$$

(b) Input data
Observed values $\{x_i\}$, N=16, observed values $\{y_i\}$ and M=23.

(c) Main program

```
      PROGRAM B5TEMH
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( N = 16, M = 23 )
      DIMENSION A(N+M),R(N+M),IWK(3*(N+M))
!
      WRITE(6,6000)
      IERR = 0
      DO 100 I=1,N+M
         READ(5,*) A(I)
  100 CONTINUE
      WRITE(6,6010) N,M
      WRITE(6,6020) 'A',(A(I),I=1,N)
      WRITE(6,6020) 'B',(A(N+I),I=1,M)
      CALL D5TEMH(A,N,M,R,U,Z,P,IWK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040) U
      WRITE(6,6050) Z
      WRITE(6,6060) P
!
      STOP
 6000 FORMAT(  ' *** D5TEMH ***',/,&
             /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'N = ',I6,5X,'M = ',I6)
 6020 FORMAT( /,7X,'OBSERVATIONS ',A,/,/,&
                 5(6X,5(2X,F11.2),/))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
             /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'U-VALUE          = ',D15.8)
 6050 FORMAT( /,7X,'NORMALIZED VALUE = ',D15.8)
 6060 FORMAT( /,7X,'PROBABILITY = ',D15.8,2X,&
                 'SIGNIFICANT AT 0.05 LEVEL')
      END
```

(d) Output results

```
*** D5TEMH ***

  ** INPUT **

      N =     16     M =     23
```

```
    OBSERVATIONS  A

           13.00          12.00          12.00          10.00          10.00
           10.00          10.00           9.00           8.00           8.00
            7.00           7.00           7.00           7.00           7.00
            6.00

    OBSERVATIONS  B

           17.00          16.00          15.00          15.00          15.00
           14.00          14.00          14.00          13.00          13.00
           13.00          12.00          12.00          12.00          12.00
           11.00          11.00          10.00          10.00          10.00
            8.00           8.00           6.00

** OUTPUT **

    IERR =      0

    U-VALUE          =  0.64000000D+02

    NORMALIZED VALUE = -0.34509547D+01

    PROBABILITY =  0.27930370D-03  SIGNIFICANT AT 0.05 LEVEL
```

### 8.3.4 D5TESP, R5TESP
### Spearman's Rank Correlation Test

(1) **Function**

The D5TESP or R5TESP obtains Spearman's rank correlation coefficient and tests the correlation between two samples.

When $(X, Y)$ is assumed to be a pair of random variables and $n$ pairs of observed values $(x_i, y_i),\ (i = 1, \cdots, n)$ are given as the actual values for them, this subroutine tests the null hypothesis $\mathrm{H}_0$: "$X$ and $Y$ are independent" against the alternative hypothesis $\mathrm{H}_1$: "There is a positive correlation between $X$ and $Y$" or $\mathrm{H}_1$: "There is a negative correlation between $X$ and $Y$."

The value for the $n$ pairs of observed values $(x_i, y_i),\ (i = 1, \cdots, n)$ is defined by the equations shown below. First, the two samples are ranked individually and represented by $a_i$ and $b_i$, respectively. The mean rank is assigned for equal ranks.

Spearman's rank correlation coefficient :

- No equal ranks exist

$$r_s = 1 - \frac{6\displaystyle\sum_{i=1}^{n}(a_i - b_i)^2}{n^3 - n}$$

- Equal ranks exist

$$r_s = \frac{(n^3 - n - S_1) + (n^3 - n - S_2) - 12\displaystyle\sum_{i=1}^{n}(a_i - b_i)^2}{2\sqrt{(n^3 - n - S_1)(n^3 - n - S_2)}}$$

Here, $S_1$, which is the correction factor of the first sample, and $S_2$, which is the correction factor of the second sample, are defined as follows.

$$S_1 = \sum({t_1}^3 - t_1)$$

$$S_2 = \sum({t_2}^3 - t_2)$$

$t$ : Number of values having equal ranks according to assigned ranks

Test statistic :

$$T = r_s\sqrt{\frac{n-2}{1-{r_s}^2}}$$

(The $T$ obeys a $t$ distribution with $n - 2$ degrees of freedom.)

(2) **Usage**

Double precision:

CALL D5TESP (A, N, B, IDF, R1, R2, RS, T, IWK, IERR)

Single precision:

CALL R5TESP (A, N, B, IDF, R1, R2, RS, T, IWK, IERR)

## (3) Arguments

D:Double precision real Z:Double precision complex I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | N | Input | Observed values of sample A $\{x_i\}$ |
| 2 | N | I | 1 | Input | Number of observed values in each of the two samples $n$ |
| 3 | B | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | N | Input | Observed values of sample B $\{y_i\}$ |
| 4 | IDF | I | 1 | Output | Degrees of freedom |
| 5 | R1 | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | N | Output | Ranks assigned to observed values of sample A $\{a_i\}$ |
| 6 | R2 | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | N | Output | Ranks assigned to observed values of sample B $\{b_i\}$ |
| 7 | RS | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | 1 | Output | Spearman's rank correlation coefficient $r_s$ |
| 8 | T | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | 1 | Output | Value for computing the significance of RS according to a $t$ distribution $T$ |
| 9 | IWK | I | 3×N | Work | Work area |
| 10 | IERR | I | 1 | Output | Error indicator |

## (4) Restrictions

(a) N≥10

## (5) Error indicator

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 4000 | RS was -1 or 1. | |

(6) **Notes**

(a) The number of observed values in the two samples must be equal.

(b) The rankings R1 and R2 correspond to the samples A and B, respectively.

(c) When this subroutine is used, the test is a one-tailed test. However, when this procedure is used for a two-tailed test, if the number of degrees of freedom is $n$ and the significance level is $\alpha$, he null hypothesis is rejected when the either of the following relationships holds, based on the value of $t_0\left(n, \frac{\alpha}{2}\right)$ :

$$t \geq t_0\left(n, \frac{\alpha}{2}\right)$$

or

$$t \leq -t_0\left(n, \frac{\alpha}{2}\right)$$

(7) **Example**

(a) Problem

Obtain Spearman's rank correlation coefficient and test the correlation between the two samples for the following set of observed values of two samples.

$$\begin{aligned}\{(x_i, y_i)\} &= \{(93,53),(98,46),(76,28),(28,25),(103,65),(99,80),(98,73),\\ &\quad (71,44),(74,51),(116,82),(97,54)\}\end{aligned}$$

(b) Input data

Set of observed values $\{(x_i, y_i)\}$ and N=11.

(c) Main program

```
      PROGRAM B5TESP
!
      IMPLICIT REAL(8)(A-H,O-Z)
      PARAMETER( N = 11 )
      DIMENSION A(N),B(N),R1(N),R2(N),IWK(3*N)
!
      WRITE(6,6000)
      IERR = 0
      DO 100 I=1,N
         READ(5,*) A(I),B(I)
  100 CONTINUE
      WRITE(6,6010) N
      DO 110 I=1,N
         WRITE(6,6020) I,A(I),B(I)
  110 CONTINUE
      CALL D5TESP(A,N,B,IDF,R1,R2,RS,T,IWK,IERR)
      WRITE(6,6030) IERR
      WRITE(6,6040)
      DO 120 I=1,N
         WRITE(6,6020) I,R1(I),R2(I)
  120 CONTINUE
      WRITE(6,6050) IDF
      WRITE(6,6060) RS
      WRITE(6,6070) T
!
      STOP
 6000 FORMAT(  ' *** D5TESP ***',/,&
              /,3X,'** INPUT **')
 6010 FORMAT( /,7X,'N = ',I6,/,&
              /,7X,'OBSERVATIONS',/,&
              /,11X,'NO.',9X,'A',12X,'B',/,&
               9X,31('-'))
 6020 FORMAT(  7X,I6,2(2X,F11.2))
 6030 FORMAT( /,3X,'** OUTPUT **',/,&
              /,7X,'IERR = ',I6)
 6040 FORMAT( /,7X,'TIED RANKED',/,&
              /,11X,'NO.',9X,'A',12X,'B',/,&
               9X,31('-'))
 6050 FORMAT( /,7X,'DEGREE OF FREEDOM',21X,'= ',5X,I6)
 6060 FORMAT( /,7X,'SPEARMAN RANK CORRELATION COEFFICIENT = ',D15.8)
 6070 FORMAT( /,7X,'STATISTICAL VALUE = ',D15.8,2X,&
                    'SIGNIFICANT AT 0.05 LEVEL')
      END
```

(d) Output results

```
*** D5TESP ***
  ** INPUT **
      N =     11
      OBSERVATIONS
         NO.          A             B
      -------------------------------
          1        93.00         53.00
          2        98.00         46.00
          3        76.00         28.00
          4        28.00         25.00
          5       103.00         65.00
          6        99.00         80.00
          7        98.00         73.00
          8        71.00         44.00
          9        74.00         51.00
         10       116.00         82.00
         11        97.00         54.00
  ** OUTPUT **
      IERR =      0
      TIED RANKED
         NO.          A             B
      -------------------------------
          1         5.00          6.00
          2         7.50          4.00
          3         4.00          2.00
          4         1.00          1.00
          5        10.00          8.00
          6         9.00         10.00
          7         7.50          9.00
          8         2.00          3.00
          9         3.00          5.00
         10        11.00         11.00
         11         6.00          7.00
      DEGREE OF FREEDOM                     =           9
      SPEARMAN RANK CORRELATION COEFFICIENT =  0.86105007D+00
      STATISTICAL VALUE =  0.50797398D+01  SIGNIFICANT AT 0.05 LEVEL
```

# Chapter 9

# MULTIVARIATE ANALYSIS

## 9.1 INTRODUCTION

When several sets of various individuals are observed, an analysis of that data is called **multivariate analysis**. This library provides the following functions for performing a multivariate analysis.

- Principal Component Analysis
- Factor Analysis
- Canonical Correlation Analysis
- Discriminant Analysis
- Cluster Analysis

### 9.1.1 Explanation

(1) **Principal Component Analysis**

The objective of a principal component analysis is to explain numerous variates according to a small number of variates. Let the mean vector and variance-covariance matrix of the probability vector $\boldsymbol{x} = [x_1, \cdots, x_n]^T$ be represented by $\boldsymbol{\mu}$ and $\Sigma$, respectively. Consider the following linear combination of $\boldsymbol{x}$ according to the vector $\boldsymbol{c}$.

$$y = \boldsymbol{c}^T(\boldsymbol{x} - \boldsymbol{\mu}), \boldsymbol{c}^T\boldsymbol{c} = 1$$

The $n$ variates $y_i = \boldsymbol{c_i}^T(\boldsymbol{x} - \boldsymbol{\mu})$ that satisfy the following conditions :

- $y_1$ maximizes the variance of $y$ in relation to $\boldsymbol{c}$.
- When $y_1, \cdots, y_k$ are defined for $k < n$, $y_{k+1} = \boldsymbol{c_{k+1}}^T(\boldsymbol{x} - \boldsymbol{\mu})$ maximizes the variance of $y$ in relation to $\boldsymbol{c}$ based on $Cov(y, y_i) = 0, i = 1, \cdots, k$.

are called the $i$th principal components of $x$, and the $\boldsymbol{c_i}$ are called the principal component vectors. The $\boldsymbol{c_i}$ also constitute the orthonormal eigenvectors of $\Sigma$. That is, if the eigenvalues of $\Sigma$ are given by $\lambda_1 \leq \cdots \leq \lambda_n$, the corresponding orthonormal eigenvector is $C = [\boldsymbol{c_1}, \cdots, \boldsymbol{c_n}]$. The contribution ratio of the $i$th principal component, which is defined as follows:

$$\frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i}$$

represents the proportion of the total variation of the variance of $y_i$. Also, the cumulative contribution ratio up to the $k$th principal component is given as follows.

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i}$$

The principal component score vector of individual $j$ is defined as follows.

$$\boldsymbol{y_j} = C^T(\boldsymbol{x_j} - \boldsymbol{\mu})$$

Generally, since $\boldsymbol{\mu}$ and $\Sigma$ are unknown, the sample mean vector and sample variance-covariance matrix are used as estimates instead. Also, since the principal components are not invariant based on a scaling transformation, in practice, the principal component analysis is performed for a standardized variate in place of $\boldsymbol{x}$.

(2) **Factor Analysis**

The objective of a factor analysis is to explain the correlation relationships among several variables according to "factors" (which are fewer in number than the number of variables). Let the observed values of $m$ variates $x_i$ $(i = 1, 2, \cdots, m)$ consisting of $n$ observed values be represented by $x_{i,j}$ $(i = 1, 2, \cdots, m;\ j = 1, 2, \cdots, n)$, and let the mean and variance of each variate be represented by $\overline{x_i}$ and $\sigma_i^2$, respectively. Let the vector $\boldsymbol{z}_j = [z_{1,j}, z_{2,j}, \cdots, z_{m,j}]^T$ consist of the standardized variables $z_{i,j}$ that correspond to each variate, where the $z_{i,j}$ are defined as follows:

$$z_{i,j} = \frac{x_{i,j} - \overline{x_i}}{\sigma_i}$$

Assume that the vector $\boldsymbol{z}_j$ has the following kind of structure.

$$\boldsymbol{z}_j = F\boldsymbol{f}_j + \boldsymbol{u}_j$$

Here, $F = (a_{i,k})$ $(i = 1, 2, \cdots, m;\ k = 1, 2, \cdots, l)$ $\boldsymbol{f}_j = [f_{1,j}, f_{2,j}, \cdots, f_{l,j}]^T$, and $\boldsymbol{u}_j = [u_{1,j}, u_{2,j}, \cdots, u_{l,j}]^T$ represent the following values, respectively:

- $a_{i,k}$: Factor loading (unknown constant) related to the $k$th factor of the $i$th variable
- $f_{k,j}$: $k$th common factor score of $j$th observed value
- $u_{k,j}$: Characteristic factor score related to the $k$th variable of the $j$th observed value

At this time, the correlation coefficient matrix between the standardized variables $R$ can be decomposed as follows.

$$R = FF^T + D = R^* + D$$

Here, $D$ is a diagonal matrix whose $i$th principal diagonal term element is called the **specificity** related to the $i$th variable. The $i$th principal diagonal term element of $R^*$ is called the **communality** of the $i$th variable. Let the eigenvalues of $R^*$ be $\lambda_i$, and let the corresponding eigenvectors be $W = [\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m] = (w_{i,j})$. To form an $l$ factor model, represent the **factor matrix** $F$ as follows.

$$F = [\sqrt{\lambda_1}\boldsymbol{w}_1, \cdots, \sqrt{\lambda_l}\boldsymbol{w}_l]$$

The **factor scores** $H = [\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_n]^T$ are obtained by solving the following equations.

$$\begin{aligned} H &= ZW \\ (F^T F)W &= F \end{aligned}$$

Here, $Z = [\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_n]^T$. Generally, the factor matrix does not take a suitable form for interpreting the actual contents of each factor. Therefore, a method is considered of applying a suitable orthogonal rotation (**factor orthogonal rotation**) to the factor matrix to make the results easier to interpret. When performing an orthogonal rotation of the factor matrix, you need to determine the kind of factor structure to aim for when performing the rotation. A representative example of an orthogonal rotation criterion is the **varimax criterion**. With the **varimax method**, the factor matrix is orthogonally rotated so that the sum for all factors of the squared variance of the factor loading of each factor (varimax criterion) is maximized. The procedure is as follows.

① Obtain the communality before rotation.

$$h_i^2 = \sum_{j=1}^{k} a_{ij}^2$$

$(i = 1, 2, \cdots, m;$ (Number of variates), $\quad j = 1, 2, \cdots, k;$ (Number of factors))

$A = (a_{ij});$ Factor loading matrix

② Normalize the factor loading matrix.

$$b_{ij} = a_{ij}/\sqrt{h_i^2} \quad (i = 1, 2, \cdots, m; \quad j = 1, 2, \cdots, k)$$

③ Perform an orthogonal rotation that maximizes the variance of the factor loading matrix, which is shown below.

$$V_c = \sum_{j=1}^{k} \frac{m\sum_{i=1}^{m}(b_{ij}^2)^2 - (\sum_{i=1}^{m} b_{ij}^2)^2}{m^2} \quad (c = 1, 2, \cdots, r(\text{Maximum number of iterations}))$$

Let the transformation matrix $T_{jj'}$ for the pair $(j, j')$ among $k$ factors be as follows.

$$
T_{jj'} = \begin{array}{cc}
\begin{array}{ccc} \quad\quad\quad\quad & \text{j} - \text{th column} \quad\quad\quad\quad & \text{j}' - \text{th column} \end{array} & \\
\left[\begin{array}{ccccccccccc}
1 & & & \vdots & & & & \vdots & & & \\
& \ddots & & \vdots & & & & \vdots & & 0 & \\
& & 1 & \vdots & & & & \vdots & & & \\
\cdots & \cdots & \cdots & \cos\theta & \cdots & \cdots & \cdots & -\sin\theta & \cdots & \cdots & \cdots \\
& & & \vdots & 1 & & & \vdots & & & \\
& & & \vdots & & \ddots & & \vdots & & & \\
& & & \vdots & & & 1 & \vdots & & & \\
\cdots & \cdots & \cdots & \sin\theta & \cdots & \cdots & \cdots & \cos\theta & \cdots & \cdots & \cdots \\
& & & \vdots & & & & \vdots & 1 & & \\
& 0 & & \vdots & & & & \vdots & & \ddots & \\
& & & \vdots & & & & \vdots & & & 1
\end{array}\right] &
\begin{array}{c} \\ \\ \\ \text{j} - \text{th row} \\ \\ \\ \\ \text{j}' - \text{th row} \\ \\ \\ \\ \end{array}
\end{array}
$$

At this time, let $u_i$, $v_i$, $p$, $q$, $r$ and $s$ be defined as follows.

$$u_i = a_{ij}/h_i, \quad v_i = a_{ij'}/h_i$$

$$p = \sum_{i=1}^{m}(u_i^2 - v_i^2)$$

$$q = 2\sum_{i=1}^{m} u_i v_i$$

$$r = \sum_{i=1}^{m}\{(u_i^2 - v_i^2)^2 - 4u_i^2 v_i^2\}$$

$$s = 4\sum_{i=1}^{m} u_i v_i (u_i^2 - v_i^2)$$

Then, $\theta$ is given by the following equation.

$$\theta = \frac{1}{4}\tan^{-1}\frac{s - \dfrac{pq}{m}}{r - \dfrac{p^2 - q^2}{m}} \quad (-\frac{\pi}{4} \leq \theta \leq \frac{\pi}{4})$$

The rotation for the combination of all factors is given as follows:

$$AT_{1,2}T_{1,3}\cdots T_{1,k}T_{2,3}\cdots T_{2,k}\cdots T_{k-1,k}$$

and $V_c$ is calculated at this time. This value is compared with $V_{c-1}$ of the previous time, and the calculation ends if the following condition is satisfied.

$$|V_c - V_{c-1}| \leq \varepsilon$$

If this condition is not satisfied, rotation is repeated. Here, $\varepsilon$ is $10^{-7}$ for single precision and $10^{-12}$ for double precision. Also $c$ denotes number of rotation iterations.

④ When the final $V_c$ is calculated, the difference of the communality for the factor matrix after rotation, which is denoted by $f_i$ as follows :

$$f_i = \sum_{j=1}^{k} a_{ij}^2 \quad (i = 1, 2, \cdots, m)$$

and the communality before rotation is calculated as follows.

$$d_i = h_i^2 - f_i$$

(3) **Canonical Correlation Analysis**

When observed values $x_{i,j}$ $(i = 1, 2, \cdots, n;\ j = 1, 2, \cdots, l+m)$ for $l+m$ variables $x_j$ $(j = 1, 2, \cdots, l+m)$ concerning $n$ subjects are given, a canonical correlation analysis is a method in which these $l+m$ variables are divided into two groups according to some criteria, and the relationships between these two groups are analyzed according to the canonical correlation coefficients. Let the mean and variance of the observed values for each variable be represented by $\overline{x_j}$ and $\sigma_j^2$, respectively. The standardized results $u_{i,j}$ corresponding to each variable are defined as follows.

$$u_{i,j} = \frac{x_{i,j} - \overline{x_j}}{\sigma_j}$$

Now, assume that the (standardized) observed values are divided into the following two groups.
First group: $u_{i,j}(i = 1, 2, \cdots, n; j = 1, 2, \cdots, l)$
Second group: $u_{i,j}(i = 1, 2, \cdots, n; j = l+1, 2, \cdots, l+m)$
At this time, consider the composite variables $z$ and $w$ determined as follows by the values.

$$\begin{aligned} z_i &= \textstyle\sum_{j=1}^{l} p_i u_{i,j} \\ w_i &= \textstyle\sum_{j=l+1}^{l+m} q_i u_{i,j} \end{aligned}$$

The coefficients $p_i$ and $q_i$ are determined so that the values of the variances $\sigma_z^2$ and $\sigma_w^2$ of the composite variables $z$ and $w$, respectively, are 1. At this time, the composite variables $z$ and $w$ are called **canonical variates**, and the correlation coefficients $\rho_{zw}$ of $z$ and $w$ are called the **canonical correlation coefficients**. Since $\sigma_z^2 = \sigma_w^2 = 1$, the following relationship holds.

$$\rho_{zw} = \sigma_{zw} = \sum_{i=1}^{n} \frac{z_i w_i}{n}$$

Here, $\sigma_{zw}$ is the covariance of the composite variables $z$ and $w$. A canonical correlation analysis determines the coefficients $\boldsymbol{p} = \{p_i\}$ and $\boldsymbol{q} = \{q_i\}$ so that the values of the canonical correlation coefficients are maximized and views the strength of the relationship between the two groups that were divided according to the canonical correlation coefficients at that time. Now, let the correlation coefficient matrix of the first group be $S$ (size: $l \times l$), the correlation coefficient matrix of the second group be $T$ (size: $m \times m$), and the correlation coefficient matrix of the first and second groups be $R$ (size: $l \times m$). For indeterminate constants $\lambda$ and $\mu$ the following equations are obtained by using Lagrange's method of indeterminate coefficients.

$$\begin{aligned} R\boldsymbol{q} &= \lambda S\boldsymbol{p} \\ R^T\boldsymbol{p} &= \mu T\boldsymbol{q} \end{aligned}$$

Now, from the following relationships:

$$\begin{aligned} \sigma_z^2 &= \boldsymbol{p}^T S\boldsymbol{p} = 1 \\ \sigma_w^2 &= \boldsymbol{q}^T T\boldsymbol{q} = 1 \end{aligned}$$

we find that:

$$\begin{aligned}\lambda &= \mu = \rho_{zw}\\ S^{-1}RT^{-1}R^T\boldsymbol{p} &= \lambda^2\boldsymbol{p}\\ \boldsymbol{q} &= \lambda^{-1}T^{-1}R^T\boldsymbol{p}\end{aligned}$$

and $\lambda$, that is $\rho_{zw}$, is obtained from the maximum eigenvalues of the matrix $S^{-1}RT^{-1}R^T$. Now, $\boldsymbol{p}$ and $\boldsymbol{q}$ should be determined so that the following relationships are satisfied among the corresponding eigenvectors.

$$\begin{aligned}&\boldsymbol{p}^TS\boldsymbol{p} = 1\\ &\boldsymbol{q}^TT\boldsymbol{q} = 1\end{aligned}$$

The number of nonzero canonical correlation coefficients is said to be the number of dimensions of a canonical variate. The number of dimensions can be determined by sequentially performing hypothesis tests that test the following null hypothesis:

$$H_k : \lambda_{k+1} = \cdots = \lambda_l = 0$$

against the following alternative hypothesis:

$$K_k : H_k \text{ is not true}$$

That is, if $H_0, \cdots, H_{k-1}$ are rejected and $H_k$ is adopted, the number of dimensions is assumed to be $k$. The test is performed by using the following fact. If Wilks' $\Lambda$, which is defined by the following equation, is used:

$$\Lambda_k = \prod_{i=k+1}^{l} (1 - {\lambda_i}^2)$$

based on hypothesis $H_k$, the following $\chi_k^2$ values:

$$\chi_k^2 = -\{n - 0.5(l + m + 1)\} \log_e \Lambda_k$$

asymptotically obey a $\chi^2$ distribution with $(l - k)(m - k)$ degrees of freedom.

(4) **Discriminant Analysis**

A discriminant analysis deals with the problem of discriminating the population to which a certain individual belongs among the $k$ populations $\pi_1, \cdots, \pi_k$ based on observed values for that individual. A prerequisite of this problem is that the individual for which the population is to be discriminated belongs to some population among $\pi_1, \cdots, \pi_k$. When the various populations are normal populations of order $p$ represented by $N(\boldsymbol{\nu_1}, \Sigma), \cdots, N(\boldsymbol{\nu_k}, \Sigma)$, the following kind of linear function (**linear discriminant function**):

$$y^{(i)}(\boldsymbol{x}) = \boldsymbol{\nu_i}^T\Sigma^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{\nu_i}^T\Sigma^{-1}\boldsymbol{\nu_i}$$

can be used to discriminate:

$$\max_j y^j(\boldsymbol{x}) = y^{j_m}(\boldsymbol{x}) \Rightarrow \boldsymbol{x} \in \pi_{j_m}$$

When the parameters are unknown, these estimates can be used for discrimination.

(5) **Cluster Analysis**

From a statistical data analysis viewpoint, classification means to provide a scale for representing the similarities and differences seen among the classification subjects and dividing those subjects into several groups (clusters) according to that scale. With this meaning, classification can be called a **clustering method** or **cluster generation method**. Cluster analysis uses an (individual)× (variate) multivariate characteristic value data matrix for the data handled as the classification subjects. At this time, both the individuals and the variates are treated as classification subjects. In either case, the cluster generation process requires a measure for representing the similarities or differences of the classification subjects. This measure is called the **similarity measure** or **dissimilarity measure**. If individuals or variates having $n$ characteristics are to be used as classification subjects when the (individual)× (variate) multivariate characteristic value data matrix $(a_{ik})$ or $(a_{ki})$ $(i = 1, 2, \cdots, n; k = 1, 2, \cdots, p)$ is given, the following measures, for example, can be used as the dissimilarity measure $d_{ij}$ $(i, j = 1, 2, \cdots, n)$.

- Euclidean quadratic distance

$$d_{ij} = \sum_{k=1}^{p} (a_{ik} - a_{jk})^2 \ (i, j = 1, \cdots, n)$$

- Standardized Euclidean quadratic distance

$$d_{ij} = \sum_{k=1}^{p} \frac{(a_{ik} - a_{jk})^2}{s_k^2} \ (i, j = 1, \cdots, n)$$

  Here, $s_k^2$, which is the variance of the variate, is defined by the following equation.

$$s_k^2 = \frac{1}{n-1} \sum_{l=1}^{n} (a_{lk} - \bar{a_k})^2 (\bar{a_k} = \frac{1}{n} \sum_{l=1}^{n} a_{lk})$$

  This is the same as obtaining the Euclidean quadratic distance when the variance of each variate is standardized to 1.

- Generalized distance of Mahalanobis

$$d_{ij} = \sum_{k=1}^{p} \sum_{m=1}^{p} (a_{ik} - a_{jk}) v_{km} (a_{im} - a_{jm}) \ (i, j = 1, \cdots, n)$$

  Here, $v_{km}$ is the $(k, m)$ element of the inverse matrix of the variance-covariance matrix of the individuals and variates.

- Minkowski distance

$$d_{ij} = \{\sum_{k=1}^{p} \mid a_{ik} - a_{jk} \mid^r\}^{1/r} \ (r \geq 1.0; i, j = 1, \cdots, n)$$

Also, to perform the classification, a measure must be defined for use when individuals or variates are merged as a cluster. When a new cluster $t$ is created by merging cluster $p$ and cluster $q$, the following measures are used as the dissimilarity measure $d_{tr}$ between cluster $t$ and a separate arbitrary cluster $r$. Here, $n_p$ represents the number of subjects contained in cluster $p$.

- Nearest neighbor method

$$d_{tr} = \min(d_{pr}, d_{qr})$$

- Furtherest neighbor method

$$d_{tr} = \max(d_{pr}, d_{qr})$$

- Group mean method

$$d_{tr} = (n_p d_{pr} + n_q d_{qr})/(n_p + n_q)$$

- Center of gravity method

$$d_{tr} = \frac{n_p}{n_p + n_q} d_{pr} + \frac{n_q}{n_p + n_q} d_{qr} - \frac{n_p n_q}{(n_p + n_q)^2} d_{pq}$$

- Median method

$$d_{tr} = \frac{1}{2} d_{pr} + \frac{1}{2} d_{qr} - \frac{1}{4} d_{pq}$$

- Ward's method

$$d_{tr} = \frac{n_p + n_r}{n_t + n_r} d_{pr} + \frac{n_q + n_r}{n_t + n_r} d_{qr} - \frac{n_r}{n_t + n_r} d_{pq}$$

- Variable method

$$d_{tr} = \frac{1 - \beta}{2} d_{pr} + \frac{1 - \beta}{2} d_{qr} + \beta d_{pq} \ (-\frac{1}{4} \leq \beta \leq 0)$$

The center of gravity method, median method, and Ward's method assume that the dissimilarity measure is given by the (standardized) Euclidean quadratic distance.

### 9.1.2 Reference Bibliography

(1) Anderson, T. W. , "An Introduction to Multivariate Statistical Analysis", John Wiley & Sons, New York (1958)

## 9.2 PRINCIPAL COMPONENT ANALYSIS

### 9.2.1 D6CPCC, R6CPCC
### Principal Component Cumulative Contribution Ratio

(1) **Function**

Let the mean vector and variance-covariance matrix of the probability vector $\boldsymbol{x} = [x_1, \cdots, x_m]^T$ consisting of $m$ variates be represented by $\boldsymbol{\mu}$ and $\Sigma$, respectively. If the eigenvalues of $\Sigma$ are given by $\lambda_1 \leq \cdots \leq \lambda_m$, and the corresponding orthonormal eigenvector is given by $C = [\boldsymbol{c_1}, \cdots, \boldsymbol{c_m}]$, the contribution ratio of the $i$th principal component, which is defined as follows:

$$\frac{\lambda_i}{\displaystyle\sum_{i=1}^{m} \lambda_i}$$

represents the proportion of the total variation of the variance of $y_i$. Also, the cumulative contribution ratio up to the $k$th principal component is given as follows.

$$c_k = \frac{\displaystyle\sum_{i=1}^{k} \lambda_i}{\displaystyle\sum_{i=1}^{m} \lambda_i}$$

When the eigenvalues $\lambda_1 \leq \cdots \leq \lambda_m$ are given, and the given criterion value is represented by $s$, the D6CPCC or R6CPCC obtains the minimum $k = k_m$ value and the cumulative contribution ratio $c_k$ $(k = 1, 2, \cdots, k_m)$ that satisfies the following relationship.

$$c_k \geq s$$

(2) **Usage**

Double precision:

CALL D6CPCC (A, M, CONS, CP, NUM, IERR)

Single precision:

CALL R6CPCC (A, M, CONS, CP, NUM, IERR)

### (3) Arguments

D:Double precision real  Z:Double precision complex  I: {INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer}
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | {D / R} | M | Input | Eigenvalues $\lambda_i$ (See Note (a)) |
| 2 | M | I | 1 | Input | Number of variates $m$ |
| 3 | CONS | I | 1 | Input | Criterion value $s$ |
| 4 | CP | {D / R} | M | Output | Cumulative contribution ratio ck values $c_k$ ($i = 1, 2, \cdots, k_m$) |
| 5 | NUM | I | 1 | Output | $k_m$ value |
| 6 | IERR | I | 1 | Output | Error indicator |

### (4) Restrictions

(a) $\mathrm{M} \geq 1$

### (5) Error indicator

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |

### (6) Notes

(a) The eigenvalues must be arranged in ascending order.

### 9.2.2 D6CPSC, R6CPSC Principal Component Scores

(1) **Function**

Let the mean vector and variance-covariance matrix of the probability vector $\boldsymbol{x} = [x_1, \cdots, x_m]^T$ consisting of $m$ variates be represented by $\boldsymbol{\mu}$ and $\Sigma$, respectively. If the eigenvalues of $\Sigma$ are given by $\lambda_1 \leq \cdots \leq \lambda_m$, and the corresponding orthonormal eigenvector is given by $C = [\boldsymbol{c_1}, \cdots, \boldsymbol{c_m}]$, the principal component score vector of individual $j$ is defined as follows.

$$\boldsymbol{y_l} = C^T(\boldsymbol{x_j} - \boldsymbol{\mu}) \ (j, l = 1, 2, \cdots, m)$$

Given $n$ observed values for $m$ variates $x_{ij}$ $(i = 1, 2, \cdots, n;\ j = 1, 2, \cdots, m)$ and the mean $\mu_j$ $(j = 1, 2, \cdots, m)$ of each variate and orthonormal eigenvectors of the variance-covariance matrix $\boldsymbol{c_j} = (c_{jl})$ $(j = 1, 2, \cdots, m; l = 1, 2, \cdots, m)$, the D6CPSC or R6CPSC obtains the principal component scores of each individual. Here, $k(k \leq m)$ represents the number of principal components to be obtained. Also, the scores are obtained for standardized variates as follows, where $\sigma_j$ is the standard deviation of each variate.

$$y_{il} = \sum_{j=1}^{m} \frac{c_{jl}(x_{ij} - \mu_j)}{\sigma_j} \ (i = 1, 2, \cdots, m; l = 1, 2, \cdots, k)$$

(2) **Usage**

Double precision:
CALL D6CPSC (A, MA, M, N, NUM, X1, SD, EV, MEV, Z, IERR)
Single precision:
CALL R6CPSC (A, MA, M, N, NUM, X1, SD, EV, MEV, Z, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: { INTEGER(4) as for 32bit Integer, INTEGER(8) as for 64bit Integer }
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | {D, R} | MA,M | Input | Observation data matrix $(x_{ij})$ (See Note (a)) |
| 2 | MA | I | 1 | Input | Adjustable dimension of array A |
| 3 | M | I | 1 | Input | Number of variates $m$ |
| 4 | N | I | 1 | Input | Number of observed values $n$ |
| 5 | NUM | I | 1 | Input | Number of principal components $k$ |
| 6 | X1 | {D, R} | M | Input | Mean of each variate $\mu_j$ |
| 7 | SD | {D, R} | M | Input | Standard deviation of each variate $\sigma_j$ |
| 8 | EV | {D, R} | MEV,M | Input | Matrix composed of eigenvectors $(c_{jl})$ (See Note (a)) |
| 9 | MEV | I | 1 | Input | Adjustable dimension of array EV |
| 10 | Z | {D, R} | MA,NUM | Output | Matrix composed of principal component scores $(y_{il})$ (See Note (a)) |
| 11 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $2 \leq \mathrm{N} \leq \mathrm{MA}$

(b) $1 \leq \mathrm{M} \leq \mathrm{MEV}$

(c) $\mathrm{NUM} \geq 1$

(d) $\mathrm{SD(i)} \geq$ Unit for determining error$(\mathrm{i} = 1, \cdots, \mathrm{M})$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |

(6) **Notes**

(a) The matrices $(x_{ij})$ and $(c_{jl})$ are stored in arrays A and EV, respectively, as real matrices (two-dimensional array type). The matrix $(y_{il})$ is stored in array Z as a real matrix (two-dimensional array type). For the method of the matrix data storage, see Appendix A.

(7) **Example**

(a) Problem

Obtain the eigenvalues and eigenvectors based on the correlation coefficient matrix, and then obtain the cumulative contribution ration of the eigenvalues and the principal component scores.

(b) Input data

Observation data matrix stored in array A:

$$\begin{bmatrix} 90.0 & 91.0 & 98.0 & 90.0 \\ 92.0 & 97.0 & 94.0 & 92.0 \\ 94.0 & 93.0 & 90.0 & 97.0 \\ 97.0 & 94.0 & 95.0 & 95.0 \\ 99.0 & 105.0 & 94.0 & 106.0 \\ 102.0 & 103.0 & 103.0 & 107.0 \\ 104.0 & 95.0 & 110.0 & 110.0 \\ 105.0 & 106.0 & 107.0 & 104.0 \\ 108.0 & 109.0 & 105.0 & 100.0 \\ 109.0 & 107.0 & 104.0 & 99.0 \end{bmatrix}$$

Correlation coefficient matrix stored in array R:

$$\begin{bmatrix} 1.0000 & 0.8000 & 0.7650 & 0.6400 \\ 0.8000 & 1.0000 & 0.4500 & 0.4575 \\ 0.7650 & 0.4500 & 1.0000 & 0.5800 \\ 0.6400 & 0.4575 & 0.5800 & 1.0000 \end{bmatrix}$$

X1(1) = 100.0,
X1(2) = 100.0,
X1(3) = 100.0,
X1(4) = 100.0,
SD(1) = 6.6667,
SD(2) = 6.6667,
SD(3) = 6.6667,
SD(4) = 6.6667,
MA = 10, MEV = 4, M = 4, N = 10 and CONS = 0.80.

(c) Main program

```
      PROGRAM B6CPSC
! *** EXAMPLE OF D6CPCC,D6CPSC ***
      IMPLICIT REAL(8) (A-H,O-Z)
      PARAMETER(MEV=4,MA=10,M=4,N=10,CONS=0.8D0)
      DIMENSION A(MA,M),X1(M),SD(M),R(M,M),EV(MEV,M),E(M),Z(MA,M),&
                W1(M),CP(M)
!
      READ(5,*) ((A(I,J),I=1,N),J=1,M)
      READ(5,*) ((R(I,J),I=1,M),J=1,M)
      READ(5,*) (X1(I),I=1,M)
      READ(5,*) (SD(I),I=1,M)
      WRITE(6,2000) MA,MEV,M,N,CONS
      DO 10 I=1,N
        WRITE(6,2010) (A(I,J),J=1,M)
   10 CONTINUE
      WRITE(6,2060)
      WRITE(6,2010) (X1(I),I=1,M)
      WRITE(6,2070)
      WRITE(6,2010) (SD(I),I=1,M)
      DO 20 J=1,M
        DO 30 I=1,M
          EV(I,J)=R(I,J)
   30   CONTINUE
   20 CONTINUE
      CALL DCSMAA(EV,MEV,M,E,W1,IERR)
      WRITE(6,2020) (E(I),I=1,M)
```

```
      WRITE(6,2030)
      DO 40 I=1,M
        WRITE(6,2010) (EV(I,J),J=1,M)
   40 CONTINUE
!
      CALL D6CPCC(E,M,CONS,CP,NUM,IERR)
      WRITE(6,3000) IERR
      WRITE(6,3020) NUM,(CP(I),I=1,NUM)
      CALL D6CPSC(A,MA,M,N,NUM,X1,SD,EV,MEV,Z,IERR)
      WRITE(6,3010) IERR
      WRITE(6,3030)
      DO 50 I=1,N
        WRITE(6,2010) (Z(I,J),J=1,NUM)
   50 CONTINUE
!
      STOP
!
 2000 FORMAT(' ',/,/,' ** INPUT **',&
      /,/,10X,'MA =',I4,6X,'MEV  =',I4,7X,'  M =',I4,7X,&
       /,10X,'N  =',I4,6X,'CONS =',F7.2,&
      /,/,7X,'*DATA*')
 2010 FORMAT(9X,4F10.4)
 2020 FORMAT(' ',/,/,6X,'*EIGENVALUES*',/,9X,(7F10.4))
 2030 FORMAT(' ',/,/,6X,'*EIGENVECTORS*')
 2060 FORMAT(' ',/,/,6X,'*MEAN OF VARIABLES (X1) *')
 2070 FORMAT(' ',/,/,6X,'*STANDARD DEVIATION (SD) *')
 3000 FORMAT(' ',/,/,' ** OUTPUT(D/R6CPCC) **',/,/,'   IERR = ',I4)
 3010 FORMAT(' ',/,/,' ** OUTPUT(D/R6CPSC) **',/,/,'   IERR = ',I4)
 3020 FORMAT(' ',/,/,6X,'NUM =',I4,/,/,6X,'*CUMULATIVE RATIO (CP) *',&
      /,(7F18.4))
 3030 FORMAT(' ',/,/,6X,'*PRINCIPAL COMPONENT SCORE (Z) *')
      END
```

(d) Output results

```
** INPUT **

         MA =  10      MEV  =   4         M =   4
         N  =  10      CONS =   0.80

      *DATA*
           90.0000   91.0000   98.0000   90.0000
           92.0000   97.0000   94.0000   92.0000
           94.0000   93.0000   90.0000   97.0000
           97.0000   94.0000   95.0000   95.0000
           99.0000  105.0000   94.0000  106.0000
          102.0000  103.0000  103.0000  107.0000
          104.0000   95.0000  110.0000  110.0000
          105.0000  106.0000  107.0000  104.0000
          108.0000  109.0000  105.0000  100.0000
          109.0000  107.0000  104.0000   99.0000


     *MEAN OF VARIABLES (X1) *
          100.0000  100.0000  100.0000  100.0000


     *STANDARD DEVIATION (SD) *
            6.6667    6.6667    6.6667    6.6667


     *EIGENVALUES*
            0.0923    0.4352    0.6096    2.8630


     *EIGENVECTORS*
            0.7911    0.1574   -0.1748    0.5647
           -0.4676   -0.1613   -0.7279    0.4748
           -0.3872    0.6555    0.4236    0.4908
           -0.0747   -0.7208    0.5101    0.4634


** OUTPUT(D/R6CPCC) **

  IERR =    0


     NUM =   2

     *CUMULATIVE RATIO (CP) *
           0.7157            0.8681


** OUTPUT(D/R6CPSC) **

  IERR =    0


     *PRINCIPAL COMPONENT SCORE (Z) *
           -2.3304    0.3526
           -1.8891   -0.4561
           -1.9516    0.0566
           -1.3971    0.0334
            0.2467   -0.4419
```

```
1.0905    0.3462
1.4140    1.8416
1.6443   -0.0353
1.6868   -0.8746
1.4859   -0.8225
```

## 9.3 FACTOR ANALYSIS

### 9.3.1 D6FALD, R6FALD Factor Loading Matrix

(1) **Function**

The D6FALD or R6FALD obtains the factor loading matrix and communality (contribution ratio of principal components) based on the eigenvalues and eigenvectors.

Factor loading matrix $(a_{ij})$ :

$$a_{ij} = \sqrt{\lambda_j} v_{ij} (i = 1, 2, \cdots, m \text{ (Number of variates)}; \; j = 1, 2, \cdots, k \text{ (Number of factors)}; \; k \leq m)$$

Here, $\lambda_j$ $(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0)$ is eigenvalues and $v_{ij}$ is the $i$th component of the eigenvector for eigenvalue $\lambda_j$.

Communality

$$h_i^2 = \sum_{j=1}^{k} a_{ij}^2$$

(2) **Usage**

Double precision:

CALL D6FALD (E, M, EV, LME, NUM, FM, LMF, OC, IERR)

Single precision:

CALL R6FALD (E, M, EV, LME, NUM, FM, LMF, OC, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: {INTEGER(4) as for 32bit Integer, INTEGER(8) as for 64bit Integer}

R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | E | {D, R} | M | Input | Eigenvalues $\lambda_j$ (See Notes (a) and (b)) |
| 2 | M | I | 1 | Input | Number of eigenvalues $m$ |
| 3 | EV | {D, R} | LME, M | Input | Matrix composed of eigenvectors corresponding to each eigenvalue $(v_{ij})$ (See Note (a)) |
| 4 | LME | I | 1 | Input | Adjustable dimension of array EV |
| 5 | NUM | I | 1 | Input | Number of factors $k$ |
| 6 | FM | {D, R} | LMF, NUM | Output | Factor loading matrix $(a_{ij})$ |
| 7 | LMF | I | 1 | Input | Adjustable dimension of array FM |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 8 | OC | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | M | Output | Initial communality (contribution ratio for variates) $h_i^2$ |
| 9 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) 1≤NUM≤M≤LME, LMF

(b) E(i)(i=1, $\cdots$ M) must be arranged in ascending order.

(c) E(M−NUM+1)≥0

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |

(6) **Notes**

(a) The eigenvalues must be arranged in ascending order. Also, the eigenvectors must be arranged corresponding to the eigenvalues. Matrix composed of eigenvectors ($v_{ij}$) and factor loading matrix ($a_{ij}$) are stored in arrays EV and FM as real matrices (two-dimensional array type) (See Appendix A).

(b) The largest NUM eigenvalues and corresponding eigenvectors are used for the calculation.

### 9.3.2 D6FAVR, R6FAVR Rotation According to the Varimax Criterion

(1) **Function**

The D6FAVR or R6FAVR performs orthogonal rotations of the factor loading matrix according to the varimax criterion

Also obtain the following values.

Communality

$$h_i^2 = \sum_{j=1}^{k} a_{ij}^2 \quad (i = 1, 2, \cdots, m \text{ (Number of variates)};\ j = 1, 2, \cdots, k \text{ (Number of factors)})$$

Here, $A = (a_{ij})$ is the factor loading matrix.

Variance of factor loading matrix

$$V_c = \sum_{j=1}^{k} \frac{m \sum_{i=1}^{m} (b_{ij}^2)^2 - (\sum_{i=1}^{m} b_{ij}^2)^2}{m^2} \quad (c = 1, 2, \cdots, r \text{(Maximum number of orthogonal rotations)})$$

Here, $b_{ij}$ is defined as follows.

$$b_{ij} = a_{ij} / \sqrt{h_i^2} \quad (i = 1, 2, \cdots, m; \quad j = 1, 2, \cdots, k)$$

(2) **Usage**

Double precision:

CALL D6FAVR (FM, LMF, M, NUM, IC, COM, LMC, V, IERR)

Single precision:

CALL R6FAVR (FM, LMF, M, NUM, IC, COM, LMC, V, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: { INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer }

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | FM | {D R} | LMF, NUM | Input | Factor loading matrix before rotations (See Note (a)) |
| | | | | Output | Factor loading matrix after final rotation |
| 2 | LMF | I | 1 | Input | Adjustable dimension of array FM |
| 3 | M | I | 1 | Input | Number of variates $m$ |
| 4 | NUM | I | 1 | Input | Number of factors $k$ |
| 5 | IC | I | 1 | Input | Maximum number of orthogonal rotations (See Note (b)) |
| | | | | Output | Actual number of orthogonal rotations |
| 6 | COM | {D R} | LMC, 3 | Output | COM(i, 1):Communality before rotations<br>COM(i, 2):Communality after final rotation<br>COM(i, 3):<br>(Communality before rotations)<br>− (Communality after final rotation)<br>$(\mathrm{i} = 1, \cdots, \mathrm{M})$ |
| 7 | LMC | I | 1 | Input | Adjustable dimension of array COM |
| 8 | V | {D R} | 0:IC | Output | Variance of factor matrix for each rotation<br>V(0): Variance before rotations<br>V(i): Variance after i-th rotation<br>(i=1, ⋯, IC) |
| 9 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) 1≤NUM≤M≤LMF, LMC

(b) IC ≥ 1

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 5000 | The solution did not converge even though the assigned maximum number of orthogonal rotations was reached. | The factor loading matrix, communality, and variance at that time are returned. |

(6) **Notes**

(a) Factor loading matrix $(a_{ij})$ are stored in array FM as real matrix (two-dimensional array type) (See Appendix A).

(b) A value of approximately 50 is suitable for IC.

(7) **Example**

(a) Problem
Obtain the eigenvalues, eigenvectors, and factor loading matrix of the following correlation coefficient matrix:

$$A = \begin{bmatrix} 1.00000 & 0.80000 & 0.76500 & 0.64000 \\ 0.80000 & 1.00000 & 0.45000 & 0.45750 \\ 0.76500 & 0.45000 & 1.00000 & 0.58000 \\ 0.64000 & 0.45750 & 0.58000 & 1.00000 \end{bmatrix}$$

and performs orthogonal rotations of the factor loading matrix according to the varimax criterion and obtain the factor loading matrix after rotations and so on.

(b) Input data
Correlation coefficient matrix $A$, M=4, LME=5, NUM=2, LMF=5, IC=10 and LMC=5.

(c) Main program

```
      PROGRAM B6FAVR
! *** EXAMPLE OF D6FALD , D6FAVR ***
      IMPLICIT REAL(8) (A-H,O-Z)
      PARAMETER (M=4,LME=5,NUM=2,LMF=5,MAXIC=10,LMC=5)
      DIMENSION A(M),EV(LME,M),FM(LMF,NUM),OC(M)
      DIMENSION W1(M),COM(LMC,3),V(0:MAXIC)
!
      READ(5,*) ((EV(I,J),I=1,M),J=1,M)
      WRITE(6,1000)
      IC=MAXIC
      DO 10 I=1,M
        WRITE(6,2000) (EV(I,J),J=1,M)
   10 CONTINUE
      WRITE(6,2100) M,LME,NUM,LMF,IC,LMC
!
      WRITE(6,3000)
      CALL DCSMAA(EV,LME,M,A,W1,IERR)
      WRITE(6,3050) IERR
      WRITE(6,3100)
      WRITE(6,2000) (A(I),I=1,M)
      WRITE(6,3200)
      DO 11 I=1,M
        WRITE(6,2000) (EV(I,J),J=1,M)
   11 CONTINUE
!
      CALL D6FALD(A,M,EV,LME,NUM,FM,LMF,OC,IERR)
      WRITE(6,4000) IERR
      WRITE(6,4100)
      DO 12 I=1,M
        WRITE(6,4200) (FM(I,J),J=1,NUM)
   12 CONTINUE
      WRITE(6,4300)
      DO 13 I=1,M
        WRITE(6,4400) OC(I)
   13 CONTINUE
      CALL D6FAVR(FM,LMF,M,NUM,IC,COM,LMC,V,IERR)
      WRITE(6,5000) IERR
      WRITE(6,4100)
      DO 14 I=1,M
        WRITE(6,4200) (FM(I,J),J=1,NUM)
   14 CONTINUE
      WRITE(6,5100) IC
      DO 15 I=0,IC
        WRITE(6,5200) I,V(I)
   15 CONTINUE
      WRITE(6,5300)
      DO 16 I=1,M
        WRITE(6,5400) I,COM(I,1),COM(I,2),COM(I,3)
   16 CONTINUE
!
      STOP
!
 1000 FORMAT('  *** D6FALD , D6FAVR ***',/,/,'    ** INPUT DATA **',/,/,&
```

```
             7X,'CORRELATION MATRIX')
 2000 FORMAT(5X,4(D15.5))
 2100 FORMAT(' ',/,7X,'M   = ',I3,5X,'LME = ',I3,5X,'NUM = ',I3,&
             /,7X,'LMF = ',I3,5X,'IC  = ',I3,5X,'LMC = ',I3)
 3000 FORMAT(' ',/,4X,'** OUTPUT **')
 3050 FORMAT(' ',/,5X,'* DCSMAA *',/,/,7X,'IERR = ',I4)
 3100 FORMAT(' ',/,7X,'A(EIGEN VALUE)')
 3200 FORMAT(' ',/,7X,'EV(EIGEN VECTOR)')
 4000 FORMAT(' ',/,5X,'* D6FALD *',/,/,7X,'IERR = ',I4)
 4100 FORMAT(' ',/,7X,'FM(FACTOR LOADING MATRIX)')
 4200 FORMAT(5X,2(D15.5))
 4300 FORMAT(' ',/,7X,'OC(COMMUNALITIES)')
 4400 FORMAT(5X,D15.5)
 5000 FORMAT(' ',/,5X,'* D6FAVR *',/,/,7X,'IERR = ',I4)
 5100 FORMAT(' ',/,7X,'IC = ',I5)
 5200 FORMAT(7X,'V(',I2,') = ',D15.5)
 5300 FORMAT(' ',/,7X,'COM(COMMUNALITIES)',/,8X,'VARIABLE',4X,&
      '(ORIGINAL)',7X,'(FINAL)',5X,'(DIFFERENCE)')
 5400 FORMAT(7X,I9,3(D15.5))
!
      END
```

(d) Output results

```
 *** D6FALD , D6FAVR ***

   ** INPUT DATA **

      CORRELATION MATRIX
        0.10000D+01    0.80000D+00    0.76500D+00    0.64000D+00
        0.80000D+00    0.10000D+01    0.45000D+00    0.45750D+00
        0.76500D+00    0.45000D+00    0.10000D+01    0.58000D+00
        0.64000D+00    0.45750D+00    0.58000D+00    0.10000D+01

      M   =   4     LME =   5     NUM =   2
      LMF =   5     IC  =  10     LMC =   5

   ** OUTPUT **

    * DCSMAA *

      IERR =    0

      A(EIGEN VALUE)
        0.92284D-01    0.43517D+00    0.60958D+00    0.28630D+01

      EV(EIGEN VECTOR)
        0.79110D+00    0.15737D+00   -0.17478D+00    0.56467D+00
       -0.46764D+00   -0.16135D+00   -0.72787D+00    0.47485D+00
       -0.38715D+00    0.65553D+00    0.42364D+00    0.49084D+00
       -0.74704D-01   -0.72075D+00    0.51008D+00    0.46341D+00

    * D6FALD *

      IERR =    0

      FM(FACTOR LOADING MATRIX)
        0.95543D+00   -0.13646D+00
        0.80345D+00   -0.56829D+00
        0.83052D+00    0.33076D+00
        0.78410D+00    0.39825D+00

      OC(COMMUNALITIES)
        0.93147D+00
        0.96849D+00
        0.79917D+00
        0.77342D+00

    * D6FAVR *

      IERR =    0

      FM(FACTOR LOADING MATRIX)
        0.62188D+00   -0.73806D+00
        0.22093D+00   -0.95900D+00
        0.83984D+00   -0.30634D+00
        0.85015D+00   -0.22507D+00

      IC =     4
      V( 0) =     0.25744D-01
      V( 1) =     0.19491D+00
      V( 2) =     0.23676D+00
      V( 3) =     0.26226D+00
      V( 4) =     0.26226D+00

      COM(COMMUNALITIES)
       VARIABLE    (ORIGINAL)       (FINAL)     (DIFFERENCE)
               1    0.93147D+00    0.93147D+00    0.00000D+00
               2    0.96849D+00    0.96849D+00   -0.11102D-15
               3    0.79917D+00    0.79917D+00   -0.22204D-15
               4    0.77342D+00    0.77342D+00    0.11102D-15
```

## 9.4 CANONICAL CORRELATION ANALYSIS

### 9.4.1 D6CVAN, R6CVAN<br>Canonical Correlation Analysis

(1) **Function**

To perform a canonical correlation analysis for two groups of observed values, the D6CVAN or R6CVAN performs the following processing.
Given the correlation coefficient matrix of the first group $R_{11}$ (size: $m_1 \times m_1$),
the correlation coefficient matrix of the second group $R_{22}$ (size: $m_2 \times m_2$),
and the correlation coefficient matrix of the first and second groups $R_{12}$ (size: $m_1 \times m_2$),
solve the following eigenvalue problem to perform a correlation analysis.

$$R_{11}{}^{-1} R_{12} R_{22}{}^{-1} R_{12}{}^{T} \boldsymbol{p} = \lambda^2 \boldsymbol{p}$$
$$\boldsymbol{q} = \lambda^{-1} R_{22}{}^{-1} R_{12}{}^{T} \boldsymbol{p}$$

Define the following kind of matrix $R$ by collecting together the correlation coefficient matrices.

$$\begin{bmatrix} R_{11} & R_{12}{}^{T} \\ R_{12} & R_{22} \end{bmatrix}$$

Next, from the eigenvalues $\lambda_i{}^2$ and eigenvectors $\boldsymbol{p_i}$ and $\boldsymbol{q_i}$ that were found, obtain the canonical correlation coefficients, Wilks' $\Lambda$, as well as the canonical coefficients of each group, which are defined by the following equations.
Canonical correlation coefficients=$\lambda_i$ $(i = 1, \cdots, m)$
However, $\lambda_1 > \lambda_2 > \cdots > \lambda_m$, $m = min(m_1, m_2)$.
Wilks' $\Lambda$:

$$\Lambda_k = \prod_{i=k+1}^{l} (1 - \lambda_i{}^2)$$

Canonical coefficients of first group: $(\boldsymbol{p_1}, \boldsymbol{p_2}, \cdots, \boldsymbol{p_m})$
Canonical coefficients of second group: $(\boldsymbol{q_1}, \boldsymbol{q_2}, \cdots, \boldsymbol{q_m})$
The number of nonzero canonical correlation coefficients is said to be the number of dimensions of a canonical variate. The number of dimensions can be determined by sequentially performing hypothesis tests that test the following null hypothesis:

$$H_k : \lambda_{k+1} = \cdots = \lambda_l = 0$$

against the following alternative hypothesis:

$$K_k : H_k \text{ is not true}$$

That is, if $H_0, \cdots, H_{k-1}$ are rejected and $H_k$ is adopted, the number of dimensions is assumed to be $k$. The test is performed by using the following fact. If Wilks' $\Lambda$ is used, based on hypothesis $H_k$, the following $\chi_k^2$ values:

$$\chi_k^2 = -\{n - 0.5(m_1 + m_2 + 1)\} \log_e \Lambda_k$$

asymptotically obey a $\chi^2$ distribution with $(m_1 - k)(m_2 - k)$ degrees of freedom.

(2) **Usage**

Double precision:

CALL D6CVAN (N, M1, M2, R, MR, CO, CO1, MCO1, CO2, MCO2, E, WIL, CHI, NDF, W1, IERR)

Single precision:

CALL R6CVAN (N, M1, M2, R, MR, CO, CO1, MCO1, CO2, MCO2, E, WIL, CHI, NDF, W1, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Number of observed values |
| 2 | M1 | I | 1 | Input | Number of variates of first group |
| 3 | M2 | I | 1 | Input | Number of variates of second group |
| 4 | R | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | See Contents | Input | Correlation coefficient matrix $R$<br>**Size**: $\text{MR}, (\text{M1} + \text{M2})$ |
| 5 | MR | I | 1 | Input | Adjustable dimension of array R |
| 6 | CO | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | M1 | Output | Canonical correlation coefficients |
| 7 | CO1 | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | MCO1, M1 | Output | Canonical coefficient matrix of first group |
| 8 | MCO1 | I | 1 | Input | Adjustable dimension of array CO1 |
| 9 | CO2 | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | MCO2, M1 | Output | Canonical coefficient matrix of second group |
| 10 | MCO2 | I | 1 | Input | Adjustable dimension of array CO2 |
| 11 | E | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | M1 | Output | Eigenvalues |
| 12 | WIL | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | M1 | Output | Wilks' $\Lambda$ |
| 13 | CHI | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | M1 | Output | $\chi^2$ values $\chi_k^2$ |
| 14 | NDF | I | M1 | Output | Number of degrees of freedom for $\chi^2$ test |
| 15 | W1 | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | See Contents | Work | Work area<br>**Size**: $\text{MR} \times (\text{M1} + \text{M2})$ |
| 16 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) N≥2

(b) M1≤M2

(c) M1+M2≤MR

(d) 2≤M1≤MCO1, 2≤M2≤MCO2

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 4000+$i$ | When performing an LU decomposition, the pivot was 0.0 during the processing of the $i$-th step. | |
| 5000 | The solution did not converge during the step for obtaining the eigenvalues. | Processing is aborted. |
| 6000 | The eigenvalue was smaller than the unit for determining error. | |

(6) **Notes**
None

### 9.4.2 D6CVSC, R6CVSC<br>Canonical Variate Scores

(1) **Function**

The D6CVSC or R6CVSC obtains the canonical variate values (scores) of the observed values based on the canonical coefficients (matrix). The canonical variate $z_i$ of the observed values $x_{i,j}$ $(i = 1, 2, \cdots, n;\ j = 1, 2, \cdots, m_1)$ for $m_1$ variables $x_j$ $(j = 1, 2, \cdots, m_1)$ concerning $n$ subjects is defined by the following equation.

$$z_i = \sum_{j=1}^{m} p_i \frac{x_{i,j} - \overline{x_j}}{\sigma_j}$$

Here, $\overline{x_j}$ and $\sigma_j^2$ represent the mean and variance of the observed values and $\boldsymbol{p} = \{p_i\}$ represents the canonical coefficient vector.

(2) **Usage**

Double precision:

CALL D6CVSC (N, M1, M2, A, MA, CO1, MCO1, CO2, MCO2, X1, SD, Z, IERR)

Single precision:

CALL R6CVSC (N, M1, M2, A, MA, CO1, MCO1, CO2, MCO2, X1, SD, Z, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: { INTEGER(4) as for 32bit Integer; INTEGER(8) as for 64bit Integer }
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | N | I | 1 | Input | Number of observed values |
| 2 | M1 | I | 1 | Input | Number of variates of first group |
| 3 | M2 | I | 1 | Input | Number of variates of second group |
| 4 | A | {D, R} | See Contents | Input | Observed value matrix<br>**Size**: MA, (M1 + M2) |
| 5 | MA | I | 1 | Input | Adjustable dimension of arrays A and Z |
| 6 | CO1 | {D, R} | MCO1, M1 | Input | Canonical coefficient matrix of first group |
| 7 | MCO1 | I | 1 | Input | Adjustable dimension of array CO1 |
| 8 | CO2 | {D, R} | MCO2, M1 | Input | Canonical coefficient matrix of second group |
| 9 | MCO2 | I | 1 | Input | Adjustable dimension of array CO2 |
| 10 | X1 | {D, R} | M1+M2 | Input | Mean of each variate |
| 11 | SD | {D, R} | M1+M2 | Input | Standard deviation of each variate |
| 12 | Z | {D, R} | See Contents | Output | Canonical variate values<br>**Size**: MA, (2 × M1) |
| 13 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) 2≤N≤MA

(b) 2≤M1≤MCO1, 2≤M2≤MCO2

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |

(6) **Notes**

(a) The canonical variate values are stored in array Z as the following kind of real matrix (two-dimensional array type) (See Appendix A).

$$\begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,\mathrm{M1}} & v_{1,1} & v_{1,2} & \cdots & v_{1,\mathrm{M1}} \\ u_{2,1} & \ddots & & \vdots & v_{2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ u_{\mathrm{N},1} & \cdots & \cdots & u_{\mathrm{N},\mathrm{M1}} & v_{\mathrm{N},1} & \cdots & \cdots & v_{\mathrm{N},\mathrm{M1}} \end{bmatrix}$$

Here, $u_{k,i}$ and $v_{k,j}$ have the following meanings.

$u_{k,i}$ $(i = 1, 2, \cdots, \mathrm{M1}; k = 1, 2, \cdots, \mathrm{N})$: Canonical variate of first group for $i$-th canonical correlation coefficient

$v_{k,j}$ $(j = 1, 2, \cdots, \mathrm{M1}; k = 1, 2, \cdots, \mathrm{N})$: Canonical variate of second group for $j$-th canonical correlation coefficient

(7) **Example**

(a) Problem

Obtain the canonical coefficient matrix from the following observed value matrix.

$$A = \begin{bmatrix} 90.0 & 91.0 & 95.0 & 103.0 & 75.0 \\ 97.0 & 98.0 & 98.0 & 92.0 & 76.0 \\ 93.0 & 92.0 & 97.0 & 106.0 & 77.0 \\ 99.0 & 90.0 & 99.0 & 108.0 & 78.0 \\ 102.0 & 97.0 & 101.0 & 105.0 & 79.0 \\ 100.0 & 100.0 & 100.0 & 100.0 & 80.0 \\ 103.0 & 99.0 & 103.0 & 103.0 & 81.0 \\ 106.0 & 98.0 & 101.0 & 101.0 & 82.0 \\ 111.0 & 90.0 & 99.0 & 104.0 & 83.0 \\ 108.0 & 98.0 & 103.0 & 102.0 & 84.0 \\ 104.0 & 97.0 & 105.0 & 99.0 & 85.0 \end{bmatrix}$$

Also obtain the canonical correlation coefficients and canonical variates.

(b) Input data

Observed value matrix $A$, N=11, M1=2, M2=3, MR=5, MA=11, MCO1=3 and MCO2=3.

(c) Main program

```
      PROGRAM B6CVSC
! *** EXAMPLE OF D6CVAN , D6CVSC ***
      IMPLICIT REAL(8) (A-H,O-Z)
      PARAMETER (N=11,M1=2,M2=3,MR=5,MCO1=3,MCO2=3,MA=11)
      DIMENSION A(MA,M1+M2),R(MR,M1+M2)
      DIMENSION CO(M1),E(M1),WIL(M1),CHI(M1),NDF(M1)
      DIMENSION CO1(MCO1,M1),CO2(MCO2,M1),W1(MR,M1+M2)
      DIMENSION X1(M1+M2),SD(M1+M2),Z(MA,2*M1),STAT(M1+M2,5)
!
      M=M1+M2
      READ(5,*) ((A(I,J),I=1,N),J=1,M)
      WRITE(6,1000) N,M1,M2,MR,MA,MCO1,MCO2
      WRITE(6,3000)
      DO 10 I=1,N
        WRITE(6,3100) (A(I,J),J=1,M)
   10 CONTINUE
      CALL D2BAMS(A,MA,N,M,NS,STAT,0,IERR)
      DO 20 I=1,M
        X1(I)=STAT(I,2)
        SD(I)=STAT(I,5)
   20 CONTINUE
      CALL D2CCMT(A,MA,N,M,NS,X1,R,MR,0,W1,IERR)
      WRITE(6,3200) (X1(I),I=1,M)
      WRITE(6,3300) (SD(I),I=1,M)
```

```
      WRITE(6,3400)
      DO 30 I=1,M
        WRITE(6,3100) (R(I,J),J=1,M)
   30 CONTINUE
      CALL D6CVAN&
      (N,M1,M2,R,MR,CO,CO1,MCO1,CO2,MCO2,E,WIL,CHI,NDF,W1,IERR)
      WRITE(6,5000) IERR
      WRITE(6,4100)
      DO 40 I=1,M1
        WRITE(6,4200) CO(I),E(I),WIL(I),CHI(I),NDF(I)
   40 CONTINUE
      WRITE(6,4300)
      DO 50 J=1,M1
        WRITE(6,4700) (CO1(I,J),I=1,M1)
   50 CONTINUE
      WRITE(6,4400)
      DO 60 J=1,M1
        WRITE(6,4800) (CO2(I,J),I=1,M2)
   60 CONTINUE
!
      CALL D6CVSC(N,M1,M2,A,MA,CO1,MCO1,CO2,MCO2,X1,SD,Z,IERR)
      WRITE(6,5500) IERR
      WRITE(6,4500)
      DO 70 I=1,N
        WRITE(6,4700) (Z(I,J),J=1,M1)
   70 CONTINUE
      WRITE(6,4600)
      DO 80 I=1,N
        WRITE(6,4700) (Z(I,J),J=M1+1,2*M1)
   80 CONTINUE
!
      STOP
!
 1000 FORMAT('  *** D6CVAN , D6CVSC ***',/,/,'    ** INPUT DATA **',/,/,&
             7X,'N    = ',I4,5X,'M1   = ',I4,5X,'M2 =   ',I4,/,&
             7X,'MR   = ',I4,5X,'MA   = ',I4,/,&
             7X,'MCO1 = ',I4,5X,'MCO2 = ',I4)
 3000 FORMAT(' ',/,7X,'A(OBSERVATIONS MATRIX)')
 3100 FORMAT(7X,5(D12.4,3X))
 3200 FORMAT(' ',/,7X,'X1(MEAN OF VARIABLES)',&
             /,7X,5(D12.4,3X))
 3300 FORMAT(' ',/,7X,'SD(STANDARD DEVIATION)',&
             /,7X,5(D12.4,3X))
 3400 FORMAT(' ',/,7X,'R(CORRELATION MATRIX)')
 4100 FORMAT(13X,'CO',13X,'E',13X,'WIL',&
              12X,'CHI',8X,'NDF')
 4200 FORMAT(7X,3(D12.4,3X),D12.4,3X,I3)
 4300 FORMAT(' ',/,7X,'CO1(CANONICAL COEFFICIENT OF THE FIRST SET)')
 4400 FORMAT(' ',/,7X,'CO2(CANONICAL COEFFICIENT OF THE SECOND SET)')
 4500 FORMAT(' ',/,7X,'Z(CANONICAL SCORE OF THE FIRST SET)')
 4600 FORMAT(' ',/,7X,'Z(CANONICAL SCORE OF THE SECOND SET)')
 4700 FORMAT(7X,D12.4,3X,D12.4)
 4800 FORMAT(7X,2(D12.4,3X),D12.4)
 5000 FORMAT(' ',/,4X,'** OUTPUT **',/,/,5X,'* D6CVAN *',/,/,7X,'IERR = ',I4)
 5500 FORMAT(' ',/,5X,'* D6CVSC *',/,/,7X,'IERR = ',I4)
!
      END
```

(d) Output results

```
  *** D6CVAN , D6CVSC ***

    ** INPUT DATA **

       N    =   11     M1   =    2     M2 =      3
       MR   =    5     MA   =   11
       MCO1 =    3     MCO2 =    3

       A(OBSERVATIONS MATRIX)
         0.9000D+02     0.9100D+02     0.9500D+02     0.1030D+03     0.7500D+02
         0.9700D+02     0.9800D+02     0.9800D+02     0.9200D+02     0.7600D+02
         0.9300D+02     0.9200D+02     0.9700D+02     0.1060D+03     0.7700D+02
         0.9900D+02     0.9000D+02     0.9900D+02     0.1080D+03     0.7800D+02
         0.1020D+03     0.9700D+02     0.1010D+03     0.1050D+03     0.7900D+02
         0.1000D+03     0.1000D+03     0.1000D+03     0.1000D+03     0.8000D+02
         0.1030D+03     0.9900D+02     0.1030D+03     0.1030D+03     0.8100D+02
         0.1060D+03     0.9800D+02     0.1010D+03     0.1010D+03     0.8200D+02
         0.1110D+03     0.9000D+02     0.9900D+02     0.1040D+03     0.8300D+02
         0.1080D+03     0.9800D+02     0.1030D+03     0.1020D+03     0.8400D+02
         0.1040D+03     0.9700D+02     0.1050D+03     0.9900D+02     0.8500D+02

       X1(MEAN OF VARIABLES)
         0.1012D+03     0.9545D+02     0.1001D+03     0.1021D+03     0.8000D+02

       SD(STANDARD DEVIATION)
         0.6274D+01     0.3857D+01     0.2914D+01     0.4253D+01     0.3317D+01

       R(CORRELATION MATRIX)
         0.1000D+01     0.2566D+00     0.6937D+00    -0.8176D-02     0.8794D+00
         0.2566D+00     0.1000D+01     0.6100D+00    -0.5819D+00     0.3284D+00
         0.6937D+00     0.6100D+00     0.1000D+01    -0.1218D+00     0.8485D+00
        -0.8176D-02    -0.5819D+00    -0.1218D+00     0.1000D+01    -0.1418D-01
         0.8794D+00     0.3284D+00     0.8485D+00    -0.1418D-01     0.1000D+01
```

```
** OUTPUT **

 * D6CVAN *

   IERR =    0
         CO             E              WIL             CHI         NDF
     0.8913D+00     0.7944D+00     0.6687D-01     0.2164D+02     6
     0.8214D+00     0.6748D+00     0.3252D+00     0.8985D+01     2

   CO1(CANONICAL COEFFICIENT OF THE FIRST SET)
     0.8736D+00     0.3115D+00
    -0.5543D+00     0.9866D+00

   CO2(CANONICAL COEFFICIENT OF THE SECOND SET)
     0.1592D+00    -0.1801D+00     0.8392D+00
     0.1332D+01    -0.5502D+00    -0.1337D+01

 * D6CVSC *

   IERR =    0

   Z(CANONICAL SCORE OF THE FIRST SET)
    -0.1917D+01    -0.1517D+00
    -0.3767D+00     0.1021D+01
    -0.1418D+01    -0.1609D+00
    -0.7444D+00    -0.1203D+01
     0.2388D+00     0.3231D+00
     0.2026D+00     0.1267D+01
     0.5396D+00     0.7464D+00
     0.8765D+00     0.2255D+00
     0.9265D+00    -0.2263D+01
     0.1155D+01     0.4885D-01
     0.5173D+00     0.1464D+00

   Z(CANONICAL SCORE OF THE SECOND SET)
    -0.1582D+01    -0.4291D+00
    -0.6989D+00     0.1962D+01
    -0.1093D+01    -0.7092D+00
    -0.8159D+00    -0.4568D+00
    -0.3266D+00     0.4423D+00
     0.8359D-01     0.2289D+00
     0.3734D+00     0.8090D+00
     0.6019D+00    -0.2495D+00
     0.6186D+00    -0.1955D+01
     0.1175D+01    -0.2709D+00
     0.1664D+01     0.6282D+00
```

## 9.5 DISCRIMINANT ANALYSIS

### 9.5.1 D6DAFN, R6DAFN
### Discriminant Functions

(1) **Function**

Given $g$ groups of observed values ${x_{l,i}}^{(k)}$ $(l = 1, 2, \cdots, n_k; i = 1, 2, \cdots, m; k = 1, 2, \cdots, g)$ for which $\boldsymbol{\nu_k} = (\overline{x}_{\cdot i}^{(k)})$ where ${\overline{x}_{\cdot i}}^{(k)}$ is defined as follows:

$$ {\overline{x}_{\cdot i}}^{(k)} = \frac{1}{n_k} \sum_{l=1}^{n_k} x_{l,i}^{(k)} $$

is the mean vector of the variates of the $k$-th group and $\Sigma = (\sigma_{i,j})$ where $\sigma_{i,j}$ is defined as follows:

$$ \sigma_{i,j} = \frac{\displaystyle\sum_{k=1}^{g} \sum_{l=1}^{n_k} (x_{l,i}^{(k)} - {\overline{x}_{\cdot i}}^{(k)})(x_{l,j}^{(k)} - {\overline{x}_{\cdot j}}^{(k)})}{\sum_{k=1}^{g} (n_k - 1)} $$

is the variance-covariance matrix over all groups and for which the populations for the observed values are normal populations of order $m$ represented by $N(\boldsymbol{\nu_1}, \Sigma), \cdots, N(\boldsymbol{\nu_k}, \Sigma)$, the D6DAFN or R6DAFN obtains the coefficients of the linear discriminant function of $\boldsymbol{u}$ defined by the following equation.

$$ y^{(p)}(\boldsymbol{u}) = {\boldsymbol{\nu_k}}^T \Sigma^{-1} \boldsymbol{u} - \frac{1}{2} {\boldsymbol{\nu_k}}^T \Sigma^{-1} \boldsymbol{\nu_k} \quad (p = 1, 2, \cdots, g) $$

It also obtains the quantity $D^2$, which is defined by the following equation.

$$ D^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} {\sigma_{i,j}}^{-1} \sum_{l=1}^{g} n_k ({\overline{x}_{\cdot i}}^{(k)} - \overline{x}_{\cdot i})({\overline{x}_{\cdot j}}^{(k)} - \overline{x}_{\cdot j}) $$

Here, $\overline{x}_{\cdot i}$, which represents the mean of the variates over all groups, is defined by the following equation.

$$ \overline{x}_{\cdot i} = \sum_{k=1}^{g} \frac{n_k {\overline{x}_{\cdot i}}^{(k)}}{\displaystyle\sum_{k=1}^{g} n_k} $$

${\sigma_{i,j}}^{-1}$ is an element of $\Sigma^{-1}$. The subroutine also obtains the maximum value ${y_{p_m}}^{(l,k)}$ related to $p = 1, 2, \cdots, g$ of the value $y^{(p)}(\boldsymbol{u}^{(l,k)})$ of the discriminant function of ${u_i}^{(l,k)} = {x_{l,i}}^{(k)}$ $(i = 1, 2, \cdots, m)$ and the value ${p_m}^{(l,k)}$ of $p$ at that time. In addition, it obtains the maximum probability of the discriminant function, which is defined by the following equation.

$$ P^{(l,k)} = \frac{1}{\displaystyle\sum_{k=1}^{g} \exp(y^{(k)}(\boldsymbol{u}^{(l,k)}) - {y_{p_m}}^{(l,k)})} \quad (l = 1, 2, \cdots, n_k; k = 1, 2, \cdots, g) $$

(2) **Usage**

Double precision:

CALL D6DAFN (A, MA, M, N, K, X1, MX1, C, TM, DIST, CO, MCO, P, NUM, IW, W1, IERR)

Single precision:

CALL R6DAFN (A, MA, M, N, K, X1, MX1, C, TM, DIST, CO, MCO, P, NUM, IW, W1, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: $\left\{\begin{array}{l}\text{INTEGER(4) as for 32bit Integer}\\ \text{INTEGER(8) as for 64bit Integer}\end{array}\right\}$
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | MA,M | Input | Observation data matrix $(x_{l,i}{}^{(k)})$ (See Note (a)) |
| 2 | MA | I | 1 | Input | Adjustable dimension of array A |
| 3 | M | I | 1 | Input | Number of variates $m$ |
| 4 | N | I | K | Input | Number of observed values of each group $(n_k)$ |
| 5 | K | I | 1 | Input | Number of groups $g$ |
| 6 | X1 | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | MX1,K | Input | Mean of variates of each group $(\overline{x}_{\cdot i}^{(k)})$(See Note (a)) |
| | | | | Output | Discriminant function value $y^{(g)}(\boldsymbol{u}^{(i,g)})$ (See Note (a)) |
| 7 | MX1 | I | 1 | Input | Adjustable dimension of array X1 |
| 8 | C | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | MX1,M | Input | Variance-covariance matrix $\Sigma^{-1}$ |
| 9 | TM | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | M | Input | Mean of variates over all groups $\overline{x}_{\cdot i}$ |
| 10 | DIST | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | 1 | Output | $D^2$ value |
| 11 | CO | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | MCO,K | Output | Discriminant function coefficients $\boldsymbol{\nu_k}^T\Sigma^{-1}$ and $-\frac{1}{2}\boldsymbol{\nu_k}^T\Sigma^{-1}\boldsymbol{\nu_k}$ (See Note (a)) |
| 12 | MCO | I | 1 | Input | Adjustable dimension of array CO |
| 13 | P | $\left\{\begin{array}{c}\text{D}\\ \text{R}\end{array}\right\}$ | See Contents | Output | Probability $P^{(l,k)}$ $(l = 1,2,\cdots,n_k; k = 1,2,\cdots,g)$ related to the maximum discriminant function of each sample over all groups (See Note (a))<br>**Size:** N(1) $+\cdots+$ N(K) |
| 14 | NUM | I | See Contents | Output | Number $p_m{}^{(l,k)}$ $(l = 1,2,\cdots,n_k; k = 1,2,\cdots,g)$ of the discriminant function having the largest probability (See Note (a))<br>**Size:** N(1) $+\cdots+$ N(K) |
| 15 | IW | I | M | Work | Work area |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 16 | W1 | $\left\{\begin{matrix}\text{D}\\\text{R}\end{matrix}\right\}$ | M+2 | Work | Work area |
| 17 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $1 \leq \mathrm{M} \leq \mathrm{MX1}$

(b) $\mathrm{K} \geq 2$

(c) $\mathrm{N(i)} \geq 2(\mathrm{i} = 1, \cdots, \mathrm{K})$

(d) $\mathrm{N(1)} + \cdots + \mathrm{N(K)} \leq \mathrm{MA}$

(e) $\mathrm{MCO} \geq \mathrm{M} + 1$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 3040 | Restriction (e) was not satisfied. | |

(6) **Notes**

(a) Consider $g$ groups for which there are $m$ variates and $n_k$ $(k = 1, 2, \cdots, g)$ observed values of each variate and assume that each observed value is given by ${x_{l,i}}^{(k)}$ $(l = 1, 2, \cdots, n_k; i = 1, 2, \cdots, m; k = 1, 2, \cdots, g)$. The observation data is stored in array A as the following kind of real matrix (two-dimensional array type) (See Appendix A).

$$\begin{bmatrix} x_{1,1}{}^{(1)} & x_{1,2}{}^{(1)} & \cdots & x_{1,m}{}^{(1)} \\ x_{2,1}{}^{(1)} & x_{2,2}{}^{(1)} & \cdots & x_{2,m}{}^{(1)} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n_1,1}{}^{(1)} & x_{n_1,2}{}^{(1)} & \cdots & x_{n_1,m}{}^{(1)} \\ x_{1,1}{}^{(2)} & x_{1,2}{}^{(2)} & \cdots & x_{1,m}{}^{(2)} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n_2,1}{}^{(2)} & x_{n_2,2}{}^{(2)} & \cdots & x_{n_2,m}{}^{(2)} \\ \vdots & \vdots & \cdots & \vdots \\ x_{1,1}{}^{(g)} & x_{1,2}{}^{(g)} & \cdots & x_{1,m}{}^{(g)} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n_g,1}{}^{(g)} & x_{n_g,2}{}^{(g)} & \cdots & x_{n_g,m}{}^{(g)} \end{bmatrix}$$

The means of the variates of each group are stored in array X1 as the real matrix (two-dimensional array type) $E = (e_{i,k})$ $(i = 1, 2, \cdots, m; k = 1, 2, \cdots, g)$, which is defined as follows (See Appendix A).

$$e_{i,k} = \overline{x}_{\cdot i}^{(k)}$$

On output, the data corresponding to the first row of matrix $E$ is replaced by the values of the discriminant function $y^{(p)}(\boldsymbol{u}^{(n_g,g)})$ $(p = 1, 2, \cdots, g)$. Also, the discriminant function coefficients are stored in array CO as the real matrix (two-dimensional array type) $C = (c_{i,k})$ $(i = 1, 2, \cdots, m+1; k = 1, 2, \cdots, g)$, which is defined as follows (See Appendix A).

$$c_{i,k} = (\boldsymbol{\nu_k}^T \Sigma^{-1})_i \ (i = 1, 2, \cdots, m)$$

$$c_{m+1,k} = (-\frac{1}{2}\boldsymbol{\nu_k}^T \Sigma^{-1} \boldsymbol{\nu_k})$$

The input time data of arrays C and X1 used by this subroutine can be generated from array A by using the subroutine 4.3.2 $\begin{Bmatrix} \text{D2VCGR} \\ \text{R2VCGR} \end{Bmatrix}$. (Where, $\mathrm{T} = \mathrm{N}(1) + \cdots + \mathrm{N}(\mathrm{K})$)

## 9.5.2 D6DASC, R6DASC
## Discriminant Function Scores

(1) **Function**

Assume there are $g$ groups, the number of observed values consisting of $m$ variates of each group is $n_k$ ($k = 1, 2, \cdots, g$), the variance-covariance matrix over all groups is $\Sigma$, and the populations for the observed values are normal populations of order $m$ represented by $N(\boldsymbol{\nu_1}, \Sigma), \cdots, N(\boldsymbol{\nu_k}, \Sigma)$. Given the observed values ${x_{l,i}}^{(k)}$ ($l = 1, 2, \cdots, n_k; i = 1, 2, \cdots, m; k = 1, 2, \cdots, g$) of the $g$ groups and the coefficients $C = (c_{i,k})$ ($i = 1, 2, \cdots, m+1; k = 1, 2, \cdots, g$) of the linear discriminant function of the $m$ -dimensional vector $\boldsymbol{u}$ defined by the following equation:

$$y^{(p)}(\boldsymbol{u}) = {\boldsymbol{\nu_k}}^T \Sigma^{-1} \boldsymbol{u} - \frac{1}{2} {\boldsymbol{\nu_k}}^T \Sigma^{-1} \boldsymbol{\nu_k} \quad (p = 1, 2, \cdots, g; k = 1, 2, \cdots, g)$$

where the $c_{i,k}$ are defined as follows:

$$c_{i,k} = ({\boldsymbol{\nu_k}}^T \Sigma^{-1})_i \ (i = 1, 2, \cdots, m)$$

$$c_{m+1,k} = (-\frac{1}{2} {\boldsymbol{\nu_k}}^T \Sigma^{-1} \boldsymbol{\nu_k})$$

the D6DASC or R6DASC obtains the value (discriminant score) ${z_{l,i}}^{(p)} = y^{(p)}(\boldsymbol{u}^{(l,k)})$ ($p = 1, 2, \cdots, g$) of the discriminant function corresponding to each observed value ${u_i}^{(l,k)} = {x_{l,i}}^{(k)}$ ($i = 1, 2, \cdots, m$).

(2) **Usage**

Double precision:
CALL D6DASC (A, MA, M, N, K, CO, MCO, Z, IERR)
Single precision:
CALL R6DASC (A, MA, M, N, K, CO, MCO, Z, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: { INTEGER(4) as for 32bit Integer }
R:Single precision real C:Single precision complex { INTEGER(8) as for 64bit Integer }

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | {D R} | MA,M | Input | $({x_{l,i}}^{(k)})$ (See Note (a)) |
| 2 | MA | I | 1 | Input | Adjustable dimension of array A |
| 3 | M | I | 1 | Input | Number of variates $m$ |
| 4 | N | I | K | Input | Number of observed values of each group ($n_k$) |
| 5 | K | I | 1 | Input | Number of groups $g$ |
| 6 | CO | {D R} | MCO,K | Input | Discriminant function coefficient $c_{i,k}$ (See Note (a)) |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 7 | MCO | I | 1 | Input | Adjustable dimension of array CO |
| 8 | Z | $\begin{Bmatrix} \text{D} \\ \text{R} \end{Bmatrix}$ | MA,K | Output | Discriminant function score $y^{(p)}(\boldsymbol{u}^{(l,k)})$ |
| 9 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $2 \leq \text{M} + 1 \leq \text{MCO}$

(b) $\text{K} \geq 2$

(c) $\text{N(i)} \geq 2 \quad (\text{i} = 1, \cdots, \text{K})$

(d) $\text{N(1)} + \cdots + \text{N(K)} \leq \text{MA}$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |

(6) **Notes**

(a) Consider $g$ groups for which there are $m$ variates and $n_k$ $(k = 1, 2, \cdots, g)$ observed values of each variate and assume that each observed value is given by ${x_{l,i}}^{(k)}$ $(l = 1, 2, \cdots, n_k; i = 1, 2, \cdots, m; k = 1, 2, \cdots, g)$. The observation data is stored in array A as the following kind of real matrix (two-dimensional array type).

$$\begin{bmatrix} {x_{1,1}}^{(1)} & {x_{1,2}}^{(1)} & \cdots & {x_{1,m}}^{(1)} \\ {x_{2,1}}^{(1)} & {x_{2,2}}^{(1)} & \cdots & {x_{2,m}}^{(1)} \\ \vdots & \vdots & \cdots & \vdots \\ {x_{n_1,1}}^{(1)} & {x_{n_1,2}}^{(1)} & \cdots & {x_{n_1,m}}^{(1)} \\ {x_{1,1}}^{(2)} & {x_{1,2}}^{(2)} & \cdots & {x_{1,m}}^{(2)} \\ \vdots & \vdots & \cdots & \vdots \\ {x_{n_2,1}}^{(2)} & {x_{n_2,2}}^{(2)} & \cdots & {x_{n_2,m}}^{(2)} \\ \vdots & \vdots & \cdots & \vdots \\ {x_{1,1}}^{(g)} & {x_{1,2}}^{(g)} & \cdots & {x_{1,m}}^{(g)} \\ \vdots & \vdots & \cdots & \vdots \\ {x_{n_g,1}}^{(g)} & {x_{n_g,2}}^{(g)} & \cdots & {x_{n_g,m}}^{(g)} \end{bmatrix}$$

Also, the discriminant function coefficients are stored in array CO as the real matrix (two-dimensional array type) $C = (c_{i,k})$ $(i = 1, 2, \cdots, m+1; k = 1, 2, \cdots, g)$, which is defined as follows.

$$c_{i,k} = ({\boldsymbol{\nu_k}}^T \Sigma^{-1})_i \;\; (i = 1, 2, \cdots, m)$$

$$c_{m+1,k} = (-\frac{1}{2}{\boldsymbol{\nu_k}}^T \Sigma^{-1} \boldsymbol{\nu_k})$$

The values (discriminant scores) ${z_{l,i}}^{(p)} = y^{(p)}(\boldsymbol{u}^{(l,k)})$ $(p = 1, 2, \cdots, g)$ of the discriminant functions corresponding to the observed values ${u_i}^{(l,k)} = {x_{l,i}}^{(k)}$ $(i = 1, 2, \cdots, m)$ are stored in array Z as a real matrix (two-dimensional array type), which is defined as follows.

$$\begin{bmatrix} {z_{1,1}}^{(1)} & {z_{1,1}}^{(2)} & \cdots & {z_{1,1}}^{(g)} \\ {z_{2,1}}^{(1)} & {z_{2,1}}^{(2)} & \cdots & {z_{2,1}}^{(g)} \\ \vdots & \vdots & \cdots & \vdots \\ {z_{n_1,1}}^{(1)} & {z_{n_1,1}}^{(2)} & \cdots & {z_{n_1,1}}^{(g)} \\ {z_{1,2}}^{(1)} & {z_{1,2}}^{(2)} & \cdots & {z_{1,2}}^{(g)} \\ \vdots & \vdots & \cdots & \vdots \\ {z_{n_2,2}}^{(1)} & {z_{n_2,2}}^{(2)} & \cdots & {z_{n_2,2}}^{(g)} \\ \vdots & \vdots & \cdots & \vdots \\ {z_{1,g}}^{(1)} & {z_{1,g}}^{(2)} & \cdots & {z_{1,g}}^{(g)} \\ \vdots & \vdots & \cdots & \vdots \\ {z_{n_g,g}}^{(1)} & {z_{n_g,g}}^{(2)} & \cdots & {z_{n_g,g}}^{(g)} \end{bmatrix}$$

For the method to store the matrix data, see Appendix A.

(7) **Example**

(a) Problem

Read observation data consisting of three groups and obtain the variance-covariance matrix. Based on this variance-covariance matrix, obtain the generalized distance of Mahalanobis ($D^2$), discriminant function, criteria values for discriminating, and discriminant scores.

(b) Input data

Observation data matrix:

$$A = \begin{bmatrix} 10.0 & 3.0 & 7.0 \\ 11.0 & 5.0 & 8.0 \\ 12.0 & 7.0 & 6.0 \\ 14.0 & 4.0 & 9.0 \\ 17.0 & 12.0 & 8.0 \\ 18.0 & 11.0 & 6.0 \\ 18.0 & 13.0 & 7.0 \\ 11.0 & 4.0 & 11.0 \\ 12.0 & 6.0 & 12.0 \\ 13.0 & 8.0 & 10.0 \\ 15.0 & 5.0 & 6.0 \\ 18.0 & 10.0 & 13.0 \end{bmatrix}$$

N(1) = 4,
N(2) = 3,
N(3) = 5,
K = 3, M = 3, MA = 12,
MX1 = 3 and MCO = 4.

(c) Main program

```
      PROGRAM B6DASC
! *** EXAMPLE OF D6DAFN,D6DASC ***
      IMPLICIT REAL(8) (A-H,O-Z)
!
      PARAMETER(NA=12,LM=3,MX1=LM,LK=3,MCO=LM+1)
!
      DIMENSION A(NA,LM),X1(MX1,LK),TM(LM)
      DIMENSION NUM(NA),N(LK),IW(LM)
      DIMENSION C(MX1,LM),CO(MCO,LK),P(NA),Z(NA,LK),W1(LM+2)
      DIMENSION NS(LK),WK(LM*LM*LK+LM)
!
      WRITE(6,1000)
!
      READ(5,*) MA,M,K
      READ(5,*) ((A(I,J),J=1,M),I=1,MA)
      READ(5,*) (N(I),I=1,K)
!
      WRITE(6,2000) MA,M,K
      DO 10 I=1,K
        WRITE(6,2010) I,N(I)
   10 CONTINUE
!
      WRITE(6,2020)
      NT=0
      DO 30 I=1,K
        NT=NT+N(I)
        WRITE(6,2030) I,(J,J=1,M)
        DO 20 L=1,N(I)
          WRITE(6,2040) NT-N(I)+L,(A(NT-N(I)+L,J),J=1,M)
   20   CONTINUE
   30 CONTINUE
!
      ISW=0
!
      CALL D2VCGR(A,NA,M,N,K,NS,TM,X1,MX1,C,MX1,ISW,WK,IERR)
!
      WRITE(6,3000) IERR
      DO 40 J=1,K
        WRITE(6,3010) J,(I,I=1,M)
        WRITE(6,3020) (X1(I,J),I=1,M)
   40 CONTINUE
!
      WRITE(6,3030)
      WRITE(6,3040) (I,TM(I),I=1,M)
!
      WRITE(6,3050)
      DO 50 I=1,M
        WRITE(6,3060) (C(I,J),J=1,M)
   50 CONTINUE
!
      CALL D6DAFN(A,NA,M,N,K,X1,MX1,C,TM,DIST,CO,MCO,P,NUM,&
                  IW,W1,IERR)
!
      WRITE(6,4000) IERR
      WRITE(6,4010) DIST
      WRITE(6,4020)
      DO 60 I=1,M
        WRITE(6,4030) I,(CO(I,J),J=1,K)
   60 CONTINUE
      WRITE(6,4040) (CO(M+1,J),J=1,K)
!
      WRITE(6,4050)
      L1=1
      L2=N(1)
      DO 80 I=1,K
        WRITE(6,4060) I
        DO 70 J=L1,L2
          WRITE(6,4070) J,P(J),NUM(J)
   70   CONTINUE
        IF(I-K.LT.0) THEN
          L1=L1+N(I)
          L2=L2+N(I+1)
        END IF
   80 CONTINUE
!
      CALL D6DASC(A,MA,M,N,K,CO,MCO,Z,IERR)
      WRITE(6,5000) IERR
      WRITE(6,5010) (J,J=1,K)
      DO 90 I=1,NT
        WRITE(6,5020) I,(Z(I,J),J=1,K)
   90 CONTINUE
!
      STOP
 1000 FORMAT(' *** D2VCGR, D6DAFN, D6DASC ***',/)
 2000 FORMAT('  ** INPUT **',/,/,&
             '     MA=',I2,',  M =',I2,',  K =',I2,/)
 2010 FORMAT('     N(',I1,')=',I2)
 2020 FORMAT(' ',/,'     *OBSERVATION DATA*')
 2030 FORMAT(' ',/,'     (GROUP',I2,')',/,&
             12X,10I6)
 2040 FORMAT(8X,I2,3X,10F6.1)
 3000 FORMAT(' ',/,' ** OUTPUT(D2VCGR) **',/,/,&
```

```
              '      IERR = ',I4,/,/,&
              '      *MEAN OF VARIABLES (EACH GROUP)*')
 3010 FORMAT(' ',/,'      (GROUP',I2,')',/,&
              10X,7I8)
 3020 FORMAT(12X,7F8.2)
 3030 FORMAT(' ',/,'      *TOTAL MEAN OF VARIABLES*')
 3040 FORMAT(9X,'TM(',I2,') =',F9.4)
 3050 FORMAT(' ',/,'      *VARIANCE COVARIANCE MATRIX*')
 3060 FORMAT(6X,7F9.4)
 4000 FORMAT(' ',/,' ** OUTPUT(D6DAFN) **',/,/,&
              '      IERR = ',I4)
 4010 FORMAT(' ',/,'      *MAHARANOBIS DISTANCE*',/,&
              9X,'DIST =',F9.4)
 4020 FORMAT(' ',/,'      *DISCRIMINANT COEFFICIENT*')
 4030 FORMAT(15X,I2,7F9.4)
 4040 FORMAT('           CONSTANT',7F9.4)
 4050 FORMAT(' ',/,'      *EVALUATION OF CLASSIFICATION*')
 4060 FORMAT(' ',/,'      (GROUP',I2,')',/,&
              12X,' MAXIMUM PROBABILITY    MAXIMUM FUNCTION NO.')
 4070 FORMAT(8X,I2,5X,F13.5,14X,I3)
 5000 FORMAT(' ',/,' ** OUTPUT(D6DASC) **',/,/,&
              '      IERR = ',I4)
 5010 FORMAT(' ',/,'      *DISCRIMINANT SCORE*',/,&
              21X,'(  GROUP  )',/,&
              8X,7I10)
 5020 FORMAT(8X,I2,7F10.4)
      END
```

(d) Output results

```
*** D2VCGR, D6DAFN, D6DASC ***

 ** INPUT **

    MA=12,  M = 3,  K = 3

    N(1)= 4
    N(2)= 3
    N(3)= 5

    *OBSERVATION DATA*

    (GROUP 1)
                1     2     3
        1     10.0   3.0   7.0
        2     11.0   5.0   8.0
        3     12.0   7.0   6.0
        4     14.0   4.0   9.0

    (GROUP 2)
                1     2     3
        5     17.0  12.0   8.0
        6     18.0  11.0   6.0
        7     18.0  13.0   7.0

    (GROUP 3)
                1     2     3
        8     11.0   4.0  11.0
        9     12.0   6.0  12.0
       10     13.0   8.0  10.0
       11     15.0   5.0   6.0
       12     18.0  10.0  13.0

** OUTPUT(D2VCGR) **

    IERR =    0

    *MEAN OF VARIABLES (EACH GROUP)*

    (GROUP 1)
                1       2       3
             11.75    4.75    7.50

    (GROUP 2)
                1       2       3
             17.67   12.00    7.00

    (GROUP 3)
                1       2       3
             13.80    6.60   10.40

    *TOTAL MEAN OF VARIABLES*
        TM( 1) =  14.0833
        TM( 2) =   7.3333
        TM( 3) =   8.5833

    *VARIANCE COVARIANCE MATRIX*
        4.4685   2.3722   0.4333
        2.3722   3.7722   1.1444
        0.4333   1.1444   4.0222

** OUTPUT(D6DAFN) **

    IERR =    0
```

```
     *MAHARANOBIS DISTANCE*
         DIST =  39.2991

     *DISCRIMINANT COEFFICIENT*
                    1   3.1259   3.5141   3.4876
                    2  -1.2806   0.6108  -1.2193
                    3   1.8923   1.1879   2.5568
         CONSTANT -22.4189 -38.8641 -33.3364

     *EVALUATION OF CLASSIFICATION*

     (GROUP 1)
             MAXIMUM PROBABILITY    MAXIMUM FUNCTION NO.
         1           0.92156                1
         2           0.78818                1
         3           0.84896                1
         4           0.59235                3

     (GROUP 2)
             MAXIMUM PROBABILITY    MAXIMUM FUNCTION NO.
         5           0.99663                2
         6           0.99816                2
         7           0.99986                2

     (GROUP 3)
             MAXIMUM PROBABILITY    MAXIMUM FUNCTION NO.
         8           0.64966                3
         9           0.85401                3
        10           0.70803                3
        11           0.76530                1
        12           0.98186                3

 ** OUTPUT(D6DASC) **

     IERR =    0

     *DISCRIMINANT SCORE*
                     (  GROUP  )
                 1         2         3
         1   18.2436    6.4250   15.7798
         2   20.7005   12.3488   19.3856
         3   17.4806   14.7087   15.3209
         4   33.2509   23.4682   33.6247
         5   30.4912   37.7093   31.7761
         6   31.1132   38.2367   31.3694
         7   30.4442   40.6463   31.4875
         8   27.6578   15.3017   28.2754
         9   30.1147   21.2255   31.8812
        10   26.8948   23.5854   27.8165
        11   29.4194   24.0293   28.2225
        12   45.6396   45.9414   50.4865
```

## 9.6 CLUSTER ANALYSIS

### 9.6.1 D6CLDS, R6CLDS Dissimilarity Measures

(1) **Function**

If individuals or variates having $n$ characteristics are to be used as classification subjects when the (individual) $\times$ (variate) multivariate characteristic value data matrix $(a_{ik})$ or $(a_{ki})$ $(i = 1, 2, \cdots, n; k = 1, 2, \cdots, p)$ is given, the D6CLDS or R6CLDS obtains the dissimilarity measure $d_{ij}$ $(i, j = 1, 2, \cdots, n)$ between the $i$-th and $j$-th individual or variate by using the measures described below.

- Euclidean quadratic distance

$$d_{ij} = \sum_{k=1}^{p} (a_{ik} - a_{jk})^2 \ (i, j = 1, \cdots, n)$$

- Standardized Euclidean quadratic distance

$$d_{ij} = \sum_{k=1}^{p} \frac{(a_{ik} - a_{jk})^2}{s_k^2} \ (i, j = 1, \cdots, n)$$

Here, $s_k^2$, which is the variance of the variate, is defined by the following equation.

$$s_k^2 = \frac{1}{n-1} \sum_{l=1}^{n} (a_{lk} - \bar{a_k})^2 (\bar{a_k} = \frac{1}{n} \sum_{l=1}^{n} a_{lk})$$

This is the same as obtaining the Euclidean quadratic distance when the variance of each variate is standardized to 1.

- Generalized distance of Mahalanobis

$$d_{ij} = \sum_{k=1}^{p} \sum_{m=1}^{p} (a_{ik} - a_{jk}) v_{km} (a_{im} - a_{jm}) \ (i, j = 1, \cdots, n)$$

Here, $v_{km}$ is the $(k, m)$ element of the inverse matrix of the variance-covariance matrix of the individuals and variates.

- Minkowski distance

$$d_{ij} = \{\sum_{k=1}^{p} \mid a_{ik} - a_{jk} \mid^r\}^{1/r} \ (r \geq 1.0; i, j = 1, \cdots, n)$$

(2) **Usage**

Double precision:

CALL D6CLDS (A, LX, LY, NX, NY, ML, DISS, ISW, W1, IERR)

Single precision:

CALL R6CLDS (A, LX, LY, NX, NY, ML, DISS, ISW, W1, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: { INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer }
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | {D / R} | LX,LY | Input | Observation data $a_{ik}$ or $a_{ki}$ $(i = 1, 2, \cdots, n;\ k = 1, 2, \cdots, p)$ (See Note (a)) |
| 2 | LX | I | 1 | Input | Adjustable dimension of array A |
| 3 | LY | I | 1 | Input | Second dimension of array A |
| 4 | NX | I | 1 | Input | Number of rows of observation data matrix ($n$ or $p$) |
| 5 | NY | I | 1 | Input | Number of columns of observation data matrix ($p$ or $n$) |
| 6 | ML | I | 1 | Input | MAX(LX, LY) |
| 7 | DISS | {D / R} | See Contents | Input | When ISW=14 or 24 (Minkowski distance), enter the value of $r$ ($r \geq 1.0$) for DISS(1, 1). Otherwise, this setting is unnecessary.<br>**Size:**<br>When NX=$n$, the size is (ML,NX)<br>When NY=$n$, the size is (ML,NY) |
| | | | | Output | Dissimilarity measure matrix (real symmetric matrix) |
| 8 | ISW | I | 1 | Input | Dissimilarity measure calculation processing switch<br>ISW=11 or 21:<br>Euclidean quadratic distance<br>ISW=12 or 22:<br>Standardized Euclidean quadratic distance<br>ISW=13 or 23:<br>(Generalized) distance of Mahalanobis<br>ISW=14 or 24:<br>Minkowski distance<br>For ISW=1x, NX=$n$<br>For ISW=2x, NY=$n$ |
| 9 | W1 | {D / R} | See Contents | Work | Work area<br>**Size:**<br>When NX=$n$, the size is NY $\times$ (ML + 1) + 2<br>When NY=$n$, the size is NX $\times$ (ML + 1) + 2 |
| 10 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $1 < \mathrm{NX} \leq \mathrm{LX}, 1 < \mathrm{NY} \leq \mathrm{LY}$

(b) $\mathrm{ML} = \mathrm{MAX}(\mathrm{LX}, \mathrm{LY})$

(c) $\mathrm{ISW} \in \{11, 12, 13, 14, 21, 22, 23, 24\}$

(d) When ISW = 14 or 24, $\mathrm{DISS}(1, 1) \geq 1.0$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |
| 4000 | For the entered observation data, the calculation cannot be performed using the (generalized) distance of Mahalanobis. (see Section 9.1.1) | |

(6) **Notes**

(a) When the Euclidean quadratic distance or Minkowski distance is used for the dissimilarity measure calculation and the units of the observation data subjects differ, the data should be standardized in advance so that a suitable positional relationship between the subjects is reflected (see the example).

(7) **Example**

(a) Problem
Standardize the following observation data matrix:

$$A = \begin{bmatrix} 0.321 & 119 & 6 & 40 & 6 & 6 & 12 & 29 \\ 0.301 & 112 & 9 & 38 & 4 & 2 & 3 & 31 \\ 0.288 & 133 & 15 & 53 & 4 & 5 & 12 & 30 \\ 0.280 & 112 & 9 & 47 & 4 & 2 & 10 & 24 \\ 0.261 & 109 & 3 & 21 & 1 & 3 & 13 & 21 \\ 0.256 & 107 & 8 & 34 & 4 & 2 & 17 & 38 \\ 0.253 & 95 & 16 & 57 & 4 & 6 & 7 & 37 \\ 0.250 & 100 & 13 & 46 & 4 & 3 & 13 & 37 \\ 0.249 & 92 & 17 & 48 & 6 & 2 & 7 & 23 \\ 0.227 & 88 & 12 & 45 & 4 & 0 & 3 & 33 \end{bmatrix}$$

let the number of rows be the number of subjects, and obtain the dissimilarity measure matrix by using the Euclidean quadratic distance.

(b) Input data
Observation data matrix $A$, LX = 11, LY = 9, NX = 10, NY = 8, ML = 11 and ISW = 11.

(c) Main program

```
      PROGRAM B6CLDS
! *** EXAMPLE OF D6CLDS ***
      IMPLICIT REAL(8) (A-H,O-Z)
      PARAMETER(LX=11,LY=9,NX=10,NY=8,ML=11,ISW=11)
      DIMENSION A(LX,LY),DISS(ML,NX),W1(NY*(ML+1)+2),AV(NY),C(NY)
!
      READ(5,*) ((A(I,J),J=1,8),I=1,10)
      WRITE(6,1000)
      WRITE(6,1010) 'A','B','C','D','E','F','G','H'
      DO 10 I=1,NX
        WRITE(6,1020) I,(A(I,J),J=1,NY)
   10 CONTINUE
!
! *** NORMALIZATION OF DATA ***
      DO 11 J=1,NY
        SUM = 0.0D0
        DO 12 I=1,NX
          SUM = SUM + A(I,J)
   12   CONTINUE
        AV(J) = SUM/DBLE(NX)
   11 CONTINUE
      DO 40 I=1,NY
        SUM = 0.0D0
        DO 41 J=1,NX
          SUM = SUM + (A(J,I)-AV(I))**2
   41   CONTINUE
        C(I) = SUM/(NX-1.0D0)
   40 CONTINUE
      DO 20 I=1,NX
        DO 21 J=1,NY
          A(I,J) = (A(I,J)-AV(J))/SQRT(C(J))
   21   CONTINUE
   20 CONTINUE
!
! *** *** *** *** *** *** *** *** *** ***
!
      WRITE(6,1030) ISW,LX,LY,NX,NY,ML
      CALL D6CLDS(A,LX,LY,NX,NY,ML,DISS,ISW,W1,IERR)
      WRITE(6,2000) IERR
      WRITE(6,2010) (I,I=1,NX)
      DO 30 I=1,NX
        WRITE(6,2020) I,(DISS(I,J),J=1,NX)
   30 CONTINUE
!
      STOP
!
 1000 FORMAT('  *** D6CLDS ***',/,/,3X,'** INPUT **',/)
 1010 FORMAT('    I',8(4X,A,3X),/,2X,'--+',64('-'))
 1020 FORMAT(' ',I3,'I',8F8.3)
 1030 FORMAT(' ',/,3X,'ISW =',I4,/,3X,' LX =',I4,5X,' LY =',I4,/,3X,&
              ' NX =',I4,5X,' NY =',I4,5X,' ML =',I4,/,/)
 2000 FORMAT('    ** OUTPUT **',/,/,' IERR = ',I4,/)
 2010 FORMAT('    I',10(2X,I2,3X),/,2X,'--+',70('-'))
 2020 FORMAT(' ',I3,'I',10F7.3)
!
      END
```

(d) Output results

```
 *** D6CLDS ***

  ** INPUT **

   I    A       B       C       D       E       F       G       H
  --+----------------------------------------------------------------
   1I   0.321 119.000   6.000  40.000   6.000   6.000  12.000  29.000
   2I   0.301 112.000   9.000  38.000   4.000   2.000   3.000  31.000
   3I   0.288 133.000  15.000  53.000   4.000   5.000  12.000  30.000
   4I   0.280 112.000   9.000  47.000   4.000   2.000  10.000  24.000
   5I   0.261 109.000   3.000  21.000   1.000   3.000  13.000  21.000
   6I   0.256 107.000   8.000  34.000   4.000   2.000  17.000  38.000
   7I   0.253  95.000  16.000  57.000   4.000   6.000   7.000  37.000
   8I   0.250 100.000  13.000  46.000   4.000   3.000  13.000  37.000
   9I   0.249  92.000  17.000  48.000   6.000   2.000   7.000  23.000
  10I   0.227  88.000  12.000  45.000   4.000   0.000   3.000  33.000

  ISW =   11
   LX =   11      LY =    9
   NX =   10      NY =    8      ML =   11


   ** OUTPUT **

IERR =     0

   I   1      2      3      4      5      6      7      8      9     10
  --+----------------------------------------------------------------------
   1I  0.000 11.433 10.328 10.369 26.272 16.179 21.456 17.143 23.111 33.885
   2I 11.433  0.000 12.656  4.942 18.958 13.443 16.062 11.351 14.159 11.996
   3I 10.328 12.656  0.000  8.019 28.695 15.941 12.338 10.772 18.421 27.228
   4I 10.369  4.942  8.019  0.000 14.354 10.042 14.833  7.866  9.039 12.615
   5I 26.272 18.958 28.695 14.354  0.000 16.356 37.180 22.925 33.453 28.873
```

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 6I | 16.179 | 13.443 | 15.941 | 10.042 | 16.356 | 0.000 | 17.770 | 3.919 | 19.886 | 15.906 |
| 7I | 21.456 | 16.062 | 12.338 | 14.833 | 37.180 | 17.770 | 0.000 | 5.752 | 12.359 | 13.710 |
| 8I | 17.143 | 11.351 | 10.772 | 7.866 | 22.925 | 3.919 | 5.752 | 0.000 | 10.465 | 8.982 |
| 9I | 23.111 | 14.159 | 18.421 | 9.039 | 33.453 | 19.886 | 12.359 | 10.465 | 0.000 | 8.556 |
| 10I | 33.885 | 11.996 | 27.228 | 12.615 | 28.873 | 15.906 | 13.710 | 8.982 | 8.556 | 0.000 |

### 9.6.2 D6CLAN, R6CLAN Cluster Analysis (Input Dissimilarity Measure or Similarity Measure)

(1) **Function**

Given the dissimilarity measure $d_{ij}$ or similarity measure $1.0 - d_{ij}$ $(i, j = 1, 2, \cdots, n)$, the D6CLAN or R6CLAN performs a cluster analysis based on this measure. When a new cluster $t$ is created by merging cluster $p$ and cluster $q$, the following measures are used as the dissimilarity measure $d_{tr}$ between cluster $t$ and a separate arbitrary cluster $r$. Here, $n_p$ represents the number of subjects contained in cluster $p$.

- Nearest neighbor method

$$d_{tr} = \min(d_{pr}, d_{qr})$$

- Furtherest neighbor method

$$d_{tr} = \max(d_{pr}, d_{qr})$$

- Group mean method

$$d_{tr} = (n_p d_{pr} + n_q d_{qr})/(n_p + n_q)$$

- Center of gravity method

$$d_{tr} = \frac{n_p}{n_p + n_q} d_{pr} + \frac{n_q}{n_p + n_q} d_{qr} - \frac{n_p n_q}{(n_p + n_q)^2} d_{pq}$$

- Median method

$$d_{tr} = \frac{1}{2} d_{pr} + \frac{1}{2} d_{qr} - \frac{1}{4} d_{pq}$$

- Ward's method

$$d_{tr} = \frac{n_p + n_r}{n_t + n_r} d_{pr} + \frac{n_q + n_r}{n_t + n_r} d_{qr} - \frac{n_r}{n_t + n_r} d_{pq}$$

- Variable method

$$d_{tr} = \frac{1 - \beta}{2} d_{pr} + \frac{1 - \beta}{2} d_{qr} + \beta d_{pq} \ (-\frac{1}{4} \leq \beta \leq 0)$$

The center of gravity method, median method, and Ward's method assume that the dissimilarity measure is given by the (standardized) Euclidean quadratic distance.

(2) **Usage**

Double precision:
 CALL D6CLAN (A, LNA, NC, IPOS, LNP, DIS, ISW1, ISW2, IW, IERR)
Single precision:
 CALL R6CLAN (A, LNA, NC, IPOS, LNP, DIS, ISW1, ISW2, IW, IERR)

(3) **Arguments**

| | | |
|---|---|---|
| D:Double precision real | Z:Double precision complex | I: { INTEGER(4) as for 32bit Integer, INTEGER(8) as for 64bit Integer } |
| R:Single precision real | C:Single precision complex | |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | {D, R} | LNA,NC | Input | Dissimilarity measure or similarity measure matrix (real symmetric matrix) (two-dimensional array type, upper triangular type) (See Note (a)) |
| | | | | Output | Entered contents are not saved. |
| 2 | LNA | I | 1 | Input | Adjustable dimension of array A |
| 3 | NC | I | 1 | Input | Second dimension of array A |
| 4 | IPOS | I | LNP, 2 | Output | Information indicating cluster merging (See Note (b))<br>This information indicates that the i-th merge was a merging of the IPOS(i, 1)-th cluster and the IPOS(i, 2)-th cluster $(\mathrm{i} = 1, \cdots, \mathrm{NC} - 1)$ |
| 5 | LNP | I | 1 | Input | Adjustable dimension of array IPOS |
| 6 | DIS | {D, R} | NC − 1 | Input | When ISW2=7, enter the value of $\beta$ $(-1/4 \leq \beta \leq 0)$ for DIS(1) (Default value:$\beta = -1/4$). Otherwise, this setting is unnecessary. |
| | | | | Output | Cluster merging distance information (See Note (c))<br>This information indicates that clusters were merged using distance DIS(i) during the i-th merge $(\mathrm{i} = 1, \cdots, \mathrm{NC} - 1)$ |
| 7 | ISW1 | I | 1 | Input | Processing switch (See Note (c))<br>ISW1=1: When the dissimilarity measure matrix is given for array A<br>ISW1=2: When the similarity measure matrix is given for array A |
| 8 | ISW2 | I | 1 | Input | Analysis method selection switch<br>ISW2=1: Nearest neighbor method<br>ISW2=2: Furtherest neighbor method<br>ISW2=3: Group mean method<br>ISW2=4: Center of gravity method<br>ISW2=5: Median method<br>ISW2=6: Ward's method<br>ISW2=7: Variable method (input $\beta$) |
| 9 | IW | I | NC | Work | Work area |
| 10 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $1 < \mathrm{NC} \leq \mathrm{LNA}$

(b) LNP$\geq$NC$-$1

(c) ISW1=1 or ISW1=2

(d) $1 \leq$ ISW2 $\leq 7$

(e) When ISW2 $= 7$, $-1/4 \leq$ DIS(1) $\leq 0$
(Except when entering 1.0 to set the default value)

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (e) was not satisfied. | Processing is performed with the default value set. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 3020 | Restriction (c) was not satisfied. | |
| 3030 | Restriction (d) was not satisfied. | |

(6) **Notes**

(a) The matrix assigned for array A must be a real symmetric matrix. This subroutine uses only the upper triangular portion to perform the processing.

(b) Information indicating which cluster was merged with which cluster during which merge is stored in array IPOS. However, the lowest cluster number of the two clusters before merging is assigned to the merged cluster. For example, the cluster formed by merging cluster 3 and cluster 5 is named cluster 3.

(c) When the matrix $(a_{ij})$ given for array A is a similarity measure matrix, the subroutine performs the analysis after internally converting this to a dissimilarity measure matrix by using the following equation.

$$a_{ij} = 1.0 - a_{ij}$$

Therefore, in this case, the distances calculated by using the values that were converted by using the equation shown above are stored in DIS(i)($\mathrm{i} = 1, \cdots, \mathrm{NC} - 1$).

(7) **Example**

(a) Problem

Perform a cluster analysis using the following dissimilarity measure matrix as an index.

$$A = \begin{bmatrix} 0.000 & 11.432 & 10.328 & 10.369 & 26.271 & 16.179 & 21.455 & 17.142 & 23.111 & 33.883 \\ 11.432 & 0.000 & 12.656 & 4.942 & 18.957 & 13.443 & 16.062 & 11.350 & 14.159 & 11.996 \\ 10.328 & 12.656 & 0.000 & 8.019 & 28.694 & 15.942 & 12.338 & 10.771 & 18.421 & 27.228 \\ 10.369 & 4.942 & 8.019 & 0.000 & 14.353 & 10.042 & 14.833 & 7.866 & 9.040 & 12.615 \\ 26.271 & 18.957 & 28.694 & 14.353 & 0.000 & 16.356 & 37.180 & 22.924 & 33.453 & 28.872 \\ 16.179 & 13.443 & 15.942 & 10.042 & 16.356 & 0.000 & 17.771 & 3.919 & 19.886 & 15.906 \\ 21.455 & 16.062 & 12.338 & 14.833 & 37.180 & 17.771 & 0.000 & 5.752 & 12.359 & 13.710 \\ 17.142 & 11.350 & 10.771 & 7.866 & 22.924 & 3.919 & 5.752 & 0.000 & 10.465 & 8.982 \\ 23.111 & 14.159 & 18.421 & 9.040 & 33.453 & 19.886 & 12.359 & 10.465 & 0.000 & 8.556 \\ 33.883 & 11.996 & 27.228 & 12.615 & 28.872 & 15.906 & 13.710 & 8.982 & 8.556 & 0.000 \end{bmatrix}$$

(b) Input data

Dissimilarity measure matrix $A$, ISW1 = 1, ISW2 = 3, LNA = 11, NC = 10 and LNP = 9.

(c) Main program

```
      PROGRAM B6CLAN
! *** EXAMPLE OF D6CLAN ***
      IMPLICIT REAL(8) (A-H,O-Z)
      PARAMETER(LNA=11,NC=10,LNP=9,ISW1=1,ISW2=3)
      DIMENSION A(LNA,NC),DIS(NC-1),IPOS(LNP,2),IW(NC)
!
      READ(5,*) ((A(I,J),J=1,NC),I=1,NC)
      WRITE(6,1000)
      WRITE(6,1010) (I,I=1,NC)
      DO 10 I=1,NC
        WRITE(6,1020) I,(A(I,J),J=1,NC)
   10 CONTINUE
!
      WRITE(6,1030) ISW1,ISW2,LNA,NC,LNP
      CALL D6CLAN(A,LNA,NC,IPOS,LNP,DIS,ISW1,ISW2,IW,IERR)
      WRITE(6,2000) IERR
!
      DO 20 I=1,NC-1
        WRITE(6,2010) IPOS(I,1),IPOS(I,2),DIS(I)
   20 CONTINUE
!
      STOP
!
 1000 FORMAT('  *** D6CLAN ***',/,/,3X,'** INPUT **',/)
 1010 FORMAT('    I',10(I4,3X),/,2X,'--+',70('-'))
 1020 FORMAT(' ',I3,'I',10F7.3)
 1030 FORMAT(' ',/,3X,'ISW1 =',I4,5X,' ISW2 =',I4,/,&
              3X,' LNA =',I4,5X,' NC =',I4,5X,' LNP =',I4,/,/)
 2000 FORMAT('    ** OUTPUT **',/,/,' IERR = ',I4,/)
 2010 FORMAT(11X,I3,' ---',I3,5X,F8.4)
!
      END
```

(d) Output results

```
 *** D6CLAN ***

  ** INPUT **

   I   1      2      3      4      5      6      7      8      9     10
  --+----------------------------------------------------------------------
  1I  0.000 11.432 10.328 10.369 26.271 16.179 21.455 17.142 23.111 33.883
  2I 11.432  0.000 12.656  4.942 18.957 13.443 16.062 11.350 14.159 11.996
  3I 10.328 12.656  0.000  8.019 28.694 15.942 12.338 10.771 18.421 27.228
  4I 10.369  4.942  8.019  0.000 14.353 10.042 14.833  7.866  9.040 12.615
  5I 26.271 18.957 28.694 14.353  0.000 16.356 37.180 22.924 33.453 28.872
  6I 16.179 13.443 15.942 10.042 16.356  0.000 17.771  3.919 19.886 15.906
  7I 21.455 16.062 12.338 14.833 37.180 17.771  0.000  5.752 12.359 13.710
  8I 17.142 11.350 10.771  7.866 22.924  3.919  5.752  0.000 10.465  8.982
  9I 23.111 14.159 18.421  9.040 33.453 19.886 12.359 10.465  0.000  8.556
 10I 33.883 11.996 27.228 12.615 28.872 15.906 13.710  8.982  8.556  0.000

  ISW1 =    1      ISW2 =    3
   LNA =   11      NC =   10      LNP =    9


   ** OUTPUT **

IERR =    0

            6 ---  8       3.9190
            2 ---  4       4.9420
            9 --- 10       8.5560
```

```
1 ---  3      10.3280
1 ---  2      10.6190
6 ---  7      11.7615
6 ---  9      13.5513
1 ---  6      15.8938
1 ---  5      25.2289
```

### 9.6.3 D6CLDA, R6CLDA Cluster Analysis (Input Observation Data, Use Dissimilarity Measure)

(1) **Function**

If individuals or variates having $n$ characteristics are to be used as classification subjects when the (individual) $\times$ (variate) multivariate characteristic value data matrix $(a_{ik})$ or $(a_{ki})$ $(i = 1, 2, \cdots, n;\ k = 1, 2, \cdots, p)$ is given, the D6CLDA or R6CLDA obtains the dissimilarity measure $d_{ij}$ $(i, j = 1, 2, \cdots, n)$ between the $i$-th and $j$-th individual or variate by using the measures described below and performs a cluster analysis using this as an index.

- Euclidean quadratic distance

$$d_{ij} = \sum_{k=1}^{p} (a_{ik} - a_{jk})^2 \ (i, j = 1, \cdots, n)$$

- Standardized Euclidean quadratic distance

$$d_{ij} = \sum_{k=1}^{p} \frac{(a_{ik} - a_{jk})^2}{s_k^2} \ (i, j = 1, \cdots, n)$$

Here, $s_k^2$, which is the variance of the variate, is defined by the following equation.

$$s_k^2 = \frac{1}{n-1} \sum_{l=1}^{n} (a_{lk} - \bar{a_k})^2 (\bar{a_k} = \frac{1}{n} \sum_{l=1}^{n} a_{lk})$$

This is the same as obtaining the Euclidean quadratic distance when the variance of each variate is standardized to 1.

- Generalized distance of Mahalanobis

$$d_{ij} = \sum_{k=1}^{p} \sum_{m=1}^{p} (a_{ik} - a_{jk}) v_{km} (a_{im} - a_{jm}) \ (i, j = 1, \cdots, n)$$

Here, $v_{km}$ is the $v_{km}$ element of the inverse matrix of the variance-covariance matrix of the individuals and variates.

- Minkowski distance

$$d_{ij} = \{\sum_{k=1}^{p} \mid a_{ik} - a_{jk} \mid^r\}^{1/r} \ (r \geq 1.0; i, j = 1, \cdots, n)$$

When a new cluster $t$ is created by merging cluster $p$ and cluster $q$, the following measures are used as the dissimilarity measure $d_{tr}$ between cluster $t$ and a separate arbitrary cluster $r$. Here, $n_p$ represents the number of subjects contained in cluster $p$.

- Nearest neighbor method

$$d_{tr} = \min(d_{pr}, d_{qr})$$

- Furtherest neighbor method

$$d_{tr} = \max(d_{pr}, d_{qr})$$

- Group mean method

$$d_{tr} = (n_p d_{pr} + n_q d_{qr})/(n_p + n_q)$$

- Center of gravity method

$$d_{tr} = \frac{n_p}{n_p + n_q} d_{pr} + \frac{n_q}{n_p + n_q} d_{qr} - \frac{n_p n_q}{(n_p + n_q)^2} d_{pq}$$

- Median method

$$d_{tr} = \frac{1}{2}d_{pr} + \frac{1}{2}d_{qr} - \frac{1}{4}d_{pq}$$

- Ward's method

$$d_{tr} = \frac{n_p + n_r}{n_t + n_r}d_{pr} + \frac{n_q + n_r}{n_t + n_r}d_{qr} - \frac{n_r}{n_t + n_r}d_{pq}$$

- Variable method

$$d_{tr} = \frac{1-\beta}{2}d_{pr} + \frac{1-\beta}{2}d_{qr} + \beta d_{pq} \ (-\frac{1}{4} \leq \beta \leq 0)$$

The center of gravity method, median method, and Ward's method assume that the dissimilarity measure is given by the (standardized) Euclidean quadratic distance.

(2) **Usage**

Double precision:

CALL D6CLDA (A, LX, LY, NX, NY, ML, NC, IPOS, LNP, DIS, ISW1, ISW2, IW, W1, IERR)

Single precision:

CALL R6CLDA (A, LX, LY, NX, NY, ML, NC, IPOS, LNP, DIS, ISW1, ISW2, IW, W1, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex I: $\left\{ \begin{array}{l} \text{INTEGER(4) as for 32bit Integer} \\ \text{INTEGER(8) as for 64bit Integer} \end{array} \right\}$
R:Single precision real C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | A | $\left\{\begin{array}{c}\text{D}\\\text{R}\end{array}\right\}$ | LX,LY | Input | Observation data $a_{ik}$ or $a_{ki}$ $(i = 1, 2, \cdots, n; k = 1, 2, \cdots, p)$ (See Note (a)) |
| 2 | LX | I | 1 | Input | Adjustable dimension of array A |
| 3 | LY | I | 1 | Input | Second dimension of array A |
| 4 | NX | I | 1 | Input | Number of rows of observation data matrix ($n$ or $p$) |
| 5 | NY | I | 1 | Input | Number of columns of observation data matrix ($p$ or $n$) |
| 6 | ML | I | 1 | Input | MAX(LX, LY) |
| 7 | NC | I | 1 | Input | Number of objects $n$ (NX or NY) |
| 8 | IPOS | I | LNP, 2 | Output | Information indicating cluster merging (See Note (b)) This information indicates that the i-th merge was a merging of the $\mathrm{IPOS(i, 1)}$-th cluster and the $\mathrm{IPOS(i, 1)}$-th cluster $(\mathrm{i} = 1, \cdots, \mathrm{NC} - 1)$ |
| 9 | LNP | I | 1 | Input | Adjustable dimension of array IPOS |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 10 | DIS | $\begin{Bmatrix} \mathrm{D} \\ \mathrm{R} \end{Bmatrix}$ | NC − 1 | Input | When ISW1=14 or 24 (Minkowski distance), enter the value of $r$ ($r \geq 1.0$) for DIS(1).<br>When ISW2 = 7, enter the value of $\beta$ ($-1/4 \leq \beta \leq 0$) for DIS(2) (Default value: $\beta = -1/4$).<br>Otherwise, this setting is unnecessary. |
| | | | | Output | Cluster merging distance information.<br>This information indicates that clusters were merged using distance DIS(i) during the i-th merge ($\mathrm{i} = 1, \cdots, \mathrm{NC} - 1$). |
| 11 | ISW1 | I | 1 | Input | Dissimilarity measure calculation processing switch<br>ISW1=11 or 21:<br>Euclidean quadratic distance<br>ISW1=12 or 22:<br>Standardized Euclidean quadratic distance<br>ISW1=13 or 23:<br>(Generalized) distance of Mahalanobis<br>ISW1=14 or 24:<br>Minkowski distance<br>For ISW1=11,12,13 or 14, NX=$n$.<br>For ISW1=21,22,23 or 24, NY=$n$. |
| 12 | ISW2 | I | 1 | Input | Analysis method selection switch<br>ISW2=1: Nearest neighbor method<br>ISW2=2: Furtherest neighbor method<br>ISW2=3: Group mean method<br>ISW2=4: Center of gravity method<br>ISW2=5: Median method<br>ISW2=6: Ward's method<br>ISW2=7: Variable method (input $\beta$) |
| 13 | IW | I | NC | Work | Work area |
| 14 | W1 | $\begin{Bmatrix} \mathrm{D} \\ \mathrm{R} \end{Bmatrix}$ | See Contents | Work | Work area<br>**Size:**<br>When ISW1 =11, 12, 13 or 14<br>the size is $\mathrm{ML} \times \mathrm{NC} + \mathrm{NY} \times (\mathrm{ML} + 1) + 2$.<br>When ISW1 =21, 22, 23 or 24<br>the size is $\mathrm{ML} \times \mathrm{NC} + \mathrm{NX} \times (\mathrm{ML} + 1) + 2$. |
| 15 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $1 < \mathrm{NX} \leq \mathrm{LX}, 1 < \mathrm{NY} \leq \mathrm{LY}$

(b) NC = NX (when ISW1 =11, 12, 13 or 14) or NY (when ISW1 =21, 22, 23 or 24)

(c) $\mathrm{LNP} \geq \mathrm{NC} - 1$

(d) $\mathrm{ML} = \mathrm{MAX}(\mathrm{LX}, \mathrm{LY})$

(e) $11 \leq \mathrm{ISW1} \leq 14$ or $21 \leq \mathrm{ISW2} \leq 24$

(f) $1 \leq \mathrm{ISW2} \leq 7$

(g) When ISW1 = 14 or 24, $\mathrm{DIS}(1) \geq 1.0$

(h) When ISW2 = 7, $-1/4 \leq \mathrm{DIS}(2) \leq 0$
(Except when entering 1.0 to set the default value)

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Restriction (h) was not satisfied. | Processing is performed with the default value set. |
| 3000 | Restriction (a) or (b) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (c) was not satisfied. | |
| 3020 | Restriction (d) was not satisfied. | |
| 3030 | Restriction (e) or (f) was not satisfied. | |
| 3040 | Restriction (g) was not satisfied. | |
| 4000 | For the entered observation data, the calculation cannot be performed using the (generalized) distance of Mahalanobis. (see Section 9.1.1) | |

(6) **Notes**

(a) When the Euclidean quadratic distance or Minkowski distance is used for the dissimilarity measure calculation and the units of the observation data subjects differ, the data should be standardized in advance so that a suitable positional relationship between the subjects is reflected.

(b) Information indicating which cluster was merged with which cluster during which merge is stored in array IPOS. However, the lowest cluster number of the two clusters before merging is assigned to the merged cluster. For example, the cluster formed by merging cluster 3 and cluster 5 is named cluster 3. Where, if NC=NX then m=NY, else if NC=NY then m=NX.

(7) **Example**

(a) Problem

From the following observation data matrix:

$$A = \begin{bmatrix} 0.321 & 119 & 6 & 40 & 6 & 6 & 12 & 29 \\ 0.301 & 112 & 9 & 38 & 4 & 2 & 3 & 31 \\ 0.288 & 133 & 15 & 53 & 4 & 5 & 12 & 30 \\ 0.280 & 112 & 9 & 47 & 4 & 2 & 10 & 24 \\ 0.261 & 109 & 3 & 21 & 1 & 3 & 13 & 21 \\ 0.256 & 107 & 8 & 34 & 4 & 2 & 17 & 38 \\ 0.253 & 95 & 16 & 57 & 4 & 6 & 7 & 37 \\ 0.250 & 100 & 13 & 46 & 4 & 3 & 13 & 37 \\ 0.249 & 92 & 17 & 48 & 6 & 2 & 7 & 23 \\ 0.227 & 88 & 12 & 45 & 4 & 0 & 3 & 33 \end{bmatrix}$$

let the number of rows be the number of subjects and perform a cluster analysis using the group mean method with the dissimilarity measure matrix (standardized Euclidean quadratic distance) as an index.

(b) Input data

Observation data matrix $A$, LX=11, LY=9, NX=10, NY=8, ML=11, NC=10, LNP=9, ISW1=12 and ISW2=3.

(c) Main program

```
      PROGRAM B6CLDA
! *** EXAMPLE OF D6CLDA ***
      IMPLICIT REAL(8) (A-H,O-Z)
      PARAMETER(LX=11,LY=9,NX=10,NY=8,ML=11,NC=10,LNP=9,ISW1=12,ISW2=3)
      DIMENSION A(LX,LY),W1(ML*NC+NY*(ML+1)+2)
      DIMENSION DIS(NC-1),IPOS(LNP,2),IW(NC)
!
      READ(5,*) ((A(I,J),J=1,8),I=1,10)
      WRITE(6,1000)
      WRITE(6,1010) 'A','B','C','D','E','F','G','H'
      DO 10 I=1,NX
        WRITE(6,1020) I,(A(I,J),J=1,NY)
   10 CONTINUE
!
      WRITE(6,1030) ISW1,ISW2,LX,LY,NX,NY,ML,NC,LNP
      CALL D6CLDA(A,LX,LY,NX,NY,ML,NC,IPOS,LNP,DIS,&
                                  ISW1,ISW2,IW,W1,IERR)
      WRITE(6,2000) IERR
!
      DO 20 I=1,NC-1
        WRITE(6,2010) IPOS(I,1),IPOS(I,2),DIS(I)
   20 CONTINUE
      STOP
!
 1000 FORMAT('  *** D6CLDA ***',/,/,3X,'** INPUT **',/)
 1010 FORMAT('    I',8(4X,A,3X),/,2X,'--+',64('-'))
 1020 FORMAT(' ',I3,'I',8F8.3)
 1030 FORMAT(' ',/,3X,'ISW1 =',I4,5X,' ISW2 =',I4,/,&
                 3X,' LX =',I4,5X,' LY =',I4,/,&
                 3X,' NX =',I4,5X,' NY =',I4,/,&
                 3X,' ML =',I4,5X,' NC =',I4,5X,' LNP =',I4,/,/)
 2000 FORMAT('    ** OUTPUT **',/,/,' IERR = ',I4,/)
 2010 FORMAT(11X,I3,' ---',I3,5X,F8.4)
!
      END
```

(d) Output results

```
 *** D6CLDA ***

  ** INPUT **

   I    A       B       C       D       E       F       G       H
 --+----------------------------------------------------------------
   1I   0.321 119.000   6.000  40.000   6.000   6.000  12.000  29.000
   2I   0.301 112.000   9.000  38.000   4.000   2.000   3.000  31.000
   3I   0.288 133.000  15.000  53.000   4.000   5.000  12.000  30.000
   4I   0.280 112.000   9.000  47.000   4.000   2.000  10.000  24.000
   5I   0.261 109.000   3.000  21.000   1.000   3.000  13.000  21.000
   6I   0.256 107.000   8.000  34.000   4.000   2.000  17.000  38.000
   7I   0.253  95.000  16.000  57.000   4.000   6.000   7.000  37.000
   8I   0.250 100.000  13.000  46.000   4.000   3.000  13.000  37.000
```

```
  9I   0.249  92.000  17.000  48.000   6.000   2.000   7.000  23.000
 10I   0.227  88.000  12.000  45.000   4.000   0.000   3.000  33.000

  ISW1 =  12      ISW2 =   3
   LX =  11       LY =   9
   NX =  10       NY =   8
   ML =  11       NC =  10      LNP =   9


   ** OUTPUT **

IERR =    0

            6 ---  8       3.9189
            2 ---  4       4.9420
            9 --- 10       8.5561
            1 ---  3      10.3283
            1 ---  2      10.6194
            6 ---  7      11.7613
            6 ---  9      13.5513
            1 ---  6      15.8940
            1 ---  5      25.2297
```

# Chapter 10

# REGRESSION ANALYSIS

## 10.1 INTRODUCTION

When statistical data consists of multiple variates and an attempt is to be made to predict the value of one of the variates from another of the variates, a **regression analysis** hypothesizes a functional relationship containing several parameters between the two variates and estimates those parameters from observed values. The parameters that are obtained are called the **regression coefficients**. The function obtained by the regression analysis gives a function for approximating the observation data based on some index. This library provides the following functions for performing regression analysis.

- Linear Regression Analysis
- Nonlinear Regression Analysis

### 10.1.1 Explanation

(1) **Linear Regression Analysis**

Assume that $n$ observed values $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ obey the following linear regression model.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (i = 1, 2, \cdots, n)$$

Here, $\varepsilon_i$ are error terms that independently obey $N(0, \sigma^2)$. Obtaining the estimates $b_0$ and $b_1$ of coefficients $\beta_0$ and $\beta_1$ by using the method of least squares is called a linear regression. $b_0$ and $b_1$ are obtained by using the following equations.

$$b_1 = \frac{\displaystyle\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{\displaystyle\sum_{i=1}^{n}(x_i - \mu_x)^2}$$

$$b_0 = \mu_y - b_1 \mu_x$$

Here, $\mu_x$ and $\mu_y$, which represent the means of $x_i$ and $y_i$, are given by the following equations.

$$\mu_x = \frac{\displaystyle\sum_{i=1}^{n} x_i}{n}, \ \mu_y = \frac{\displaystyle\sum_{i=1}^{n} y_i}{n}$$

The analysis of variance table due to the linear regression is as follows.

| | Sum of squares | Degrees of freedom | Unbiased variance | F ratio |
|---|---|---|---|---|
| Total variation | $S_T$ | $n-1$ | | |
| Variation due to regression | $S_R$ | 1 | $V_R = S_R$ | $F = \dfrac{V_R}{V_E}$ |
| Residual variation | $S_E = S_T - S_R$ | $n-2$ | $V_E = \dfrac{S_E}{n-2}$ | |

Here, $S_T$ and $S_R$ are defined as follows.

$$S_T = \sum_{i=1}^{n}(y_i - \mu_y)^2$$

$$S_R = b_1 \left( \sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y) \right)$$

The estimates ${\varepsilon_0}^2$ and ${\varepsilon_1}^2$ of the variance of the statistics $b_0$ and $b_1$, are given by the following equations:

$${\varepsilon_0}^2 = V_E \left( \frac{1}{n} + \frac{{\mu_x}^2}{\displaystyle\sum_{i=1}^{n}(x_i - \mu_x)^2} \right)$$

$$\varepsilon_1{}^2 = \frac{V_E}{\displaystyle\sum_{i=1}^{n}(x_i - \mu_x)^2}$$

and the test quantities $t_0$, which obeys a $t$ distribution with $n-2$ degrees of freedom based on the null hypothesis $H_0 : \beta_0 = 0$, and $t_1$, which obeys a $t$ distribution with $n-2$ degrees of freedom based on the null hypothesis $H_0 : \beta_1 = 0$, respectively, are given as follows.

$$t_0 = \frac{b_0}{\varepsilon_0}$$

$$t_1 = \frac{b_1}{\varepsilon_1}$$

(2) **Linear Regression Analysis (Repetitive Data)**

Assume that mi given dependent variable values $y_{ij}$ $(j = 1, 2, \cdots, m_i; i = 1, 2, \cdots, n)$ corresponding to $n$ independent variable values $x_i$ $(i = 1, 2, \cdots, n)$ have been given and obey the following linear regression model.

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (j = 1, 2, \cdots, m_i; i = 1, 2, \cdots, n)$$

Here, $\varepsilon_i$ are error terms that independently obey $N(0, \sigma^2)$. The estimates $b_0$ and $b_1$ of coefficients $\beta_0$ and $\beta_1$ are to be obtained by using the method of least squares. $b_0$ and $b_1$ are obtained by using the following equations.

$$b_1 = \frac{\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{m_i}(x_i - \mu_x)(y_{ij} - \mu_y)}{\displaystyle\sum_{i=1}^{n} m_i(x_i - \mu_x)^2}$$

$$b_0 = \mu_y - b_1\mu_x$$

Here, $\mu_x$ and $\mu_y$, which represent the (weighted) means of $x_i$ and $y_{ij}$, are given by the following equations.

$$\mu_x = \frac{\displaystyle\sum_{i=1}^{n} m_i x_i}{\displaystyle\sum_{i=1}^{n} m_i}, \quad \mu_y = \frac{\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij}}{\displaystyle\sum_{i=1}^{n} m_i}$$

The analysis of variance table due to the linear regression is as follows.

| | Sum of squares | Degrees of freedom | Unbiased variance | F ratio |
|---|---|---|---|---|
| Total variation | $S_T$ | $\sum_{i=1}^{n} m_i - 1$ | | |
| Variation due to regression | $S_R$ | 1 | $V_R = S_R$ | $F_R = \dfrac{V_R}{V_E}$ |
| Residual variation | $S_E = S_T - S_R$ | $\sum_{i=1}^{n} m_i - 2$ | $V_E = \dfrac{S_E}{\sum_{i=1}^{n} m_i - 2}$ | |
| Variation due to high order regression | $S_L = S_B - S_R$ | $n-2$ | $V_L = \dfrac{S_L}{n-2}$ | $F_L = \dfrac{V_L}{V_W}$ |
| Between-class variation | $S_B$ | $n-1$ | $V_B = \dfrac{S_B}{n-1}$ | $F_B = \dfrac{V_B}{V_W}$ |
| Within-class variation | $S_W = S_T - S_B$ | $\sum_{i=1}^{n} m_i - n$ | $V_W = \dfrac{S_W}{\sum_{i=1}^{n} m_i - n}$ | |

Here, $S_T$, $S_R$ and $S_B$ are defined as follows.

$$S_T = \sum_{i=1}^{n}\sum_{j=1}^{m_i}(y_{ij} - \mu_y)^2$$

$$S_R = b_1 \left( \sum_{i=1}^{n}\sum_{j=1}^{m_i}(x_i - \mu_x)(y_{ij} - \mu_y) \right)$$

$$S_B = \sum_{i=1}^{n} m_i \left( \frac{\sum_{j=1}^{m_i} y_{ij}}{m_i} - \mu_y \right)^2$$

The estimates ${\varepsilon_0}^2$ and ${\varepsilon_1}^2$ of the variance of the statistics $b_0$ and $b_1$, are given by the following equations:

$${\varepsilon_0}^2 = V_E \left( \frac{1}{\sum_{i=1}^{n} m_i} + \frac{{\mu_x}^2}{\sum_{i=1}^{n} m_i(x_i - \mu_x)^2} \right)$$

$${\varepsilon_1}^2 = \frac{V_E}{\sum_{i=1}^{n} m_i(x_i - \mu_x)^2}$$

and the test quantities $t_0$, which obeys a $t$ distribution with $\sum_{i=1}^{n} m_i - 2$ degrees of freedom based on the null hypothesis $H_0 : \beta_0 = 0$, and $t_1$, which obeys a $t$ distribution with $\sum_{i=1}^{n} m_i - 2$ degrees of freedom based on the null hypothesis $H_0 : \beta_1 = 0$, respectively, are given as follows.

$$t_0 = \frac{b_0}{\varepsilon_0}$$

$$t_1 = \frac{b_1}{\varepsilon_1}$$

(3) **Multiple Regression Analysis**

Assume that $n$ independent variable values $x_{ki}$ $(k = 1, 2, \cdots, n; i = 1, 2, \cdots, m)$ for $m$ variables and $n$ given dependent variable values $y_k$ $(k = 1, 2, \cdots, n)$ corresponding to them have been given and obey the following linear regression model.

$$y_k = \beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2} + \cdots + \beta_m x_{km} + \varepsilon_k \quad (k = 1, 2, \cdots, n)$$

Here, $\varepsilon_k$ are error terms that independently obey $N(0, \sigma^2)$. Obtaining the estimates $b_i$ $(i = 0, 1, \cdots, m)$ of partial correlation coefficients $\beta_i$ $(i = 0, 1, \cdots, m)$ by using the method of least squares is called a multiple regression. The $b_i$ $(i = 0, 1, \cdots, m)$ are obtained by solving the following simultaneous linear equations (normal equations) for the $m - 1$ dimensional vector $\boldsymbol{b} = (b_j)$ $(j = 1, 2, \cdots, m)$.

$$C\boldsymbol{b} = \boldsymbol{u}$$

The elements of matrix $C = (c_{ij})$ $(i, j = 1, 2, \cdots, m)$ and vector $\boldsymbol{u} = (u_j)$ $(j = 1, 2, \cdots, m)$ are defined as follows.

$$c_{ij} = \sum_{k=1}^{n} (x_{ki} - \mu_i)(x_{kj} - \mu_j) \ (i, j = 1, 2, \cdots, m)$$

$$u_j = \sum_{k=1}^{n} (x_{ki} - \mu_i)(y_j - \nu) \ (j = 1, 2, \cdots, m)$$

Here, $\mu_i$ and $\nu$ are given by the following equations.

$$\mu_i = \frac{\sum_{k=1}^{n} x_{ki}}{n} \ (i = 1, 2, \cdots, m)$$

$$\nu = \frac{\sum_{k=1}^{n} y_k}{n}$$

Now, $b_0$ is obtained as follows:

$$b_0 = \nu - \sum_{i=1}^{m} b_i \mu_i$$

and the analysis of variance table due to the linear regression is as follows.

| | Sum of squares | Degrees of freedom | Unbiased variance | F ratio |
|---|---|---|---|---|
| Total variation | $S_T$ | $n-1$ | | |
| Variation due to regression | $S_R$ | m | $V_R = \dfrac{S_R}{m}$ | $F = \dfrac{V_R}{V_E}$ |
| Residual variation | $S_E = S_T - S_R$ | $n-m-1$ | $V_E = \dfrac{S_E}{n-m-1}$ | |

Here, $S_T$ and $S_R$ are defined as follows.

$$S_T = \sum_{k=1}^{n} (y_k - \nu)^2$$

$$S_R = \sum_{i=1}^{m} b_i \left( \sum_{k=1}^{n} (x_{ki} - \mu_i)(y_i - \nu) \right)$$

Let the elements of the inverse matrix of matrix $C$ be represented by $d_{ij}$ $(i, j = 1, 2, \cdots, m)$. The estimates ${\varepsilon_i}^2$ $(i = 0, 1, \cdots, m)$ of the variance of the statistics $b_i$ $(i = 0, 1, \cdots, m)$ are given by the following equations:

$${\varepsilon_0}^2 = V_E \left( \frac{1}{n} + \sum_{i=1}^{m} \sum_{j=1}^{m} \mu_i \mu_j d_{ij} \right)$$

$${\varepsilon_i}^2 = V_E d_{ii} \ (i = 1, 2, \cdots, m)$$

and the test quantities $t_i (i = 0, 1, \cdots, m)$, which obey $t$ distributions with $n - m - 1$ degrees of freedom based on the null hypotheses $H_0 : \beta_i = 0$ $(i = 0, 1, \cdots, m)$, are given as follows.

$$t_i = \frac{b_i}{\varepsilon_i} \ (i = 0, 1, \cdots, m)$$

(4) **Polynomial Regression Analysis**

Assume that $n$ independent variable values $x_k$ $(k = 1, 2, \cdots, n)$ and $n$ given dependent variable values $y_k$ $(k = 1, 2, \cdots, n)$ corresponding to them have been given and obey the following regression model.

$$y_k = \beta_0 + \beta_1 x_k + \beta_2 {x_k}^2 + \cdots + \beta_m {x_k}^m + \varepsilon_k \quad (k = 1, 2, \cdots, n)$$

Here, $\varepsilon_k$ are error terms that independently obey $N(0, \sigma^2)$. The estimates $b_i$ $(i = 0, 1, \cdots, m)$ of partial correlation coefficients $\beta_i$ $(i = 0, 1, \cdots, m)$ are obtained by using the method of least squares. The $b_i$ $(i = 0, 1, \cdots, m)$ are obtained by solving the following simultaneous linear equations (normal equations) for the $m - 1$ dimensional vector $\boldsymbol{b} = (b_j)$ $(j = 1, 2, \cdots, m)$.

$$C\boldsymbol{b} = \boldsymbol{u}$$

The elements of matrix $C = (c_{ij})$ $(i, j = 1, 2, \cdots, m)$ and vector $\boldsymbol{u} = (u_j)$ $(j = 1, 2, \cdots, m)$ are defined as follows.

$$c_{ij} = \sum_{k=1}^{n} ({x_k}^i - \mu_i)({x_k}^j - \mu_j) \ (i, j = 1, 2, \cdots, m)$$

$$u_j = \sum_{k=1}^{n} ({x_k}^i - \mu_i)(y_j - \nu) \ (j = 1, 2, \cdots, m)$$

Here, $\mu_i$ and $\nu$ are given by the following equations.

$$\mu_i = \frac{\sum_{k=1}^{n} {x_k}^i}{n} \ (i = 1, 2, \cdots, m)$$

$$\nu = \frac{\sum_{k=1}^{n} y_k}{n}$$

Now, $b_0$ is obtained as follows:

$$b_0 = \nu - \sum_{i=1}^{m} b_i \mu_i$$

and the analysis of variance table due to the regression is as follows.

| | Sum of squares | Degrees of freedom | Unbiased variance | F ratio |
|---|---|---|---|---|
| Total variation | $S_T$ | $n-1$ | | |
| Variation due to regression | $S_R$ | m | $V_R = \frac{S_R}{m}$ | $F = \frac{V_R}{V_E}$ |
| Residual variation | $S_E = S_T - S_R$ | $n-m-1$ | $V_E = \frac{S_E}{n-m-1}$ | |

Here, $S_T$ and $S_R$ are defined as follows.

$$S_T = \sum_{k=1}^{n} (y_k - \nu)^2$$

$$S_R = \sum_{i=1}^{m} b_i \left( \sum_{k=1}^{n} ({x_k}^i - \mu_i)(y_i - \nu) \right)$$

Let the elements of the inverse matrix of matrix $C$ be represented by $d_{ij}$ $(i, j = 1, 2, \cdots, m)$. The estimates ${\varepsilon_i}^2$ $(i = 0, 1, \cdots, m)$ of the variance of the statistics $b_i$ $(i = 0, 1, \cdots, m)$ are given by the following equations:

$${\varepsilon_0}^2 = V_E \left( \frac{1}{n} + \sum_{i=1}^{m} \sum_{j=1}^{m} \mu_i \mu_j d_{ij} \right)$$

$${\varepsilon_i}^2 = V_E d_{ii} \ (i = 1, 2, \cdots, m)$$

and the test quantities $t_i (i = 0, 1, \cdots, m)$, which obey $t$ distributions with $n - m - 1$ degrees of freedom based on the null hypotheses $H_0 : \beta_i = 0$ $(i = 0, 1, \cdots, m)$, are given as follows.

$$t_i = \frac{b_i}{\varepsilon_i} \ (i = 0, 1, \cdots, m)$$

Remark: The coefficient matrix of normal equations often is rather ill-behaved, and, in particular, the solution has a tendency to quickly approach a singularity as the number of variates increases. For a polynomial regression, this is serious. Therefore, the calculation should be performed using the double-precision subroutine whenever possible.

(5) **Regression According to an Arbitrary Function**

Assume that $n$ sets of independent variables $\boldsymbol{x}_k = (x_{1k}, x_{2k}, \cdots, x_{mk})$ $(k = 1, 2, \cdots, n)$ consisting of $m$ variables and $n$ given dependent variable values $y_k$ $(k = 1, 2, \cdots, n)$ corresponding to them have been given and obey the following regression model.

$$y_k = f(\boldsymbol{x}_k; \boldsymbol{\beta}) + \varepsilon_k \ (k = 1, 2, \cdots, n)$$

Here, $\varepsilon_k$ are error terms that independently obey $N(0, \sigma^2)$. The estimates $\boldsymbol{b} = (b_1, b_2, \cdots, b_\ell)$ of the $\ell$ regression model parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_\ell)$ is obtained according to the nonlinear method of least squares, with $\boldsymbol{\beta}$ minimizing the residual variation

$$S_e = \sum_{k=1}^{n} (y_k - f(x_k; \boldsymbol{\beta}))^2,$$

and the analysis of variance table due to the regression is as follows.

| | Sum of squares | Degrees of freedom | Unbiased variance |
|---|---|---|---|
| Total variation | $S_T$ | $n$ | |
| Residual variation | $S_E$ | $n - \ell$ | $V_E = \dfrac{S_E}{n - \ell}$ |

Here, $S_T$ and $S_E$ are defined as follows.

$$S_T = \sum_{k=1}^{n} y_k^2$$

$$S_E = \sum_{i=1}^{n} (y_k - f(\boldsymbol{x}_k; \boldsymbol{b}))^2$$

In addition, the asymptotic variance-covariance matrix $V = (V_{ij})$ $(i, j = 1, \cdots, \ell)$ and standard deviation estimates $\varepsilon_i$ $(i = 1, \cdots, \ell)$ of the statistics $b_i$ $(i = 1, \cdots, \ell)$ are defined by the following equations.

$$V = S_E (J^T J)^{-1}$$

$$\varepsilon_i = \sqrt{V_{ii}} \ (i = 1, 2, \cdots, \ell)$$

Here, $J$ is the Jacobian matrix for which the $(i, j)$-th element is defined as follows.

$$J_{ij} = \left. \frac{\partial f(\boldsymbol{x}_i; \boldsymbol{\beta})}{\partial \beta_j} \right|_{\boldsymbol{\beta} = \boldsymbol{b}} \qquad (i = 1, \cdots, n; j = 1, \cdots, \ell)$$

### 10.1.2 Reference Bibliography

(1) Draper, N. R. and Smith, H. , “Applied regression analysis”, John Wiley & Sons, New York (1966)

## 10.2 LINEAR REGRESSION ANALYSIS

### 10.2.1 DNLNRG, RNLNRG
### Linear Regression Analysis

(1) **Function**

Assume that $n$ observed values $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ obey the following linear regression model.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (i = 1, 2, \cdots, n)$$

Here, $\varepsilon_i$ are error terms that independently obey $N(0, \sigma^2)$.

The DNLNRG or RNLNRG obtains:

(a) the estimates $b_0$ and $b_1$ of coefficients $\beta_0$ and $\beta_1$

(b) the statistics for creating the following analysis of variance table due to the linear regression:

| | Sum of squares | Degrees of freedom | Unbiased variance | F ratio |
|---|---|---|---|---|
| Total variation | $S_T$ | $n-1$ | | |
| Variation due to regression | $S_R$ | 1 | $V_R = S_R$ | $F = \dfrac{V_R}{V_E}$ |
| Residual variation | $S_E = S_T - S_R$ | $n-2$ | $V_E = \dfrac{S_E}{n-2}$ | |

(c) the multiple correlation coefficient $R$

$$R = \sqrt{\frac{S_R}{S_T}}$$

(d) the standard deviation estimates $\varepsilon_0$ and $\varepsilon_1$ of the statistics $b_0$ and $b_1$

(e) the test quantity $t_0$, which obeys a $t$ distribution with $n-2$ degrees of freedom based on the null hypothesis $H_0 : \beta_0 = 0$

(f) the test quantity $t_1$, which obeys a $t$ distribution with $n-2$ degrees of freedom based on the null hypothesis $H_0 : \beta_1 = 0$.

(2) **Usage**

Double precision:

CALL DNLNRG (X, Y, N, B0, B1, R, STAT, IERR)

Single precision:

CALL RNLNRG (X, Y, N, B0, B1, R, STAT, IERR)

(3) **Arguments**

D:Double precision real Z:Double precision complex
R:Single precision real C:Single precision complex
I: {INTEGER(4) as for 32bit Integer, INTEGER(8) as for 64bit Integer}

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | X | {D, R} | N | Input | Independent variate observation data $x_i$ |
| 2 | Y | {D, R} | N | Input | Dependent variate observation data $y_i$ |
| 3 | N | I | 1 | Input | Number of observed values $n$ |
| 4 | B0 | {D, R} | 3 | Output | B0(1): $\beta_0$ estimate $b_0$<br>B0(2): $\beta_0$ standard deviation estimate $\varepsilon_0$<br>B0(3): $\beta_0$ test quantity $t_0$ |
| 5 | B1 | {D, R} | 3 | Output | B1(1): $\beta_1$ estimate $b_1$<br>B1(2): $\beta_1$ standard deviation estimate $\varepsilon_1$<br>B1(3): $\beta_1$ test quantity $t_1$ |
| 6 | R | {D, R} | 2 | Output | R(1): Multiple correlation coefficient $R$<br>R(2): Contribution ratio $R^2$ |
| 7 | STAT | {D, R} | 21 | Output | Basic statistics of calculation result (See Note (a)) |
| 8 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $\mathrm{N} \geq 3$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | There was no difference between independent variates. | 0.0 is set for B0, B1, R, and STAT(1) to STAT(9). |
| 1010 | The residual is 0.0. (STAT(3)=0.0) | The positive maximum value is set for STAT(9), B0(3), and B1(3). |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |

(6) **Notes**

(a) The following kind of array is stored as a one-dimensional vector in array STAT.

| | | |
|---|---|---|
| 1: | $S_T$ | Total variation |
| 2: | $S_R$ | Variation due to regression |
| 3: | $S_E$ | Residual variation |
| 4: | $f_T = n - 1$ | Degrees of freedom of total variation |
| 5: | $f_R = 1$ | Degrees of freedom of variation due to regression |
| 6: | $f_E = n - 2$ | Degrees of freedom of residual variation |
| 7: | $V_R$ | Unbiased variance of variation due to regression |
| 8: | $V_E$ | Unbiased variance of residual variation |
| 9: | $F$ | F ratio |
| 10: | $\mu_x = \frac{\sum_{i=1}^{n} x_i}{n}$ | Mean of $x_i$ |
| 11: | $\mu_y = \frac{\sum_{i=1}^{n} y_i}{n}$ | Mean of $y_i$ |
| 12: | $\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu_x)^2}{n-1}}$ | Standard deviation of $x_i$ |
| 13: | $\sigma_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\mu_y)^2}{n-1}}$ | Standard deviation of $y_i$ |
| 14: | $s_x = \sum_{i=1}^{n} x_i$ | Sum of $x_i$ |
| 15: | $s_y = \sum_{i=1}^{n} y_i$ | Sum of $y_i$ |
| 16: | $s_{xx} = \sum_{i=1}^{n} {x_i}^2$ | Sum of squares of $x_i$ |
| 17: | $s_{yy} = \sum_{i=1}^{n} {y_i}^2$ | Sum of squares of $y_i$ |
| 18: | $s_{xy} = \sum_{i=1}^{n} x_i y_i$ | Sum of products of $x_i$ and $y_i$ |
| 19: | $S_{xx} = \sum_{i=1}^{n}(x_i - \mu_x)^2$ | Sum of squares of deviations of $x_i$ |
| 20: | $S_{yy} = \sum_{i=1}^{n}(y_i - \mu_y)^2$ | Sum of squares of deviations of $y_i$ |
| 21: | $S_{xy} = \sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)$ | Sum of products of deviations of $x_i$ and $y_i$ |

(7) **Example**

(a) Problem

Obtain the regression coefficient values, multiple correlation coefficient value, contribution ratio, and basic statistics according to a regression analysis from the following observation data.

X(1)=1.0 Y(1)=-2.0
X(2)=2.0 Y(2)=7.0
X(3)=3.0 Y(3)=34.0
X(4)=4.0 Y(4)=91.0
X(5)=5.0 Y(5)=190.0
X(6)=6.0 Y(6)=343.0
X(7)=7.0 Y(7)=562.0
X(8)=8.0 Y(8)=859.0
X(9)=9.0 Y(9)=1246.0
X(10)=10.0 Y(10)=1735.0
X(11)=11.0 Y(11)=2338.0

(b) Input data

Observation data X, Y and N=11.

(c) Main program

```
      PROGRAM BNLNRG
! *** EXAMPLE OF DNLNRG ***
      IMPLICIT REAL(8) (A-H,O-Z)
      DIMENSION X(11),Y(11),B0(3),B1(3),R(2),STAT(21)
!
      READ(5,*) N
      READ(5,*) (X(I),I=1,N)
      READ(5,*) (Y(I),I=1,N)
!
      WRITE(6,1000)
      DO 10 I=1,N
        WRITE(6,2000) I,X(I),Y(I)
   10 CONTINUE
!
      CALL DNLNRG(X,Y,N,B0,B1,R,STAT,IERR)
!
      WRITE(6,3000) IERR
      WRITE(6,3010) (B1(I),I=1,3)
      WRITE(6,3020) (B0(I),I=1,3)
      WRITE(6,3030) (R(I),I=1,2)
      WRITE(6,4000)
      WRITE(6,4010) STAT(1),INT(STAT(4)),STAT(2),INT(STAT(5)),&
                    STAT(3),INT(STAT(6))
      WRITE(6,4020)
      WRITE(6,4030) STAT(7),STAT(9),STAT(8)
      WRITE(6,5000)
      WRITE(6,5010) 'MEAN OF X                  : ',10,STAT(10)
      WRITE(6,5010) 'MEAN OF Y                  : ',11,STAT(11)
      WRITE(6,5010) 'STANDARD DEVIATION OF X    : ',12,STAT(12)
      WRITE(6,5010) 'STANDARD DEVIATION OF Y    : ',13,STAT(13)
      WRITE(6,5010) 'SUM OF X                   : ',14,STAT(14)
      WRITE(6,5010) 'SUM OF Y                   : ',15,STAT(15)
      WRITE(6,5010) 'SUM OF SQUARES OF X        : ',16,STAT(16)
      WRITE(6,5010) 'SUM OF SQUARES OF Y        : ',17,STAT(17)
      WRITE(6,5010) 'SUM OF PRODUCTS OF X AND Y : ',18,STAT(18)
      WRITE(6,5020) 'SUM OF SQUARES OF            '
      WRITE(6,5010) '    DEVIATIONS OF X        : ',19,STAT(19)
      WRITE(6,5020) 'SUM OF SQUARES OF            '
      WRITE(6,5010) '    DEVIATIONS OF Y        : ',20,STAT(20)
      WRITE(6,5020) 'SUM OF PRODUCTS OF           '
      WRITE(6,5010) '    DEVIATIONS OF X AND Y  : ',20,STAT(21)
!
      STOP
!
 1000 FORMAT(' **** INPUT DATA ****',/,/,6X,'  N','I',15X,'X',24X,'Y',&
               /,6X,'---+',50('-'))
 2000 FORMAT(6X,I3,'I',2(5X,D20.10))
 3000 FORMAT(' ',/,/,' *** OUTPUT ***',/,/,4X,'IERR=',I4,&
             /,/,3X,'*REGRESSION COEFFICIENT*',&
             /,/,21X,'R.C',13X,'S.E',12X,'T-V',/,4X,12('-'),'+',42('-'))
 3010 FORMAT(4X,'B1(X)',7X,'I',3D14.6)
 3020 FORMAT(4X,'B0(CONSTANT)','I',3D14.6)
 3030 FORMAT(' ',/,4X,'R(1)(MULTIPLE CORRELATION COEFFICIENT) =',&
               D20.10,/,4X,'R(2)(CONTRIBUTION RATIO)',15X,'=',D20.10)
 4000 FORMAT(' ',/,/,3X,'*ANALYSIS OF VARIANCE TABLE*',/,/,4X,'FACTOR OF ',&
```

```
                 /,4X,'VARIATION',2X,'I',5X,'SUM OF SQUARES',&
                 15X,'DEGREES OF FREEDOM',/,4X,11('-'),'+',55('-'))
 4010 FORMAT(4X,'TOTAL',6X,'ISTAT(1) = ',D18.10,6X,&
                                'STAT(4) = ',I3,/,15X,'I',&
                 /,4X,'DUE TO',5X,'ISTAT(2) = ',D18.10,6X,&
                  'STAT(5) = ',I3,/,4X,'REGRESSION I',/,15X,'I',&
                 /,4X,'DEVIATION  ISTAT(3) = ',D18.10,6X,&
                  'STAT(6) = ',I3,/,5X,'FROM',6X,'I',/,4X,'REGRESSION I')
 4020 FORMAT(' ',/,/,4X,'FACTOR OF ',/,4X,'VARIATION',&
                   2X,'I',3X,'UNBIASED VARIANCE',&
                   21X,'F-RATIO',/,4X,11('-'),'+',62('-'))
 4030 FORMAT(4X,'DUE TO',5X,'ISTAT(7) = ',D18.10,6X,&
                  'STAT(9) = ',D18.10,/,4X,'REGRESSION I',/,15X,'I',&
                 /,4X,'DEVIATION  ISTAT(8) = ',D18.10,6X,&
                 /,5X,'FROM',6X,'I',/,4X,'REGRESSION I')
 5000 FORMAT(' ',/,/,3X,'*STATISTICS*')
 5010 FORMAT(5X,A31,'STAT(',I2,') =',D14.6)
 5020 FORMAT(5X,A31)
      END
```

(d) Output results

```
**** INPUT DATA ****

       NI              X                         Y
     ---+--------------------------------------------------
       1I        0.1000000000D+01       -0.2000000000D+01
       2I        0.2000000000D+01        0.7000000000D+01
       3I        0.3000000000D+01        0.3400000000D+02
       4I        0.4000000000D+01        0.9100000000D+02
       5I        0.5000000000D+01        0.1900000000D+03
       6I        0.6000000000D+01        0.3430000000D+03
       7I        0.7000000000D+01        0.5620000000D+03
       8I        0.8000000000D+01        0.8590000000D+03
       9I        0.9000000000D+01        0.1246000000D+04
      10I        0.1000000000D+02        0.1735000000D+04
      11I        0.1100000000D+02        0.2338000000D+04

*** OUTPUT ***

   IERR=   0

  *REGRESSION COEFFICIENT*

                   R.C            S.E            T-V
  ------------+-----------------------------------------
  B1(X)       I  0.219600D+03  0.311249D+02  0.705544D+01
  B0(CONSTANT)I -0.644600D+03  0.211099D+03 -0.305354D+01

  R(1)(MULTIPLE CORRELATION COEFFICIENT) =    0.9202634294D+00
  R(2)(CONTRIBUTION RATIO)               =    0.8468847795D+00

  *ANALYSIS OF VARIANCE TABLE*

   FACTOR OF
   VARIATION  I     SUM OF SQUARES              DEGREES OF FREEDOM
   -----------+-------------------------------------------------------
   TOTAL      ISTAT(1) =   0.6263730000D+07      STAT(4) =  10
              I
   DUE TO     ISTAT(2) =   0.5304657600D+07      STAT(5) =   1
   REGRESSION I
              I
   DEVIATION  ISTAT(3) =   0.9590724000D+06      STAT(6) =   0
    FROM      I
   REGRESSION I


   FACTOR OF
   VARIATION  I   UNBIASED VARIANCE                       F-RATIO
   -----------+--------------------------------------------------------------
   DUE TO     ISTAT(7) =   0.5304657600D+07      STAT(9) =   0.4977926421D+02
   REGRESSION I
              I
   DEVIATION  ISTAT(8) =   0.1065636000D+06
    FROM      I
   REGRESSION I


  *STATISTICS*
      MEAN OF X                     : STAT(10) =  0.600000D+01
      MEAN OF Y                     : STAT(11) =  0.673000D+03
      STANDARD DEVIATION OF X       : STAT(12) =  0.331662D+01
      STANDARD DEVIATION OF Y       : STAT(13) =  0.791437D+03
      SUM OF X                      : STAT(14) =  0.660000D+02
      SUM OF Y                      : STAT(15) =  0.740300D+04
      SUM OF SQUARES OF X           : STAT(16) =  0.506000D+03
      SUM OF SQUARES OF Y           : STAT(17) =  0.112459D+08
      SUM OF PRODUCTS OF X AND Y    : STAT(18) =  0.685740D+05
      SUM OF SQUARES OF
          DEVIATIONS OF X           : STAT(19) =  0.110000D+03
      SUM OF SQUARES OF
          DEVIATIONS OF Y           : STAT(20) =  0.626373D+07
```

```
SUM OF PRODUCTS OF
    DEVIATIONS OF X AND Y  : STAT(20) =  0.241560D+05
```

### 10.2.2 DNLNRR, RNLNRR Linear Regression Analysis (Repetitive Data)

(1) **Function**

Assume that $m_i$ given dependent variable values $y_{ij}$ $(j = 1, 2, \cdots, m_i; i = 1, 2, \cdots, n)$ corresponding to $n$ independent variable values $x_i$ $(i = 1, 2, \cdots, n)$ have been given and obey the following linear regression model.

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (j = 1, 2, \cdots, m_i; i = 1, 2, \cdots, n)$$

Here, $\varepsilon_i$ are error terms that independently obey $N(0, \sigma^2)$.

The DNLNRR or RNLNRR obtains:

(a) the estimates $b_0$ and $b_1$ of coefficients $\beta_0$ and $\beta_1$

(b) the statistics for creating the following analysis of variance table due to the linear regression.

| | Sum of squares | Degrees of freedom | Unbiased variance | F ratio |
|---|---|---|---|---|
| Total variation | $S_T$ | $\sum_{i=1}^{n} m_i - 1$ | | |
| Variation due to regression | $S_R$ | 1 | $V_R = S_R$ | $F_R = \dfrac{V_R}{V_E}$ |
| Residual variation | $S_E = S_T - S_R$ | $\sum_{i=1}^{n} m_i - 2$ | $V_E = \dfrac{S_E}{\sum_{i=1}^{n} m_i - 2}$ | |
| Variation due to high order regression | $S_L = S_B - S_R$ | $n - 2$ | $V_L = \dfrac{S_L}{n-2}$ | $F_L = \dfrac{V_L}{V_W}$ |
| Between-class variation | $S_B$ | $n - 1$ | $V_B = \dfrac{S_B}{n-1}$ | $F_B = \dfrac{V_B}{V_W}$ |
| Within-class variation | $S_W = S_T - S_B$ | $\sum_{i=1}^{n} m_i - n$ | $V_W = \dfrac{S_W}{\sum_{i=1}^{n} m_i - n}$ | |

(c) the multiple correlation coefficient $R$

$$R = \sqrt{\frac{S_R}{S_T}}$$

(d) the standard deviation estimates $\varepsilon_0$ and $\varepsilon_1$ of the statistics $b_0$ and $b_1$

(e) the test quantity $t_0$, which obeys a $t$ distribution with $\sum_{i=1}^{n} m_i - 2$ degrees of freedom based on the null hypothesis $H_0 : \beta_0 = 0$

(f) the test quantity $t_1$, which obeys a $t$ distribution with $\sum_{i=1}^{n} m_i - 2$ degrees of freedom based on the null hypothesis $H_0 : \beta_1 = 0$

(2) **Usage**

Double precision:

CALL DNLNRR (X, N, NY, Y, MY, M, B0, B1, R, STAT, IERR)

Single precision:

CALL RNLNRR (X, N, NY, Y, MY, M, B0, B1, R, STAT, IERR)

(3) **Arguments**

D:Double precision real　Z:Double precision complex　I: { INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer }
R:Single precision real　C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | X | {D / R} | N | Input | Independent variate observation data $x_i$ |
| 2 | N | I | 1 | Input | Number of independent variate observed values $n$ |
| 3 | NY | I | N | Input | Number of repetitions of each dependent variate for an independent variate $m_i$ |
| 4 | Y | {D / R} | MY , M | Input | Dependent variate observation data $y_{ij}$ (See Note (a)) |
| 5 | MY | I | 1 | Input | Adjustable dimension of array Y |
| 6 | M | I | 1 | Input | Maximum number of repetitions of dependent variate observed values max($m_i$) (See Note (c)) |
| 7 | B0 | {D / R} | 3 | Output | B0(1): $\beta_0$ estimate $b_0$<br>B0(2): $\beta_0$ standard deviation estimate $\varepsilon_0$<br>B0(3): $\beta_0$ test quantity $t_0$ |
| 8 | B1 | {D / R} | 3 | Output | B1(1): $\beta_1$ estimate $b_1$<br>B1(2): $\beta_1$ standard deviation estimate $\varepsilon_1$<br>B1(3): $\beta_1$ test quantity $t_1$ |
| 9 | R | {D / R} | 2 | Output | R(1): Multiple correlation coefficient $R$<br>R(2): Contribution ratio $R^2$ |
| 10 | STAT | {D / R} | 33 | Output | Basic statistics of calculation result (See Note (b)) |
| 11 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $3 \leq \mathrm{N} \leq \mathrm{MY}$

(b) $1 \leq \mathrm{NY(i)} \leq \mathrm{M}$ $(\mathrm{i} = 1, 2, \cdots, \mathrm{N})$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | There was no difference between independent variates. | 0.0 is set for B0, B1, R, and STAT(1) to STAT(20). |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |

(6) **Notes**

(a) Observation data $y_{i,j}$ is stored in array Y as the following kind of real matrix (two-dimensional array type) data (See Appendix A).

$$\begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,m_1} & * & * \\ y_{2,1} & y_{2,2} & \cdots & \cdots & y_{2,m_2} & * \\ \vdots & & & \vdots & * & * \\ y_{i,1} & & \cdots & \cdots & \cdots & y_{i,m_i} \\ \vdots & & & \vdots & * & * \\ y_{n,1} & \cdots & y_{n,m_n} & * & * & * \end{bmatrix}$$

Remark: The asterisks ($*$) indicate arbitrary values.

(b) The following kind of array is stored as a one-dimensional vector in array STAT.

| | | |
|---|---|---|
| 1: | $S_T$ | Total variation |
| 2: | $S_R$ | Variation due to regression |
| 3: | $S_E$ | Residual variation |
| 4: | $f_T = \sum_{i=1}^{n} m_i - 1$ | Degrees of freedom of total variation |
| 5: | $f_R = 1$ | Degrees of freedom of variation due to regression |
| 6: | $f_E = \sum_{i=1}^{n} m_i - 2$ | Degrees of freedom of residual variation |
| 7: | $V_R$ | Unbiased variance of variation due to regression |
| 8: | $V_E$ | Unbiased variance of residual variation |
| 9: | $F_T$ | F ratio of variation due to regression |
| 10: | $S_L$ | Variation due to high-order regression |
| 11: | $S_B$ | Between-class variation |
| 12: | $S_W$ | Within-class variation |
| 13: | $f_L = n - 2$ | Degrees of freedom due to high-order regression |
| 14: | $f_B = n - 1$ | Degrees of freedom of between-class variation |
| 15: | $f_W = \sum_{i=1}^{n} m_i - n$ | Degrees of freedom of within-class variation |

| | | |
|---|---|---|
| 16: | $V_L$ | Unbiased variance of variation due to high-order regression |
| 17: | $V_B$ | Unbiased variance of between-class variation |
| 18: | $V_W$ | Unbiased variance of within-class variation |
| 19: | $F_L$ | F ratio of variation due to high-order regression |
| 20: | $F_B$ | F ratio of between-class variation |
| 21: | $\mu_x = \frac{\sum_{i=1}^{n} m_i x_i}{\sum_{i=1}^{n} m_i}$ | Weighted mean of $x_i$ |
| 22: | $\mu_y = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m_i} y_{ij}}{\sum_{i=1}^{n} m_i}$ | Grand mean of $y_{ij}$ |
| 23: | $\sigma_x = \sqrt{\frac{\sum_{i=1}^{n} m_i (x_i - \mu_x)^2}{\sum_{i=1}^{n} m_i - 1}}$ | Weighted standard deviation of $x_i$ |
| 24: | $\sigma_y = \sqrt{\frac{\sum_{i=1}^{n} \sum_{j=1}^{m_i} (y_{ij} - \mu_y)^2}{\sum_{i=1}^{n} m_i - 1}}$ | Standard deviation of $y_{ij}$ |
| 25: | $s_x = \sum_{i=1}^{n} m_i x_i$ | Weighted sum of $x_i$ |
| 26: | $s_y = \sum_{i=1}^{n} \sum_{j=1}^{m_i} y_{ij}$ | Sum of $y_{ij}$ |
| 27: | $s_{xx} = \sum_{i=1}^{n} m_i {x_i}^2$ | Weighted sum of squares of $x_i$ |
| 28: | $s_{yy} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} {y_{ij}}^2$ | Sum of squares of $y_{ij}$ |
| 29: | $s_{xy} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} x_i y_{ij}$ | Sum of products of $x_i$ and $y_{ij}$ |
| 30: | $S_{xx} = \sum_{i=1}^{n} m_i (x_i - \mu_x)^2$ | Weighted sum of squares of deviations of $x_i$ |
| 31: | $S_{yy} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} (y_{ij} - \mu_y)^2$ | Sum of squares of deviations of $y_{ij}$ |
| 32: | $S_{xy} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} (x_i - \mu_x)(y_{ij} - \mu_y)$ | Sum of products of deviations of $x_i$ and $y_{ij}$ |
| 33: | $\sum_{i=1}^{n} m_i$ | Total number of $y_{ij}$ |

(c) When M=1, STAT(10) to STAT(20) are not calculated.

(7) **Example**

(a) Problem

Obtain the regression coefficient values, multiple correlation coefficient value, contribution ratio, and basic statistics according to a regression analysis from the following observation data:

$$(x_i) = \begin{bmatrix} 55.0 \\ 65.0 \\ 75.0 \\ 80.0 \\ 85.0 \end{bmatrix}, \quad (y_{i,j}) = \begin{bmatrix} 26.0 & 29.0 & 32.0 & 0.0 \\ 32.0 & 35.0 & 0.0 & 0.0 \\ 36.0 & 35.0 & 31.0 & 34.0 \\ 33.0 & 0.0 & 0.0 & 0.0 \\ 37.0 & 35.0 & 0.0 & 0.0 \end{bmatrix}$$

and numbers of repetitions:

$$(m_i) = \begin{bmatrix} 3 \\ 2 \\ 4 \\ 1 \\ 2 \end{bmatrix}$$

(b) Input data

Observation data X and Y, numbers of repetitions NY, N=5, MY=5 and M=4.

(c) Main program

```
      PROGRAM BNLNRR
! *** EXAMPLE OF DNLNRR ***
      IMPLICIT REAL(8) (A-H,O-Z)
      DIMENSION X(5),NY(5),Y(5,4),B0(3),B1(3),R(2),STAT(33)
!
      READ(5,*) MY,N,M
      READ(5,*) (NY(I),I=1,N)
      READ(5,*) (X(I),I=1,N)
      READ(5,*) ((Y(I,J),I=1,MY),J=1,M)
!
      WRITE(6,1000) MY,N,M
      WRITE(6,1010)
      DO 10 I=1,N
        WRITE(6,2000) I,X(I),NY(I),(Y(I,J),J=1,NY(I))
   10 CONTINUE
!
      CALL DNLNRR(X,N,NY,Y,MY,M,B0,B1,R,STAT,IERR)
!
      WRITE(6,3000) IERR
      WRITE(6,3010) (B1(I),I=1,3)
      WRITE(6,3020) (B0(I),I=1,3)
      WRITE(6,3030) (R(I),I=1,2)
!
      WRITE(6,4000)
      WRITE(6,4010) STAT(1),INT(STAT(4)),STAT(2),INT(STAT(5)),&
                    STAT(3),INT(STAT(6))
      WRITE(6,4020) STAT(10),INT(STAT(13)),STAT(11),INT(STAT(14)),&
                    STAT(12),INT(STAT(15))
      WRITE(6,5000) STAT(7),STAT(9),STAT(8)
      WRITE(6,5010) STAT(16),STAT(19),STAT(17),STAT(20),STAT(18)
!
      WRITE(6,6000)
      WRITE(6,6010) 'WEIGHTED MEAN OF X           : ',21,STAT(21)
      WRITE(6,6010) 'GRAND MEAN OF Y              : ',22,STAT(22)
      WRITE(6,6020) 'WEIGHTED STANDARD              '
      WRITE(6,6010) '    DEVIATION OF X           : ',23,STAT(23)
      WRITE(6,6010) 'STANDARD DEVIATION OF Y      : ',24,STAT(24)
      WRITE(6,6010) 'WEIGHTED SUM OF X            : ',25,STAT(25)
      WRITE(6,6010) 'SUM OF Y                     : ',26,STAT(26)
      WRITE(6,6010) 'WEIGHTED SUM OF SQUARES OF X : ',27,STAT(27)
      WRITE(6,6010) 'SUM OF SQUARES OF Y          : ',28,STAT(28)
      WRITE(6,6010) 'SUM OF PRODUCTS OF X AND Y   : ',29,STAT(29)
      WRITE(6,6020) 'WEIGHTED SUM OF SQUARES OF     '
      WRITE(6,6010) '    DEVIATIONS OF X          : ',30,STAT(30)
      WRITE(6,6020) 'SUM OF SQUARES OF              '
      WRITE(6,6010) '    DEVIATIONS OF Y          : ',32,STAT(31)
      WRITE(6,6020) 'SUM OF PRODUCTS OF             '
      WRITE(6,6010) '    DEVIATIONS OF X AND Y    : ',32,STAT(32)
      WRITE(6,6010) 'TOTAL NUMBER OF Y            : ',33,STAT(33)
!
      STOP
!
 1000 FORMAT(' **** INPUT DATA ****',&
```

```
                 /,/,8X,'MY =',I2,5X,'N =',I2,5X,'M =',I2)
 1010 FORMAT(' ',/,6X,'  N',' I',6X,'X',6X,'NY',17X,'Y',&
                 /,7X,'---+',65('-'))
 2000 FORMAT(6X,I3,' I',D10.2,I5,4D12.4)
 3000 FORMAT(' ',/,/,' *** OUTPUT ***',/,/,4X,'IERR=',I4,&
             /,/,3X,'*REGRESSION COEFFICIENT*',&
             /,/,21X,'R.C',13X,'S.E',12X,'T-V',/,4X,12('-'),'+',42('-'))
 3010 FORMAT(4X,'B1(X)',7X,'I',3D14.6)
 3020 FORMAT(4X,'B0(CONSTANT)','I',3D14.6)
 3030 FORMAT(' ',/,4X,'R(1)(MULTIPLE CORRELATION COEFFICIENT) =',&
                 D20.10,/,4X,'R(2)(CONTRIBUTION RATIO)',15X,'=',D20.10)
 4000 FORMAT(' ',/,/,3X,'*ANALYSIS OF VARIANCE TABLE*',/,/,&
                  4X,'FACTOR OF ',/,4X,'VARIATION',&
                  2X,'I',5X,'SUM OF SQUARES',&
                  13X,'DEGREES OF FREEDOM',/,4X,11('-'),'+',55('-'))
 4010 FORMAT(4X,'TOTAL',6X,'ISTAT(1)  = ',D18.10,3X,&
                'STAT(4)  = ',I3,/,15X,'I',&
                /,4X,'DUE TO',5X,'ISTAT(2)  = ',D18.10,3X,&
                'STAT(5)  = ',I3,/,4X,'REGRESSION I',/,15X,'I',&
                /,4X,'DEVIATION  ISTAT(3)  = ',D18.10,3X,&
                'STAT(6)  = ',I3,/,5X,'FROM',6X,'I',/,4X,'REGRESSION I',&
                 /,15X,'I')
 4020 FORMAT(4X,'LACK OF',4X,'ISTAT(10) = ',D18.10,3X,&
                'STAT(13) = ',I3,/,4X,'FITNESS',4X,'I',/,15X,'I',&
                /,4X,'BETWEEN',4X,'ISTAT(11) = ',D18.10,3X,&
                'STAT(14) = ',I3,/,15X,'I',&
                /,4X,'WITHIN',5X,'ISTAT(12) = ',D18.10,3X,&
                'STAT(15) = ',I3)
 5000 FORMAT(' ',/,/,4X,'FACTOR OF ',/,4X,'VARIATION',&
                2X,'I',3X,'UNBIASED VARIANCE',&
                21X,'F-RATIO',/,4X,11('-'),'+',62('-'),&
                /,4X,'DUE TO',5X,'ISTAT(7)  = ',D18.10,3X,&
                'STAT(9)  = ',D18.10,/,4X,'REGRESSION I',/,15X,'I',&
                /,4X,'DEVIATION  ISTAT(8)  = ',D18.10,6X,&
                /,5X,'FROM',6X,'I',/,4X,'REGRESSION I',/,15X,'I')
 5010 FORMAT(4X,'LACK OF',4X,'ISTAT(16) = ',D18.10,3X,&
                'STAT(19) = ',D18.10,/,4X,'FITNESS',4X,'I',/,15X,'I',&
                /,4X,'BETWEEN',4X,'ISTAT(17) = ',D18.10,3X,&
                'STAT(20) = ',D18.10,/,15X,'I',&
                /,4X,'WITHIN',5X,'ISTAT(18) = ',D18.10)
 6000 FORMAT(' ',/,/,3X,'*STATISTICS*',/)
 6010 FORMAT(5X,A31,'STAT(',I2,') =',D14.6)
 6020 FORMAT(5X,A31)
      END
```

(d) Output results

```
**** INPUT DATA ****

      MY = 5     N = 5     M = 4

      N I      X      NY                  Y
      ---+-----------------------------------------------------------------
      1 I  0.55D+02    3  0.2600D+02  0.2900D+02  0.3200D+02
      2 I  0.65D+02    2  0.3200D+02  0.3500D+02
      3 I  0.75D+02    4  0.3600D+02  0.3500D+02  0.3100D+02  0.3400D+02
      4 I  0.80D+02    1  0.3300D+02
      5 I  0.85D+02    2  0.3700D+02  0.3500D+02


*** OUTPUT ***

   IERR=   0

  *REGRESSION COEFFICIENT*

                     R.C             S.E             T-V
   ------------+------------------------------------------
   B1(X)       I  0.207891D+00  0.600947D-01  0.345938D+01
   B0(CONSTANT)I  0.182777D+02  0.428021D+01  0.427028D+01

   R(1)(MULTIPLE CORRELATION COEFFICIENT) =    0.7380912731D+00
   R(2)(CONTRIBUTION RATIO)               =    0.5447787274D+00


  *ANALYSIS OF VARIANCE TABLE*

   FACTOR OF
   VARIATION  I     SUM OF SQUARES             DEGREES OF FREEDOM
   -----------+-------------------------------------------------------
   TOTAL      ISTAT(1)  =    0.1089166667D+03   STAT(4)  =  11
              I
   DUE TO     ISTAT(2)  =    0.5933548306D+02   STAT(5)  =   1
   REGRESSION I
              I
   DEVIATION  ISTAT(3)  =    0.4958118361D+02   STAT(6)  =  10
    FROM      I
   REGRESSION I
              I
   LACK OF    ISTAT(10) =    0.1108118361D+02   STAT(13) =   3
   FITNESS    I
              I
   BETWEEN    ISTAT(11) =    0.7041666667D+02   STAT(14) =   4
              I
```

```
WITHIN     ISTAT(12) =   0.3850000000D+02   STAT(15) =   7

FACTOR OF
VARIATION  I   UNBIASED VARIANCE                     F-RATIO
-----------+----------------------------------------------------------------
DUE TO     ISTAT(7)  =   0.5933548306D+02   STAT(9)  =   0.1196733896D+02
REGRESSION I
           I
DEVIATION  ISTAT(8)  =   0.4958118361D+01
 FROM      I
REGRESSION I
           I
LACK OF    ISTAT(16) =   0.3693727871D+01   STAT(19) =   0.6715868855D+00
FITNESS    I
           I
BETWEEN    ISTAT(17) =   0.1760416667D+02   STAT(20) =   0.3200757576D+01
           I
WITHIN     ISTAT(18) =   0.5500000000D+01

*STATISTICS*

 WEIGHTED MEAN OF X            : STAT(21) =  0.704167D+02
 GRAND MEAN OF Y               : STAT(22) =  0.329167D+02
 WEIGHTED STANDARD
     DEVIATION OF X            : STAT(23) =  0.111719D+02
 STANDARD DEVIATION OF Y       : STAT(24) =  0.314667D+01
 WEIGHTED SUM OF X             : STAT(25) =  0.845000D+03
 SUM OF Y                      : STAT(26) =  0.395000D+03
 WEIGHTED SUM OF SQUARES OF X  : STAT(27) =  0.608750D+05
 SUM OF SQUARES OF Y           : STAT(28) =  0.131110D+05
 SUM OF PRODUCTS OF X AND Y    : STAT(29) =  0.281000D+05
 WEIGHTED SUM OF SQUARES OF
     DEVIATIONS OF X           : STAT(30) =  0.137292D+04
 SUM OF SQUARES OF
     DEVIATIONS OF Y           : STAT(32) =  0.108917D+03
 SUM OF PRODUCTS OF
     DEVIATIONS OF X AND Y     : STAT(32) =  0.285417D+03
 TOTAL NUMBER OF Y             : STAT(33) =  0.120000D+02
```

## 10.2.3 DNLNMA, RNLNMA Multiple Regression Analysis

(1) **Function**

Assume that $n$ independent variable values $x_{ki}$ $(k = 1, 2, \cdots, n; i = 1, 2, \cdots, m)$ for $m$ variables and $n$ given dependent variable values $y_k$ $(k = 1, 2, \cdots, n)$ corresponding to them have been given and obey the following linear regression model.

$$y_k = \beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2} + \cdots + \beta_m x_{km} + \varepsilon_k \quad (k = 1, 2, \cdots, n)$$

Here, $\varepsilon_k$ are error terms that independently obey $N(0, \sigma^2)$.

The DNLNMA or RNLNMA obtains:

(a) the estimates $b_i$ $(i = 0, 1, \cdots, m)$ of partial correlation coefficients $\beta_i$ $(i = 0, 1, \cdots, m)$

(b) the statistics for creating the following analysis of variance table due to the linear regression.

| | Sum of squares | Degrees of freedom | Unbiased variance | F ratio |
|---|---|---|---|---|
| Total variation | $S_T$ | $n-1$ | | |
| Variation due to regression | $S_R$ | m | $V_R = \dfrac{S_R}{m}$ | $F = \dfrac{V_R}{V_E}$ |
| Residual variation | $S_E = S_T - S_R$ | $n-m-1$ | $V_E = \dfrac{S_E}{n-m-1}$ | |

(c) the multiple correlation coefficient $R$

$$R = \sqrt{\frac{S_R}{S_T}}$$

(d) the degree of freedom adjusted correlation coefficient $R^*$

$$R^* = \sqrt{1 - \frac{(n-1)S_E}{(n-m-1)S_T}}$$

(e) the standard deviation estimates $\varepsilon_i$ $(i = 0, 1, \cdots, m)$ of the statistics $b_i$ $(i = 0, 1, \cdots, m)$

(f) the test quantities $t_i (i = 0, 1, \cdots, m)$, which obey $t$ distributions with $n - m - 1$ degrees of freedom based on the null hypotheses $H_0 : \beta_i = 0$ $(i = 0, 1, \cdots, m)$

(2) **Usage**

Double precision:

CALL DNLNMA (X, MX, M, N, B, R, V, STAT, IW1, W1, IERR)

Single precision:

CALL RNLNMA (X, MX, M, N, B, R, V, STAT, IW1, W1, IERR)

## (3) **Arguments**

D:Double precision real  Z:Double precision complex  I: { INTEGER(4) as for 32bit Integer / INTEGER(8) as for 64bit Integer }
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | X | {D / R} | MX, M+1 | Input | Independent variate observation data $x_{ki}$ and dependent variate observation data $y_k$ (See Note (a)) |
| 2 | MX | I | 1 | Input | Adjustable dimension of array X |
| 3 | M | I | 1 | Input | Number of independent variates $m$ |
| 4 | N | I | 1 | Input | Number of observed values $n$ |
| 5 | B | {D / R} | $0:\mathrm{M}, 3$ | Output | B($i$, 1): $\beta_i$ estimate $b_i$<br>B($i$, 2): $\beta_i$ standard deviation estimate $\varepsilon_i$<br>B($i$, 3): $\beta_i$ test quantity $t$ value $t_i$<br>$(i = 0, 1, \cdots, \mathrm{M})$ |
| 6 | R | {D / R} | 3 | Output | R(1): Multiple correlation coefficient $R$<br>R(2): Contribution ratio $R^2$<br>R(3): Degree of freedom adjusted correlation coefficient $R^*$ |
| 7 | V | {D / R} | 9 | Output | Statistics for creating analysis of variance table (See Note (b)) |
| 8 | STAT | {D / R} | M+1, 5 | Output | Basic statistics of calculation result (See Note (c)) |
| 9 | IW1 | I | M+1 | Work | Work area |
| 10 | W1 | {D / R} | See Contents | Work | Work area<br>**Size:** $(\mathrm{M}+1) \times (\mathrm{M}+3)$ |
| 11 | IERR | I | 1 | Output | Error indicator |

## (4) **Restrictions**

(a) $3 \leq \mathrm{N} \leq \mathrm{MX}$

(b) $1 \leq \mathrm{M} < \mathrm{N} - 1$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Total sum of squares $\leq$ Regression sum of squares. | 0.0 is set for $b_i$ and $t_i$. |
| 2000 | Residual unbiased variance > Total unbiased variance. | 0.0 is set for $R^*$. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 3010 | Restriction (b) was not satisfied. | |
| 4000 | The inverse matrix could not be obtained. | |

(6) **Notes**

(a) Observation data $x_{ki}$ $(k = 1, 2, \cdots, n; i = 1, 2, \cdots, m)$ and $y_k$ $(k = 1, 2, \cdots, n)$ are stored in array X as the following kind of real matrix (two-dimensional array type) data (See Appendix A).

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1,m} & y_1 \\ x_{21} & \ddots & & \vdots & y_2 \\ \vdots & & \ddots & \vdots & \vdots \\ x_{n1} & \cdots & \cdots & x_{nm} & y_n \end{bmatrix}$$

(b) The following kind of a one-dimensional vector is stored in array V.

| | | |
|---|---|---|
| 1: | $S_T$ | Total variation |
| 2: | $S_R$ | Variation due to regression |
| 3: | $S_E$ | Residual variation |
| 4: | $f_T = n - 1$ | Degrees of freedom of total variation |
| 5: | $f_R = m$ | Degrees of freedom of variation due to regression |
| 6: | $f_E = n - m - 1$ | Degrees of freedom of residual variation |
| 7: | $V_R$ | Unbiased variance of variation due to regression |
| 8: | $V_E$ | Unbiased variance of residual variation |
| 9: | $F$ | F ratio |

(c) The following kind of real matrix (two-dimensional array type) is stored in array STAT (See Appendix A).

$$\begin{bmatrix} \sum_{k=1}^{n} x_{k1} & \sum_{k=1}^{n} x_{k2} & \cdots & \sum_{k=1}^{n} x_{k,m} & \sum_{k=1}^{n} y_k \\ \mu_1 & \mu_2 & \cdots & \mu_m & \nu \\ \sum_{k=1}^{n}(x_{k1}-\mu_1)^2 & \sum_{k=1}^{n}(x_{k2}-\mu_2)^2 & \cdots & \sum_{k=1}^{n}(x_{km}-\mu_m)^2 & \sum_{k=1}^{n}(y_k-\nu)^2 \\ \frac{\sum_{k=1}^{n}(x_{k1}-\mu_1)^2}{n-1} & \frac{\sum_{k=1}^{n}(x_{k2}-\mu_2)^2}{n-1} & \cdots & \frac{\sum_{k=1}^{n}(x_{km}-\mu_m)^2}{n-1} & \frac{\sum_{k=1}^{n}(y_k-\nu)^2}{n-1} \\ \sqrt{\frac{\sum_{k=1}^{n}(x_{k1}-\mu_1)^2}{n-1}} & \sqrt{\frac{\sum_{k=1}^{n}(x_{k2}-\mu_2)^2}{n-1}} & \cdots & \sqrt{\frac{\sum_{k=1}^{n}(x_{km}-\mu_m)^2}{n-1}} & \sqrt{\frac{\sum_{k=1}^{n}(y_k-\nu)^2}{n-1}} \end{bmatrix}$$

## (7) **Example**

(a) Problem

Obtain the regression coefficient values, multiple correlation coefficient value, contribution ratio, analysis of variance table, and basic statistics according to a regression analysis from the following observation data.

$$(x_{ki}|y_k) = \left[\begin{array}{rrr|r} 2.0 & 1.0 & 3.0 & -1.0 \\ 3.0 & 3.0 & -1.0 & 0.0 \\ 4.0 & -2.0 & 0.0 & 2.0 \\ 1.0 & 1.0 & -2.0 & 2.0 \\ -1.0 & 3.0 & -1.0 & 3.0 \end{array}\right]$$

(b) Input data

Observation data X, MX=5, N=5 and M=3.

(c) Main program

```
      PROGRAM BNLNMA
! *** EXAMPLE OF DNLNMA ***
      IMPLICIT REAL(8) (A-H,O-Z)
      DIMENSION IW1(4),X(5,4),B(0:3,3),R(3),V(9),STAT(4,5),W1(24)
!
          READ(5,*) MX,N,M
          READ(5,*) ((X(I,J),I=1,N),J=1,M+1)
!
          WRITE(6,1000) MX,N,M
          WRITE(6,1010) (I,X(I,M+1),(X(I,J),J=1,M),I=1,N)
!
          CALL DNLNMA(X,MX,M,N,B,R,V,STAT,IW1,W1,IERR)
!
          WRITE(6,3000) IERR
          WRITE(6,3010) (I,(B(I,J),J=1,3),I=0,M)
          WRITE(6,3030) (R(I),I=1,3)
          IV4=V(4)
          IV5=V(5)
          IV6=V(6)
          WRITE(6,3040)
          WRITE(6,3041) V(1),IV4
          WRITE(6,3042) V(2),IV5
          WRITE(6,3043) V(3),IV6
          WRITE(6,3044)
          WRITE(6,3045) V(7),V(9)
          WRITE(6,3046) V(8)
          WRITE(6,3050)
          WRITE(6,3060)&
        (I,STAT(I,1),STAT(I,2),STAT(I,3),STAT(I,4),STAT(I,5),I=1,M+1)
      STOP
!
 1000 FORMAT(1X,/,1X,'**** INPUT DATA ****',&
             /,/,5X,'MX =',I3,5X,'N =',I3,5X,'M  =',I3,&
             /,/,6X,'  N',' I',5X,'Y',4X,'I',14X,'X',&
              /,7X,'---+',10('-'),'+',29('-'))
 1010 FORMAT(6X,I3,' I',F9.3,' I',3F9.3)
 3000 FORMAT(1X,/,/,2X,'*** OUTPUT ***',/,/,4X,'IERR=',I4,&
           /,/,4X,'*REGRESSION COEFFICIENT*',&
           /,/,27X,'B(I,1)',9X,'B(I,2)',9X,'B(I,3)',&
            /,14X,'I    I',8X,'R.C.',11X,'S.E.',11X,'T-V.',&
           /,11X,8('-'),'+',45('-'))
 3010 FORMAT(14X,I1,'    I',3F15.7)
 3030 FORMAT(1X,/,/,10X,'R(1)(MULTIPLE CORRELATION COEFFICIENT)',10X,&
                    '=',F10.7,&
           /,10X,'R(2)(CONTRIBUTION RATIO)',24X,'=',F10.7,&
           /,10X,'R(3)(ADJUSTED MULTIPLE CORRELATION COEFFICIENT) =',&
                    F10.7)
 3040 FORMAT(1X,/,/,3X,'*ANALYSIS OF VARIANCE TABLE*',&
             /,/,10X,'FACTOR OF',/,10X,'VARIATION',2X,'ISUM OF SQUARES',&
                5X,'DEGREES OF FREEDOM',/,10X,11('-'),'+',37('-'))
 3041 FORMAT(10X,'TOTAL',6X,'IV(1) =',F10.7,3X,'V(4) =',I3,/,21X,'I')
 3042 FORMAT(10X,'DUE TO',5X,'IV(2) =',F10.7,3X,'V(5) =',I3,&
               /,10X,'REGRESSION I',/,21X,'I')
 3043 FORMAT(10X,'DEVIATION  IV(3) =',F10.7,3X,'V(6) =',I3,&
               /,11X,'FROM',6X,'I',/,10X,'REGRESSION I')
 3044 FORMAT(1X,/,/,10X,'FACTOR OF',/,10X,'VARIATION',2X,&
                'IUNBIASED VARIANCE',5X,'F-RATIO',&
               /,10X,11('-'),'+',37('-'))
 3045 FORMAT(10X,'DUE TO',5X,'IV(7) =',F10.7,3X,'V(9) =',F10.7,&
               /,10X,'REGRESSION I',/,21X,'I')
 3046 FORMAT(10X,'DEVIATION  IV(8) =',F10.7,&
               /,11X,'FROM',6X,'I',/,10X,'REGRESSION I')
 3050 FORMAT(1X,/,/,3X,'*STATISTICS*',&
             /,/,18X,'STAT(I,1)',4X,'STAT(I,2)',4X,'STAT(I,3)',&
                   4X,'STAT(I,4)',4X,'STAT(I,5)',&
```

```
                /,8X,'I    I',6X,'SUM',10X,'MEAN',&
                 8X,'SUM OF',6X,'VARIANCE',5X,'STANDARD',&
                /,13X,'I',31X,'SQUARES',18X,'DEVIATION',&
                /,5X,8('-'),'+',65('-'))
 3060 FORMAT(8X,I1,4X,'I',5F13.7)
      END
```

(d) Output results

```
**** INPUT DATA ****

    MX =  5     N =  5     M  =  3

      N I     Y    I              X
     ---+----------+-----------------------------
      1 I   -1.000 I    2.000    1.000    3.000
      2 I    3.000 I    3.000   -1.000    0.000
      3 I    4.000 I   -2.000    0.000    2.000
      4 I    1.000 I    1.000   -2.000    2.000
      5 I   -1.000 I    3.000   -1.000    3.000

 *** OUTPUT ***

   IERR=   0

   *REGRESSION COEFFICIENT*

                          B(I,1)         B(I,2)         B(I,3)
          I    I           R.C.           S.E.           T-V.
        --------+---------------------------------------------
          0    I       5.2055950      0.8979475      5.7972153
          1    I      -0.7442274      0.1916032     -3.8842112
          2    I       0.0386323      0.3714016      0.1040177
          3    I      -1.4702487      0.3410923     -4.3104124

         R(1)(MULTIPLE CORRELATION COEFFICIENT)          = 0.9852786
         R(2)(CONTRIBUTION RATIO)                        = 0.9707738
         R(3)(ADJUSTED MULTIPLE CORRELATION COEFFICIENT) = 0.9397315

  *ANALYSIS OF VARIANCE TABLE*

         FACTOR OF
         VARIATION  ISUM OF SQUARES     DEGREES OF FREEDOM
         -----------+------------------------------------
         TOTAL      IV(1) =20.8000000   V(4) =  4
                    I
         DUE TO     IV(2) =20.1920959   V(5) =  3
         REGRESSION I
                    I
         DEVIATION  IV(3) = 0.6079041   V(6) =  1
          FROM      I
         REGRESSION I

         FACTOR OF
         VARIATION  IUNBIASED VARIANCE     F-RATIO
         -----------+------------------------------------
         DUE TO     IV(7) = 6.7306986   V(9) =11.0719747
         REGRESSION I
                    I
         DEVIATION  IV(8) = 0.6079041
          FROM      I
         REGRESSION I

  *STATISTICS*

                   STAT(I,1)    STAT(I,2)    STAT(I,3)    STAT(I,4)    STAT(I,5)
      I    I         SUM          MEAN        SUM OF      VARIANCE     STANDARD
           I                                  SQUARES                  DEVIATION
    --------+-----------------------------------------------------------------
      1    I    7.0000000    1.4000000   17.2000000    4.3000000    2.0736441
      2    I   -3.0000000   -0.6000000    5.2000000    1.3000000    1.1401754
      3    I   10.0000000    2.0000000    6.0000000    1.5000000    1.2247449
      4    I    6.0000000    1.2000000   20.8000000    5.2000000    2.2803509
```

## 10.3 NONLINEAR REGRESSION ANALYSIS

### 10.3.1 DNNLPO Polynomial Regression Analysis

(1) **Function**

Assume that $n$ independent variable values $x_k$ $(k = 1, 2, \cdots, n)$ and $n$ given dependent variable values $y_k$ $(k = 1, 2, \cdots, n)$ corresponding to them have been given and obey the following linear regression model.

$$y_k = \beta_0 + \beta_1 x_k + \beta_2 {x_k}^2 + \cdots + \beta_m {x_k}^m + \varepsilon_k \quad (k = 1, 2, \cdots, n)$$

Here, $\varepsilon_k$ are error terms that independently obey $N(0, \sigma^2)$.

The DNNLPO obtains:

(a) the estimates $b_i$ $(i = 0, 1, \cdots, m)$ of partial correlation coefficients $\beta_i$ $(i = 0, 1, \cdots, m)$

(b) the statistics for creating the following analysis of variance table due to the regression.

| | Sum of squares | Degrees of freedom | Unbiased variance | F ratio |
|---|---|---|---|---|
| Total variation | $S_T$ | $n-1$ | | |
| Variation due to regression | $S_R$ | m | $V_R = \dfrac{S_R}{m}$ | $F = \dfrac{V_R}{V_E}$ |
| Residual variation | $S_E = S_T - S_R$ | $n-m-1$ | $V_E = \dfrac{S_E}{n-m-1}$ | |

(c) the multiple correlation coefficient $R$

$$R = \sqrt{\frac{S_R}{S_T}}$$

(d) the degree of freedom adjusted correlation coefficient $R^*$

$$R^* = \sqrt{1 - \frac{(n-1)S_E}{(n-m-1)S_T}}$$

(e) the standard deviation estimates $\varepsilon_i$ $(i = 0, 1, \cdots, m)$ of the statistics $b_i$ $(i = 0, 1, \cdots, m)$

(f) the test quantities $t_i (i = 0, 1, \cdots, m)$, which obey $t$ distributions with $n - m - 1$ degrees of freedom based on the null hypotheses $H_0 : \beta_i = 0$ $(i = 0, 1, \cdots, m)$.

(2) **Usage**

Double precision:

CALL DNNLPO (X, N, Y, M, B, R, V, STATX, STATY, IW1, W1, IERR)

Single precision:

Nothing

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: { INTEGER(4) as for 32bit Integer
R:Single precision real  C:Single precision complex  INTEGER(8) as for 64bit Integer }

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | X | D | N | Input | Independent variate observed values $x_k$ |
| 2 | N | I | 1 | Input | Number of observed values $n$ |
| 3 | Y | D | N | Input | Dependent variate observed values $y_i$ |
| 4 | M | I | 1 | Input | Polynomial degree $m$ |
| 5 | B | D | 0:M, 3 | Output | B($i$, 1): $\beta_i$ estimate $b_i$<br>B($i$, 2): $\beta_i$ standard deviation estimate $\varepsilon_i$<br>B($i$, 3): $\beta_i$ test quantity $t$ value $t_i$<br>$(i = 0, 1, \cdots, \mathrm{M})$ |
| 6 | R | D | 3 | Output | R(1): Multiple correlation coefficient $R$<br>R(2): Contribution ratio $R^2$<br>R(3): Degree of freedom adjusted correlation coefficient $R^*$ |
| 7 | V | D | 9 | Output | Statistics for creating analysis of variance table (See Note (a)) |
| 8 | STATX | D | 5 | Output | STATX(1): Independent variate sum<br>STATX(2): Independent variate mean<br>STATX(3): Independent variate sum of squares of deviations<br>STATX(4): Independent variate variance<br>STATX(5): Independent variate standard deviation |
| 9 | STATY | D | 5 | Output | STATY(1): Dependent variate sum<br>STATY(2): Dependent variate mean<br>STATY(3): Dependent variate sum of squares of deviations<br>STATY(4): Dependent variate variance<br>STATY(5): Dependent variate standard deviation |
| 10 | IW1 | I | M+1 | Work | Work area |
| 11 | W1 | D | See Contents | Work | Work area<br>**Size:** (M+1)×(M+4) |
| 12 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $2 \leq \mathrm{M} + 1 < \mathrm{N}$

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 1000 | Total sum of squares $\leq$ Regression sum of squares. | 0.0 is set for $b_i$ and $t_i$. |
| 2000 | Residual unbiased variance $\geq$ Total unbiased variance. | 0.0 is set for $R^*$. |
| 3000 | Restriction (a) was not satisfied. | Processing is aborted. |
| 4000 | The inverse matrix could not be obtained. | |

(6) **Notes**

(a) The following kind of a one-dimensional vector is stored in array V.

| | | |
|---|---|---|
| 1: | $S_T$ | Total variation |
| 2: | $S_R$ | Variation due to regression |
| 3: | $S_E$ | Residual variation |
| 4: | $f_T = n - 1$ | Degrees of freedom of total variation |
| 5: | $f_R = m$ | Degrees of freedom of variation due to regression |
| 6: | $f_E = n - m - 1$ | Degrees of freedom of residual variation |
| 7: | $V_R$ | Unbiased variance of variation due to regression |
| 8: | $V_E$ | Unbiased variance of residual variation |
| 9: | $F$ | F ratio |

(7) **Example**

(a) Problem

Obtain the regression coefficient values, multiple correlation coefficient value, contribution ratio, analysis of variance table, and basic statistics according to a regression analysis from the following observation data.

| | | | |
|---|---|---|---|
| X( 1) = | 1.0 | Y( 1) = | 10.0 |
| X( 2) = | 3.0 | Y( 2) = | 20.0 |
| X( 3) = | 5.0 | Y( 3) = | 25.0 |
| X( 4) = | 6.0 | Y( 4) = | 26.0 |
| X( 5) = | 8.0 | Y( 5) = | 36.0 |
| X( 6) = | 10.0 | Y( 6) = | 62.0 |
| X( 7) = | 11.0 | Y( 7) = | 78.0 |
| X( 8) = | 13.0 | Y( 8) = | 107.0 |
| X( 9) = | 14.0 | Y( 9) = | 118.0 |
| X(10) = | 15.0 | Y(10) = | 127.0 |

(b) Input data

Observation data X and Y, N=10 and M=4.

(c) Main program

```
      PROGRAM BNNLPO
! *** EXAMPLE OF DNNLPO ***
      IMPLICIT REAL(8) (A-H,O-Z)
!
      PARAMETER(N=10,M=4)
      DIMENSION X(10),Y(10),B(0:M,3)
      DIMENSION R(3),V(9),STATX(5),STATY(5)
      DIMENSION IW1(N+1),W1((N+1)*(N+4))
!
      READ(5,*) (X(I),I=1,N)
      READ(5,*) (Y(I),I=1,N)
      WRITE(6,1000) N,M
      WRITE(6,3000)
      DO 10 I=1,N
        WRITE(6,3100) I,X(I),Y(I)
   10 CONTINUE
      CALL DNNLPO(X,N,Y,M,B,R,V,STATX,STATY,IW1,W1,IERR)
      WRITE(6,4000) IERR
      WRITE(6,4010)
      WRITE(6,4020) (I,(B(I,J),J=1,3),I=0,M)
      WRITE(6,4040) (R(I),I=1,3)
      IV4=V(4)
      IV5=V(5)
      IV6=V(6)
      WRITE(6,4050)
      WRITE(6,4060) V(1),IV4
      WRITE(6,4070) V(2),IV5
      WRITE(6,4080) V(3),IV6
      WRITE(6,4090)
      WRITE(6,4100) V(7),V(9)
      WRITE(6,4110) V(8)
      WRITE(6,4120)
      WRITE(6,4130) (STATX(I),I=1,5)
      WRITE(6,4140) (STATY(I),I=1,5)
!
      STOP
!
 1000 FORMAT('  *** DNNLPO ***',/,/,'    ** INPUT DATA **',/,/,&
             4X,'N = ',I4,3X,'M = ',I4)
 3000 FORMAT(' ',/,4X,'OBSERVATION VALUE',/,/,5X,'N',11X,'X',18X,'Y')
 3100 FORMAT(4X,I2,3X,D16.9,3X,D16.9)
 4000 FORMAT(4X,'** OUTPUT **',/,/,4X,'IERR = ',I4)
 4010 FORMAT(/,/,4X,'*REGRESSION COEFFICIENT*',&
             /,/,21X,'B(I,1)',12X,'B(I,2)',12X,'B(I,3)',&
              /,8X,'I    I',8X,'R.C.',14X,'S.E.',14X,'T-V.',&
             /,5X,8('-'),'+',55('-'))
 4020 FORMAT(7X,I2,'    I',3D18.10)
 4030 FORMAT(11X,'CONSTANT','I',3D18.10)
 4040 FORMAT(' ',/,/,5X,'R(1)(MULTIPLE CORRELATION COEFFICIENT)',10X,&
                      '=',D18.10,&
             /,5X,'R(2)(CONTRIBUTION RATIO)',24X,'=',D18.10,&
             /,5X,'R(3)(ADJUSTED MULTIPLE CORRELATION COEFFICIENT) =',&
                       D18.10)
 4050 FORMAT(' ',/,/,4X,'*ANALYSIS OF VARIANCE TABLE*',&
             /,/,5X,'FACTOR OF',/,5X,'VARIATION',2X,'ISUM OF SQUARES',&
                 13X,'DEGREES OF FREEDOM',/,5X,11('-'),'+',52('-'))
 4060 FORMAT(5X,'TOTAL',6X,'IV(1) =',D18.10,3X,'V(4) =',I3,/,16X,'I')
 4070 FORMAT(5X,'DUE TO',5X,'IV(2) =',D18.10,3X,'V(5) =',I3,&
                /,5X,'REGRESSION I',/,16X,'I')
 4080 FORMAT(5X,'DEVIATION  IV(3) =',D18.10,3X,'V(6) =',I3,&
                /,6X,'FROM',6X,'I',/,5X,'REGRESSION I')
 4090 FORMAT(' ',/,/,5X,'FACTOR OF',/,5X,'VARIATION',2X,&
                 'IUNBIASED VARIANCE',10X,'F-RATIO',&
                /,5X,11('-'),'+',52('-'))
 4100 FORMAT(5X,'DUE TO',5X,'IV(7) =',D18.10,3X,'V(9) =',D18.10,&
                /,5X,'REGRESSION I',/,16X,'I')
 4110 FORMAT(5X,'DEVIATION  IV(8) =',D18.10,&
                /,6X,'FROM',6X,'I',/,5X,'REGRESSION I')
 4120 FORMAT(' ',/,/,4X,'*STATISTICS*',&
               /,/,8X,'I    I 1.SUM',8X,'2.MEAN',&
                 7X,'3.SUM OF',5X,'4.VARIANCE',3X,'5.STANDARD',&
                /,13X,'I',29X,'SQUARES',19X,'DEVIATION',&
                /,5X,8('-'),'+',65('-'))
 4130 FORMAT(6X,'STATX  I',5(D13.6))
 4140 FORMAT(6X,'STATY  I',5(D13.6))
!
      END
```

(d) Output results

```
  *** DNNLPO ***

    ** INPUT DATA **

    N =   10   M =    4

    OBSERVATION VALUE

     N           X                   Y
     1    0.100000000D+01    0.100000000D+02
     2    0.300000000D+01    0.200000000D+02
     3    0.500000000D+01    0.250000000D+02
```

```
 4    0.600000000D+01    0.260000000D+02
 5    0.800000000D+01    0.360000000D+02
 6    0.100000000D+02    0.620000000D+02
 7    0.110000000D+02    0.780000000D+02
 8    0.130000000D+02    0.107000000D+03
 9    0.140000000D+02    0.118000000D+03
10    0.150000000D+02    0.127000000D+03
** OUTPUT **

IERR =    0


*REGRESSION COEFFICIENT*

                 B(I,1)            B(I,2)            B(I,3)
    I    I        R.C.              S.E.              T-V.
 --------+---------------------------------------------------------
    0    I -0.4607861598D+01  0.1568666444D+01 -0.2937438751D+01
    1    I  0.1891301677D+02  0.3032062662D+01  0.6237673449D+01
    2    I -0.4973007854D+01  0.7643575879D+00 -0.6506127411D+01
    3    I  0.5505362721D+00  0.7155311132D-01  0.7694092709D+01
    4    I -0.1761845964D-01  0.2223941477D-02 -0.7922177728D+01


 R(1)(MULTIPLE CORRELATION COEFFICIENT)          =  0.9996471755D+00
 R(2)(CONTRIBUTION RATIO)                        =  0.9992944755D+00
 R(3)(ADJUSTED MULTIPLE CORRELATION COEFFICIENT) =  0.9993648262D+00


*ANALYSIS OF VARIANCE TABLE*

 FACTOR OF
 VARIATION  ISUM OF SQUARES            DEGREES OF FREEDOM
 -----------+----------------------------------------------
 TOTAL      IV(1) =  0.1743890000D+05   V(4) =  9
            I
 DUE TO     IV(2) =  0.1742659643D+05   V(5) =  4
 REGRESSION I
            I
 DEVIATION  IV(3) =  0.1230357206D+02   V(6) =  5
  FROM      I
 REGRESSION I


 FACTOR OF
 VARIATION  IUNBIASED VARIANCE         F-RATIO
 -----------+----------------------------------------------
 DUE TO     IV(7) =  0.4356649107D+04   V(9) =  0.1770481405D+04
 REGRESSION I
            I
 DEVIATION  IV(8) =  0.2460714411D+01
  FROM      I
 REGRESSION I


*STATISTICS*

    I    I 1.SUM        2.MEAN       3.SUM OF     4.VARIANCE   5.STANDARD
         I                             SQUARES                   DEVIATION
 --------+------------------------------------------------------------------
  STATX  I 0.860000D+02 0.860000D+01 0.206400D+03 0.229333D+02 0.478888D+01
  STATY  I 0.609000D+03 0.609000D+02 0.174389D+05 0.193766D+04 0.440188D+02
```

## 10.3.2 DNNLGF, RNNLGF
## Regression According to an Arbitrary Function

(1) **Function**

Assume that $n$ sets of independent variables $\boldsymbol{x}_k = (x_{1k}, x_{2k}, \cdots, x_{mk})$ $(k = 1, 2, \cdots, n)$ consisting of $m$ variables and $n$ given dependent variable values $y_k$ $(k = 1, 2, \cdots, n)$ corresponding to them have been given and obey the following regression model.

$$y_k = f(\boldsymbol{x}_k; \boldsymbol{\beta}) + \varepsilon_k \ (k = 1, 2, \cdots, n)$$

Here, $\varepsilon_k$ are error terms that independently obey $N(0, \sigma^2)$).

The DNNLGF or RNNLGF obtains:

(a) the estimates $\boldsymbol{b} = (b_1, b_2, \cdots, b_\ell)$ of the $\ell$ regression model parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_\ell)$

(b) the statistics for performing an analysis of variance for the regression model

| | Sum of squares | Degrees of freedom | Unbiased variance |
|---|---|---|---|
| Total variation | $S_T$ | $n$ | |
| Residual variation | $S_E$ | $n - \ell$ | $V_E = \dfrac{S_E}{n - \ell}$ |

(c) the asymptotic variance-covariance matrix $V = (V_{ij})$ $(i, j = 1, \cdots, \ell)$

(d) the standard deviation estimates $\varepsilon_i$ $(i = 1, \cdots, \ell)$ of the statistics $b_i$ $(i = 1, \cdots, \ell)$

(2) **Usage**

Double precision:

CALL DNNLGF (F, XD, NA, NN, NM, YD, NL, ER, NEV, X, XE, Y, C, NV, V, STAT, IW1, W1, IERR)

Single precision:

CALL RNNLGF (F, XD, NA, NN, NM, YD, NL, ER, NEV, X, XE, Y, C, NV, V, STAT, IW1, W1, IERR)

(3) **Arguments**

D:Double precision real  Z:Double precision complex  I: { INTEGER(4) as for 32bit Integer ; INTEGER(8) as for 64bit Integer }
R:Single precision real  C:Single precision complex

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 1 | F | {D, R} | – | Input | Function subprogram name of the function F(X, B) for defining the regression model $f(\boldsymbol{x}, \boldsymbol{\beta})$. |
| 2 | XD | {D, R} | NA,NM | Input | Independent variate observation values $\boldsymbol{x}_k$ (See Note (a)) |
| 3 | NA | I | 1 | Input | Adjustable dimension of array X |
| 4 | NN | I | 1 | Input | Number of observed values $m$ |
| 5 | NM | I | 1 | Input | Number of independent variables $n$ |
| 6 | YD | {D, R} | NN | Input | Dependent variate observation values $y_i$ |
| 7 | NL | I | 1 | Input | Number of parameters $\ell$ of regression model function $f(\boldsymbol{x}; \boldsymbol{\beta})$. |
| 8 | ER | {D, R} | 1 | Input | Requested precision (default value:$2\times\sqrt{(\text{units for error decision})}$) |
| 9 | NEV | I | 1 | Input | Maximum number $n$ of evaluations of function $f(\boldsymbol{x}, \boldsymbol{\beta})$ (default value:100×NN×NL) |
| | | | | Output | Actual number of function evaluations |
| 10 | X | {D, R} | NL | Input | Initial values of $\beta_i$ |
| | | | | Output | Estimated values $b_i$ of $\beta_i$ $(i =, 1, \cdots, \text{NL})$ |
| 11 | XE | {D, R} | NL | Output | Estimated standard deviation $\varepsilon_i$ of $b_i$ $(i = 1, \cdots, \text{NL})$ |
| 12 | Y | {D, R} | NN | Output | Function value $f(\boldsymbol{x}_i, \boldsymbol{b})$ according to the optimal least square solution. |

| No. | Argument | Type | Size | Input/ Output | Contents |
|---|---|---|---|---|---|
| 13 | C | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | NV,NL | Output | Asymptotic variance-covariance matrix $V$ for parameter $\boldsymbol{\beta}$. |
| 14 | NV | I | 1 | Input | Adjustable dimension of array CV. |
| 15 | V | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | 5 | Output | Statistics for creating analysis of variance table. (See Note (c)) |
| 16 | STAT | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | NL+1,5 | Output | Basic statistics of calculation result. (See Note (b)) |
| 17 | IW1 | I | $4 \times$ NL | Work | Work area |
| 18 | W1 | $\begin{Bmatrix} D \\ R \end{Bmatrix}$ | See Contents | Work | Work area<br>**Size:**NN $\times$ (2 $\times$ NL + 1) + NL $\times$ (NL + 4) |
| 19 | IERR | I | 1 | Output | Error indicator |

(4) **Restrictions**

(a) $0 <$ NN $\leq$ NA

(b) $0 <$ NL $\leq$ NV

(c) NM $> 0$

(d) $2 \leq$ NL $+ 1 <$ NN

(5) **Error indicator**

| IERR value | Meaning | Processing |
|---|---|---|
| 0 | Normal termination. | |
| 3000 | Any of restrictions (a) to (d) was not satisfied. | Processing is aborted. |
| 4000 | The linear least squares method could not be solved. | The values of X, Y, STAT and V at that time are output. |
| 4100 | The steepest descent could not be calculated. | |
| 4200 | The solution could not be corrected 2NN times consecutively. | |
| 4300 | The inverse matrix could not be obtained. | |
| 5000 | The sequence did not converge before reaching the given maximum number of evaluations. | |

(6) **Notes**

(a) Observation data $x_{ki}$ $(k = 1, 2, \cdots, n; i = 1, 2, \cdots, m)$ are stored as a real matrix (two-dimensional array type) in array XD as follows (See Appendix A).

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1,m} \\ x_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{n1} & \cdots & \cdots & x_{nm} \end{bmatrix}$$

(b) The following kind of real matrix (two-dimensional array type) is stored in array STAT (See Appendix A).

$$\begin{bmatrix} \sum_{k=1}^{n} x_{k1} & \mu_1 & \sum_{k=1}^{n}(x_{k1}-\mu_1)^2 & \dfrac{\sum_{k=1}^{n}(x_{k1}-\mu_1)^2}{n-1} & \sqrt{\dfrac{\sum_{k=1}^{n}(x_{k1}-\mu_1)^2}{n-1}} \\ \sum_{k=1}^{n} x_{k2} & \mu_2 & \sum_{k=1}^{n}(x_{k2}-\nu)^2 & \dfrac{\sum_{k=1}^{n}(x_{k2}-\mu_2)^2}{n-1} & \sqrt{\dfrac{\sum_{k=1}^{n}(x_{k2}-\mu_2)^2}{n-1}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_{k=1}^{n} x_{km} & \mu_1 & \sum_{k=1}^{n}(x_{km}-\nu)^2 & \dfrac{\sum_{k=1}^{n}(x_{km}-\mu_m)^2}{n-1} & \sqrt{\dfrac{\sum_{k=1}^{n}(x_{km}-\mu_m)^2}{n-1}} \\ \sum_{k=1}^{n} y_k & \nu & \sum_{k=1}^{n}(y_k-\nu)^2 & \dfrac{\sum_{k=1}^{n}(y_k-\nu)^2}{n-1} & \sqrt{\dfrac{\sum_{k=1}^{n}(y_k-\nu)^2}{n-1}} \end{bmatrix}$$

Here, $\mu_i$ and $\nu$ are defined as follows.

$$\mu_i = \frac{\sum_{k=1}^{n} x_{ki}}{n}$$

$$\nu = \frac{\sum_{k=1}^{n} y_k}{n}$$

(c) The following kind of a one-dimensional vector is stored in array V.

| | | |
|---|---|---|
| 1: | $S_T$ | Total variation |
| 2: | $S_E$ | Residual variation |
| 3: | $f_T = n$ | Degrees of freedom of total variation |
| 4: | $f_E = n - \ell$ | Degrees of freedom of residual variation |
| 5: | $V_E$ | Unbiased variance of residual variation |

(d) The actual name for the first argument term F must be declared by using an EXTERNAL statement in a user-created program and a function subprogram having the actual name must be created in advance. The function subprogram (in double-precision) should be created as follows.

REAL(8) FUNCTION F(X, B)
REAL(8) B, X]
DIMENSION B(∗), X(∗)
F = $f(\boldsymbol{x}, \boldsymbol{b})$ (Expression of X and coefficients B(i))

```
      RETURN
      END
```

(e) The convergence decision is made according to the following expression by solving for $\boldsymbol{b} + \Delta\boldsymbol{b}$.

$$\|\Delta\boldsymbol{b}\| \le \mathrm{ER} \times \max(1, \|\boldsymbol{b} + \Delta\boldsymbol{b}\|)$$

Here, $\Delta\boldsymbol{b}$ is the correction vector for $\boldsymbol{b}$ and $\|\boldsymbol{b}\| = \max|b_i|$
A value on the order of the default value should be taken as ER.

(f) If a default value is written in the Contents column of the Arguments table, the default value is set if an integer less than or equal to 0 is entered for an integer type or a real number less than or equal to 0.0 is entered for a real type.

(7) **Example**

(a) Problem
From the following observation data,

$$(x_{ki}|y_k) = \left[\begin{array}{rrr|r} -2.0 & -2.0 & -1.0 & 2.7 \\ -1.5 & -1.5 & -1.0 & 2.9 \\ -1.0 & -1.0 & -1.0 & 3.1 \\ -0.5 & -1.0 & -1.5 & 3.4 \\ -0.5 & -1.5 & 1.0 & 3.9 \\ 0.0 & 0.0 & 0.0 & 4.7 \\ 0.5 & 0.25 & 0.25 & 6.0 \\ 0.5 & 1.0 & 0.5 & 7.8 \\ 1.0 & 1.0 & 1.0 & 7.9 \\ 1.5 & 1.5 & 1.0 & 6.3 \\ 1.5 & 2.0 & 1.5 & 5.2 \end{array}\right]$$

perform a regression analysis according to the following regression model

$$f(\boldsymbol{x};\boldsymbol{\beta}) = \frac{\beta_3\beta_2^2}{(x_1+x_2+x_3-\beta_1)^2+\beta_2^2} + \beta_4 + \beta_5(x_1+x_2+x_3)$$

to obtain the estimates and estimate errors of the regression parameters, asymptotic variance-covariance matrix, analysis of variance table, and basic statistics.

(b) Input data
Observation data XD and YD, NN=11, NM=3, NL=5, NEV=0, initial values of regression parameters, NEV=0 and ER=0.0.

(c) Main program

```
      PROGRAM BNNLGF
! *** EXAMPLE OF DNNLGF ***
      IMPLICIT REAL(8) (A-H,O-Z)
      PARAMETER (L = 5, N = 11, NA = 20, M = 3, NV = 5)
      DIMENSION IWK(4*L)
      DIMENSION A(5),AE(5),YF(N),WK(N*(2*L+1)+L*(L+4))
      DIMENSION X(NA,M),Y(N),V(5),C(NV,L),STAT(M+1,5)
      EXTERNAL FNDANL
!
      WRITE(6,1000)
      DO 100 I = 1,N
         READ(5,*) (X(I,J),J=1,M),Y(I)
 100  CONTINUE
      READ(5,*) NEV
      READ(5,*) ER
      READ(5,*) (A(I),I=1,L)
      WRITE(6,1100) N,M,L,NEV,ER
      DO 110 I = 1,N
         WRITE(6,1200) (X(I,J),J=1,M),Y(I)
 110  CONTINUE
      WRITE(6,1300) (I,A(I),I=1,L)
```

```
      CALL DNNLGF&
      (FNDANL,X,NA,N,M,Y,L,ER,NEV,A,AE,YF,C,NV,V,STAT,IWK,WK,IERR)
      WRITE(6,1400) &
      IERR,NEV,(I,A(I),I=1,L),(I,AE(I),I=1,L),(I,YF(I),I=1,N),&
      ((C(I,J),J=1,L),I=1,L),(I,V(I),I=1,5),&
      (I,(STAT(I,J),J=1,5),I=1,M),(STAT(M+1,J),J=1,5)
!
 1000 FORMAT(' ',/,' *** DNDANL ***')
 1100 FORMAT('  ** INPUT **',/,&
             5X,'N      =',I5,/,&
             5X,'M      =',I5,/,&
             5X,'L      =',I5,/,&
             5X,'NEV    =',I5,/,&
             5X,'ER     =',D18.10)
 1200 FORMAT(5X,'(( COORDINATES (X,Y) ))',/,&
            (5X,3(2X,F5.1),4X,F5.1))
 1300 FORMAT(5X,'(( INITIAL VALUE OF COEFFICIENTS ))',/,&
            (5X,'   A(',I2,')  =',F5.1))
 1400 FORMAT('  ** OUTPUT **',/,&
             5X,'IERR   =',I5,/,&
             5X,'NEV    =',I5,/,&
             5X,'(( OPTIMIZED COEFFICIENTS ))',/,&
           5(5X,'   A(',I2,') =',D18.10,/),&
             5X,'(( ESTIMATED ERROR OF COEFFICIENTS ))',/,&
           5(5X,'   AE(',I2,') =',D18.10,/),&
             5X,'(( FUNCTION VALUE ))',/,&
          11(5X,'   YF(',I2,') =',D18.10,/),&
             5X,'(( ASYPTOTIC VARIANCE-COVARIANCE MATRIX ))',/,&
           5(5X,5(D11.3,1X),/),&
             5X,'(( ANALISYS OF VARIANCE TABLE ))',/,&
           5(5X,'   V(',I2,')  =',D18.10,/),&
             5X,'(( STATISTICS ))',/,&
           3(5X,'   X(',I2,')  :',5(D11.3,1X),/),&
             5X,'   Y       :',5(D11.3,1X))
      END


      REAL(8) FUNCTION FNDANL(X,A)
      REAL(8) X,A,F1,F2
      DIMENSION A(*),X(*)
      REAL(8) R
!
      R = X(1) + X(2) +X(3)
      F1 = A(3)*A(2)*A(2)/((R-A(1))*(R-A(1))+A(2)*A(2))
      F2 = A(4)+A(5)*R
      FNDANL = F1+F2
      RETURN
      END
```

(d) Output results

```
*** DNDANL ***
 ** INPUT **
    N      =   11
    M      =    3
    L      =    5
    NEV    =    0
    ER     =  0.0000000000D+00
    (( COORDINATES (X,Y) ))
       -2.0   -2.0   -1.0      2.7
    (( COORDINATES (X,Y) ))
       -1.5   -1.5   -1.0      2.9
    (( COORDINATES (X,Y) ))
       -1.0   -1.0   -1.0      3.1
    (( COORDINATES (X,Y) ))
       -0.5   -1.0   -1.5      3.4
    (( COORDINATES (X,Y) ))
       -0.5   -1.5    1.0      3.9
    (( COORDINATES (X,Y) ))
        0.0    0.0    0.0      4.7
    (( COORDINATES (X,Y) ))
        0.5    0.3    0.3      6.0
    (( COORDINATES (X,Y) ))
        0.5    1.0    0.5      7.8
    (( COORDINATES (X,Y) ))
        1.0    1.0    1.0      7.9
    (( COORDINATES (X,Y) ))
        1.5    1.5    1.0      6.3
    (( COORDINATES (X,Y) ))
        1.5    2.0    1.5      5.2
    (( INITIAL VALUE OF COEFFICIENTS ))
       A( 1)  =  0.0
       A( 2)  =  1.0
       A( 3)  =  6.0
       A( 4)  =  3.5
       A( 5)  =  0.2
 ** OUTPUT **
    IERR   =    0
    NEV    =  902
    (( OPTIMIZED COEFFICIENTS ))
       A( 1) =  0.2486789111D+01
       A( 2) =  0.1745170205D+01
       A( 3) =  0.4861744252D+01
       A( 4) =  0.3073046191D+01
```

```
   A( 5) =  0.1132194723D+00
(( ESTIMATED ERROR OF COEFFICIENTS ))
   AE( 1) =  0.1051397452D+00
   AE( 2) =  0.3165868493D+00
   AE( 3) =  0.5230434840D+00
   AE( 4) =  0.4125090217D+00
   AE( 5) =  0.6950610080D-01
(( FUNCTION VALUE ))
   YF( 1) =  0.2757500684D+01
   YF( 2) =  0.2948308621D+01
   YF( 3) =  0.3180048686D+01
   YF( 4) =  0.3180048686D+01
   YF( 5) =  0.3933760271D+01
   YF( 6) =  0.4677319147D+01
   YF( 7) =  0.6003344734D+01
   YF( 8) =  0.7810268963D+01
   YF( 9) =  0.7887470677D+01
   YF(10) =  0.6301151442D+01
   YF(11) =  0.5220777954D+01
(( ASYPTOTIC VARIANCE-COVARIANCE MATRIX ))
  0.111D-01   0.183D-01   0.315D-01  -0.254D-01  -0.491D-02
  0.183D-01   0.100D+00   0.115D+00  -0.120D+00  -0.182D-01
  0.315D-01   0.115D+00   0.274D+00  -0.191D+00  -0.321D-01
 -0.254D-01  -0.120D+00  -0.191D+00   0.170D+00   0.261D-01
 -0.491D-02  -0.182D-01  -0.321D-01   0.261D-01   0.483D-02
(( ANALISYS OF VARIANCE TABLE ))
   V( 1)  =  0.3001500000D+03
   V( 2)  =  0.6278727310D-01
   V( 3)  =  0.1100000000D+02
   V( 4)  =  0.8000000000D+01
   V( 5)  =  0.7848409138D-02
(( STATISTICS ))
   X( 1)  : -0.500D+00  -0.455D-01   0.137D+02   0.137D+01   0.117D+01
   X( 2)  : -0.125D+01  -0.114D+00   0.187D+02   0.187D+01   0.137D+01
   X( 3)  :  0.750D+00   0.682D-01   0.108D+02   0.108D+01   0.104D+01
   Y      :  0.539D+02   0.490D+01   0.360D+02   0.360D+01   0.190D+01
```

# Appendix A

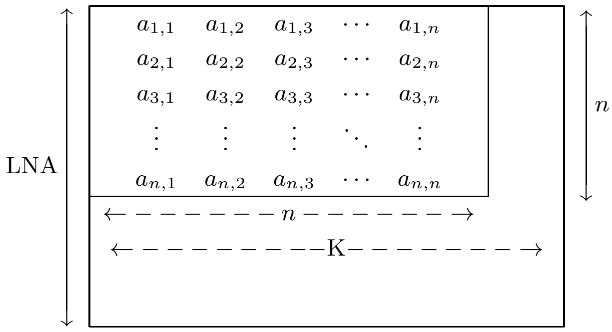
# METHODS OF HANDLING ARRAY DATA

## A.1 Methods of handling array data corresponding to matrix

Since the ASL subroutine library uses array data corresponding to matrix, this section describes various methods of handling arrays.

To call a subroutine that uses array data, you must declare that array in advance in the calling program. If the declared array is A(LNA, K), then $n \times n$ matrix $A = (a_{i,j})$ $(i = 1, 2, \cdots, n; j = 1, 2, \cdots, n)$ is stored in array A as shown in the figure below.

Matrix Storage Mode Within an Array A




**Remarks**

a. LNA $\geq n$ and K $\geq n$ must hold.

b. Matrix element $a_{i,j}$ corresponds to the array element A$(i, j)$.

Figure A−1 Matrix Storage Mode Within an Array A()

LNA is called an adjustable dimension. If a two-dimensional array is used as an argument, the adjustable must be passed to the subroutine as an argument in addition to the array name and order of the array. The matrix elements $a_{i,j}$ $(i = 1, 2, \cdots, \text{LNA}; j = 1, 2, \cdots, \text{K})$ must correspond to the array element A(i, j) $(\text{i} = 1, 2, \cdots, \text{LNA}; \text{j} = 1, 2, \cdots, \text{K})$, as follows on the main memory.

| | | | | | | |
|---|---|---|---|---|---|---|
| $a_{1,1}$ | $a_{2,1}$ | $\cdots$ | $a_{\text{LNA},1}$ | $a_{1,2}$ | $a_{2,2}$ | $\cdots$ |
| $\updownarrow$ | $\updownarrow$ | $\cdots$ | $\updownarrow$ | $\updownarrow$ | $\updownarrow$ | $\cdots$ |
| A(1,1) | A(2,1) | $\cdots$ | A(LNA,1) | A(1,2) | A(2,2) | $\cdots$ |

**Example** DAM1AD (Real matrix addition)

Add $3 \times 2$ matrices $A$ and $B$ placing the sum in matrix $C$. If you declare arrays of size (5, 4), the declaration and CALL statements are as follows.

```
      REAL(8) A(5, 4), B(5, 4), C(5, 4)
      INTEGER IERR
C
      CALL DAM1AD(A, 5, 3, 2, B, 5, C, 5, IERR)
```

Data is stored in A as follows. Data are stored in B and C in the same way.

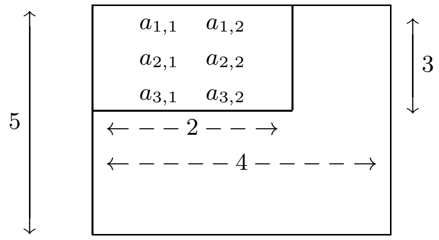


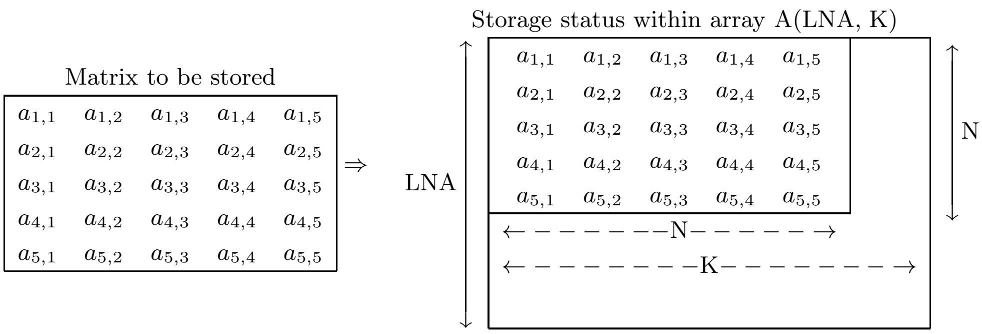

Figure A−2 Matrix Storage Mode Within an Array A

If you will be manipulating several arrays having different orders as data, you can prepare one array having LNA equal to the largest order and use that array successively for each array. However, you must always assign the LNA value as an adjustable dimension.

## A.2 Data storage modes

Matrix data storage modes differ according to the matrix type. Storage modes for each type of matrix are shown below.

### A.2.1 Real matrix (two-dimensional array type)



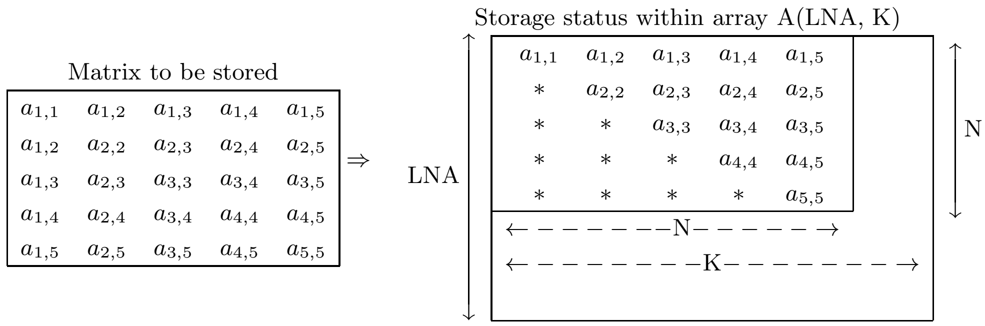

**Remarks**

a. LNA $\geq$ N and K $\geq$ N must hold.

Figure A−3 Real Matrix (Two-Dimensional Array Type) Storage Mode

### A.2.2 Real symmetric matrix and positive symmetric matrix

(1) **Two-dimensional array type, upper triangular type**




**Remarks**

a. The asterisk ($*$) indicates an arbitrary value.

b. LNA $\geq$ N and K $\geq$ N must hold.

Figure A−4 Real Symmetric Matrix (Two-dimensional Array Type) (Upper Triangular Type) Storage mode

# Appendix B

# MACHINE CONSTANTS USED IN ASL

## B.1 Units for Determining Error

The table below shows values in ASL as units for determining error in floating point calculations. The units shown in the table are numeric values determined by the internal representation of floating point data. ASL uses these units for determining convergence and zeros.

Table B−1 Units for Determining Error

| Single-precision | Double-precision |
|---|---|
| $2^{-23}(\simeq 1.19 \times 10^{-7})$ | $2^{-52}(\simeq 2.22 \times 10^{-16})$ |

**Remark:** The unit for determining error $\varepsilon$, which is also called the machine $\varepsilon$, is usually defined as the smallest positive constant for which the calculation result of $1+\varepsilon$ differs from 1 in the corresponding floating point mode. Therefore, seeing the unit for determining error enables you to know the maximum number of significant digits of an operation (on the mantissa) in that floating point mode.

## B.2 Maximum and Minimum Values of Floating Point Data

The table below shows maximum and minimum values of floating point data defined within ASL. Note that the maximum and minimum values shown below may differ from the maximum and minimum values that are actually used by the hardware for each floating point mode.

Table B−2 Maximum and Minimum Values of Floating Point Data

| | Single-precision | Double-precision |
|---|---|---|
| Maximum value | $2^{127}(2-2^{-23})$ $(\simeq 3.40 \times 10^{38})$ | $2^{1023}(2-2^{-52})$ $(\simeq 1.80 \times 10^{308})$ |
| Positive minimum value | $2^{-126}$ $(\simeq 1.17 \times 10^{-38})$ | $2^{-1022}$ $(\simeq 2.23 \times 10^{-308})$ |
| Negative maximum value | $-2^{-126}$ $(\simeq -1.17 \times 10^{-38})$ | $-2^{-1022}$ $(\simeq -2.23 \times 10^{-308})$ |
| Minimum value | $-2^{127}(2-2^{-23})$ $(\simeq -3.40 \times 10^{38})$ | $-2^{1023}(2-2^{-52})$ $(\simeq -1.80 \times 10^{308})$ |

# Index


CAM1HH : Vol.1, 85
CAM1HM : Vol.1, 82
CAM1MH : Vol.1, 79
CAM1MM : Vol.1, 76
CAN1HH : Vol.1, 97
CAN1HM : Vol.1, 94
CAN1MH : Vol.1, 91
CAN1MM : Vol.1, 88
CANVJ1 : Vol.1, 126
CARGJM : Vol.1, 37
CARSJD : Vol.1, 32
CBGMDI : Vol.2, 72
CBGMLC : Vol.2, 64
CBGMLS : Vol.2, 66
CBGMLU : Vol.2, 62
CBGMLX : Vol.2, 74
CBGMMS : Vol.2, 68
CBGMSL : Vol.2, 58
CBGMSM : Vol.2, 54
CBGNDI : Vol.2, 92
CBGNLC : Vol.2, 84
CBGNLS : Vol.2, 86
CBGNLU : Vol.2, 82
CBGNLX : Vol.2, 94
CBGNMS : Vol.2, 88
CBGNSL : Vol.2, 79
CBGNSM : Vol.2, 76
CBHEDI : Vol.2, 216
CBHELS : Vol.2, 211
CBHELX : Vol.2, 218
CBHEMS : Vol.2, 213
CBHESL : Vol.2, 203
CBHEUC : Vol.2, 209
CBHEUD : Vol.2, 207
CBHFDI : Vol.2, 199
CBHFLS : Vol.2, 194
CBHFLX : Vol.2, 201
CBHFMS : Vol.2, 196
CBHFSL : Vol.2, 186
CBHFUC : Vol.2, 192
CBHFUD : Vol.2, 190
CBHPDI : Vol.2, 165
CBHPLS : Vol.2, 160
CBHPLX : Vol.2, 167
CBHPMS : Vol.2, 162
CBHPSL : Vol.2, 152
CBHPUC : Vol.2, 158
CBHPUD : Vol.2, 156
CBHRDI : Vol.2, 182
CBHRLS : Vol.2, 177
CBHRLX : Vol.2, 184
CBHRMS : Vol.2, 179
CBHRSL : Vol.2, 169
CBHRUC : Vol.2, 175
CBHRUD : Vol.2, 173
CCGEAA : Vol.1, 160
CCGEAN : Vol.1, 164
CCGHAA : Vol.1, 318
CCGHAN : Vol.1, 323
CCGJAA : Vol.1, 325
CCGJAN : Vol.1, 329
CCGKAA : Vol.1, 331
CCGKAN : Vol.1, 335
CCGNAA : Vol.1, 166
CCGNAN : Vol.1, 169
CCGRAA : Vol.1, 311
CCGRAN : Vol.1, 316
CCHEAA : Vol.1, 205
CCHEAN : Vol.1, 208
CCHEEE : Vol.1, 216
CCHEEN : Vol.1, 220
CCHESN : Vol.1, 214
CCHESS : Vol.1, 210
CCHJSS : Vol.1, 267
CCHRAA : Vol.1, 188
CCHRAN : Vol.1, 191
CCHREE : Vol.1, 199
CCHREN : Vol.1, 203
CCHRSN : Vol.1, 197
CCHRSS : Vol.1, 193
CFC1BF : Vol.3, 58
CFC1FB : Vol.3, 54
CFC2BF : Vol.3, 117
CFC2FB : Vol.3, 113
CFC3BF : Vol.3, 145
CFC3FB : Vol.3, 141
CFCMBF : Vol.3, 87
CFCMFB : Vol.3, 83
CIBH1N : Vol.5, 142
CIBH2N : Vol.5, 144

CIBINZ : Vol.5, 127
CIBJNZ : Vol.5, 92
CIBKNZ : Vol.5, 129
CIBYNZ : Vol.5, 94
CIGAMZ : Vol.5, 179
CIGLGZ : Vol.5, 181
CLACHA : Vol.5, 345
CLNCIS : Vol.5, 361

D1CDBN : Vol.6, 72
D1CDBT : Vol.6, 114
D1CDCC : Vol.6, 147
D1CDCH : Vol.6, 75
D1CDEX : Vol.6, 132
D1CDFB : Vol.6, 100
D1CDGM : Vol.6, 107
D1CDGU : Vol.6, 135
D1CDIB : Vol.6, 117
D1CDIC : Vol.6, 78
D1CDIF : Vol.6, 104
D1CDIG : Vol.6, 111
D1CDIN : Vol.6, 69
D1CDIS : Vol.6, 97
D1CDIT : Vol.6, 91
D1CDIX : Vol.6, 85
D1CDLD : Vol.6, 138
D1CDLG : Vol.6, 144
D1CDLN : Vol.6, 141
D1CDNC : Vol.6, 81
D1CDNO : Vol.6, 66
D1CDNT : Vol.6, 94
D1CDPA : Vol.6, 126
D1CDTB : Vol.6, 88
D1CDTR : Vol.6, 123
D1CDUF : Vol.6, 120
D1CDWE : Vol.6, 129
D1DDBP : Vol.6, 150
D1DDGO : Vol.6, 154
D1DDHG : Vol.6, 159
D1DDHN : Vol.6, 162
D1DDPO : Vol.6, 157
D2BA1T : Vol.6, 173
D2BA2S : Vol.6, 179
D2BAGM : Vol.6, 191
D2BAHM : Vol.6, 199
D2BAMO : Vol.6, 195
D2BAMS : Vol.6, 186
D2BASM : Vol.6, 203
D2CCMA : Vol.6, 225
D2CCMT : Vol.6, 220
D2CCPR : Vol.6, 231
D2VCGR : Vol.6, 212
D2VCMT : Vol.6, 207
D3IECD : Vol.6, 307
D3IEME : Vol.6, 293
D3IERA : Vol.6, 290
D3IESR : Vol.6, 311
D3IESU : Vol.6, 297
D3IETC : Vol.6, 304
D3IEVA : Vol.6, 301
D3TSCD : Vol.6, 347
D3TSME : Vol.6, 326
D3TSRA : Vol.6, 317
D3TSRD : Vol.6, 321
D3TSSR : Vol.6, 350
D3TSSU : Vol.6, 331
D3TSTC : Vol.6, 342
D3TSVA : Vol.6, 338
D41WR1 : Vol.6, 363
D42WR1 : Vol.6, 383
D42WRM : Vol.6, 375
D42WRN : Vol.6, 369
D4BI01 : Vol.6, 438
D4GL01 : Vol.6, 434
D4MU01 : Vol.6, 415
D4MWRF : Vol.6, 391
D4MWRM : Vol.6, 402
D4RB01 : Vol.6, 430
D5CHEF : Vol.6, 447
D5CHMD : Vol.6, 456
D5CHMN : Vol.6, 453
D5CHTT : Vol.6, 450
D5TEMH : Vol.6, 466
D5TESG : Vol.6, 459
D5TESP : Vol.6, 470
D5TEWL : Vol.6, 462
D6CLAN : Vol.6, 518
D6CLDA : Vol.6, 523
D6CLDS : Vol.6, 513
D6CPCC : Vol.6, 482
D6CPSC : Vol.6, 484
D6CVAN : Vol.6, 495
D6CVSC : Vol.6, 498
D6DAFN : Vol.6, 503
D6DASC : Vol.6, 507
D6FALD : Vol.6, 489
D6FAVR : Vol.6, 491
DABMCS : Vol.1, 12
DABMEL : Vol.1, 15
DAM1AD : Vol.1, 47
DAM1MM : Vol.1, 64
DAM1MS : Vol.1, 56
DAM1MT : Vol.1, 67
DAM1MU : Vol.1, 53
DAM1SB : Vol.1, 50
DAM1TM : Vol.1, 70
DAM1TP : Vol.1, 109
DAM1TT : Vol.1, 73

DAM1VM : Vol.1, 100
DAM3TP : Vol.1, 111
DAM3VM : Vol.1, 103
DAM4VM : Vol.1, 106
DAMT1M : Vol.1, 59
DAMVJ1 : Vol.1, 114
DAMVJ3 : Vol.1, 118
DAMVJ4 : Vol.1, 122
DARGJM : Vol.1, 26
DARSJD : Vol.1, 21
DASBCS : Vol.1, 17
DASBEL : Vol.1, 19
DATM1M : Vol.1, 61
DBBDDI : Vol.2, 231
DBBDLC : Vol.2, 227
DBBDLS : Vol.2, 229
DBBDLU : Vol.2, 225
DBBDLX : Vol.2, 233
DBBDSL : Vol.2, 220
DBBPDI : Vol.2, 247
DBBPLS : Vol.2, 245
DBBPLX : Vol.2, 249
DBBPSL : Vol.2, 237
DBBPUC : Vol.2, 243
DBBPUU : Vol.2, 241
DBGMDI : Vol.2, 48
DBGMLC : Vol.2, 41
DBGMLS : Vol.2, 43
DBGMLU : Vol.2, 39
DBGMLX : Vol.2, 50
DBGMMS : Vol.2, 45
DBGMSL : Vol.2, 35
DBGMSM : Vol.2, 31
DBPDDI : Vol.2, 106
DBPDLS : Vol.2, 104
DBPDLX : Vol.2, 108
DBPDSL : Vol.2, 96
DBPDUC : Vol.2, 102
DBPDUU : Vol.2, 100
DBSMDI : Vol.2, 140
DBSMLS : Vol.2, 135
DBSMLX : Vol.2, 142
DBSMMS : Vol.2, 137
DBSMSL : Vol.2, 127
DBSMUC : Vol.2, 133
DBSMUD : Vol.2, 131
DBSNLS : Vol.2, 150
DBSNSL : Vol.2, 144
DBSNUD : Vol.2, 148
DBSPDI : Vol.2, 123
DBSPLS : Vol.2, 118
DBSPLX : Vol.2, 125
DBSPMS : Vol.2, 120
DBSPSL : Vol.2, 110
DBSPUC : Vol.2, 116
DBSPUD : Vol.2, 114
DBTDSL : Vol.2, 251
DBTLCO : Vol.2, 291
DBTLDI : Vol.2, 293
DBTLSL : Vol.2, 288
DBTOSL : Vol.2, 271
DBTPSL : Vol.2, 254
DBTSSL : Vol.2, 274
DBTUCO : Vol.2, 284
DBTUDI : Vol.2, 286
DBTUSL : Vol.2, 281
DBVMSL : Vol.2, 277
DCGBFF : Vol.1, 337
DCGEAA : Vol.1, 148
DCGEAN : Vol.1, 153
DCGGAA : Vol.1, 273
DCGGAN : Vol.1, 278
DCGJAA : Vol.1, 299
DCGJAN : Vol.1, 303
DCGKAA : Vol.1, 305
DCGKAN : Vol.1, 309
DCGNAA : Vol.1, 155
DCGNAN : Vol.1, 158
DCGSAA : Vol.1, 280
DCGSAN : Vol.1, 284
DCGSEE : Vol.1, 292
DCGSEN : Vol.1, 297
DCGSSN : Vol.1, 290
DCGSSS : Vol.1, 286
DCSBAA : Vol.1, 222
DCSBAN : Vol.1, 225
DCSBFF : Vol.1, 233
DCSBSN : Vol.1, 231
DCSBSS : Vol.1, 227
DCSJSS : Vol.1, 260
DCSMAA : Vol.1, 171
DCSMAN : Vol.1, 174
DCSMEE : Vol.1, 182
DCSMEN : Vol.1, 186
DCSMSN : Vol.1, 180
DCSMSS : Vol.1, 176
DCSRSS : Vol.1, 254
DCSTAA : Vol.1, 237
DCSTAN : Vol.1, 240
DCSTEE : Vol.1, 248
DCSTEN : Vol.1, 252
DCSTSN : Vol.1, 246
DCSTSS : Vol.1, 242
DFASMA : Vol.6, 256
DFC1BF : Vol.3, 49
DFC1FB : Vol.3, 45
DFC2BF : Vol.3, 108
DFC2FB : Vol.3, 104

DFC3BF : Vol.3, 135
DFC3FB : Vol.3, 131
DFCMBF : Vol.3, 77
DFCMFB : Vol.3, 73
DFCN1D : Vol.3, 161
DFCN2D : Vol.3, 170
DFCN3D : Vol.3, 177
DFCR1D : Vol.3, 187
DFCR2D : Vol.3, 196
DFCR3D : Vol.3, 203
DFCRCS : Vol.6, 254
DFCRCZ : Vol.6, 252
DFCRSC : Vol.6, 250
DFCVCS : Vol.6, 246
DFCVSC : Vol.6, 243
DFDPED : Vol.6, 262
DFDPES : Vol.6, 260
DFDPET : Vol.6, 265
DFLAGE : Vol.3, 245
DFLARA : Vol.3, 240
DFPS1D : Vol.3, 213
DFPS2D : Vol.3, 221
DFPS3D : Vol.3, 228
DFR1BF : Vol.3, 67
DFR1FB : Vol.3, 63
DFR2BF : Vol.3, 126
DFR2FB : Vol.3, 122
DFR3BF : Vol.3, 155
DFR3FB : Vol.3, 150
DFRMBF : Vol.3, 98
DFRMFB : Vol.3, 93
DFWTFF : Vol.3, 272
DFWTFT : Vol.3, 274
DFWTH1 : Vol.3, 249
DFWTH2 : Vol.3, 258
DFWTHI : Vol.3, 264
DFWTHR : Vol.3, 251
DFWTHS : Vol.3, 254
DFWTHT : Vol.3, 261
DFWTMF : Vol.3, 268
DFWTMT : Vol.3, 270
DGICBP : Vol.4, 447
DGICBS : Vol.4, 467
DGICCM : Vol.4, 422
DGICCN : Vol.4, 425
DGICCO : Vol.4, 417
DGICCP : Vol.4, 408
DGICCQ : Vol.4, 410
DGICCR : Vol.4, 412
DGICCS : Vol.4, 414
DGICCT : Vol.4, 419
DGIDBY : Vol.4, 451
DGIDCY : Vol.4, 431
DGIDMC : Vol.4, 391
DGIDPC : Vol.4, 382
DGIDSC : Vol.4, 386
DGIDYB : Vol.4, 439
DGIIBZ : Vol.4, 453
DGIICZ : Vol.4, 433
DGIIMC : Vol.4, 404
DGIIPC : Vol.4, 396
DGIISC : Vol.4, 399
DGIIZB : Vol.4, 444
DGISBX : Vol.4, 449
DGISCX : Vol.4, 429
DGISI1 : Vol.4, 470
DGISI2 : Vol.4, 474
DGISI3 : Vol.4, 482
DGISMC : Vol.4, 377
DGISPC : Vol.4, 369
DGISPO : Vol.4, 455
DGISPR : Vol.4, 458
DGISS1 : Vol.4, 487
DGISS2 : Vol.4, 491
DGISS3 : Vol.4, 498
DGISSC : Vol.4, 372
DGISSO : Vol.4, 461
DGISSR : Vol.4, 464
DGISXB : Vol.4, 435
DH2INT : Vol.4, 263
DHBDFS : Vol.4, 233
DHBSFC : Vol.4, 236
DHEMNH : Vol.4, 239
DHEMNI : Vol.4, 253
DHEMNL : Vol.4, 199
DHNANL : Vol.4, 230
DHNEFL : Vol.4, 209
DHNENH : Vol.4, 246
DHNENL : Vol.4, 221
DHNFML : Vol.4, 279
DHNFNM : Vol.4, 270
DHNIFL : Vol.4, 213
DHNINH : Vol.4, 249
DHNINI : Vol.4, 259
DHNINL : Vol.4, 226
DHNOFH : Vol.4, 242
DHNOFI : Vol.4, 256
DHNOFL : Vol.4, 205
DHNPNL : Vol.4, 217
DHNRML : Vol.4, 274
DHNRNM : Vol.4, 266
DHNSNL : Vol.4, 202
DIBAID : Vol.5, 166
DIBAIX : Vol.5, 162
DIBBEI : Vol.5, 148
DIBBER : Vol.5, 146
DIBBID : Vol.5, 168
DIBBIX : Vol.5, 164

DIBIMX : Vol.5, 121
DIBINX : Vol.5, 117
DIBJMX : Vol.5, 86
DIBJNX : Vol.5, 81
DIBKEI : Vol.5, 152
DIBKER : Vol.5, 150
DIBKMX : Vol.5, 124
DIBKNX : Vol.5, 119
DIBSIN : Vol.5, 138
DIBSJN : Vol.5, 132
DIBSKN : Vol.5, 140
DIBSYN : Vol.5, 135
DIBYMX : Vol.5, 89
DIBYNX : Vol.5, 83
DIEII1 : Vol.5, 192
DIEII2 : Vol.5, 194
DIEII3 : Vol.5, 196
DIEII4 : Vol.5, 198
DIGIG1 : Vol.5, 175
DIGIG2 : Vol.5, 177
DIICOS : Vol.5, 225
DIIERF : Vol.5, 241
DIISIN : Vol.5, 223
DILEG1 : Vol.5, 245
DILEG2 : Vol.5, 248
DIMTCE : Vol.5, 265
DIMTSE : Vol.5, 268
DIOPC2 : Vol.5, 261
DIOPCH : Vol.5, 259
DIOPGL : Vol.5, 263
DIOPHE : Vol.5, 257
DIOPLA : Vol.5, 255
DIOPLE : Vol.5, 250
DIXEPS : Vol.5, 283
DIZBS0 : Vol.5, 96
DIZBS1 : Vol.5, 98
DIZBSL : Vol.5, 105
DIZBSN : Vol.5, 100
DIZBYN : Vol.5, 103
DIZGLW : Vol.5, 252
DJTECC : Vol.6, 31
DJTEEX : Vol.6, 28
DJTEGM : Vol.6, 42
DJTEGU : Vol.6, 34
DJTELG : Vol.6, 45
DJTENO : Vol.6, 24
DJTEUN : Vol.6, 19
DJTEWE : Vol.6, 38
DKFNCS : Vol.4, 67
DKHNCS : Vol.4, 73
DKINCT : Vol.4, 52
DKMNCN : Vol.4, 78
DKSNCA : Vol.4, 46
DKSNCS : Vol.4, 41
DKSSCA : Vol.4, 61
DLARHA : Vol.5, 342
DLNRDS : Vol.5, 348
DLNRIS : Vol.5, 352
DLNRSA : Vol.5, 358
DLNRSS : Vol.5, 355
DLSRDS : Vol.5, 364
DLSRIS : Vol.5, 370
DMCLAF : Vol.5, 436
DMCLCP : Vol.5, 459
DMCLMC : Vol.5, 454
DMCLMZ : Vol.5, 447
DMCLSN : Vol.5, 430
DMCLTP : Vol.5, 465
DMCQAZ : Vol.5, 481
DMCQLM : Vol.5, 476
DMCQSN : Vol.5, 471
DMCUSN : Vol.5, 427
DMSP11 : Vol.5, 500
DMSP1M : Vol.5, 493
DMSPMM : Vol.5, 497
DMSQPM : Vol.5, 487
DMUMQG : Vol.5, 418
DMUMQN : Vol.5, 414
DMUSSN : Vol.5, 422
DMUUSN : Vol.5, 411
DNCBPO : Vol.4, 345
DNDAAO : Vol.4, 319
DNDANL : Vol.4, 328
DNDAPO : Vol.4, 324
DNGAPL : Vol.4, 340
DNLNMA : Vol.6, 550
DNLNRG : Vol.6, 537
DNLNRR : Vol.6, 543
DNNLGF : Vol.6, 560
DNNLPO : Vol.6, 555
DNRAPL : Vol.4, 334
DOFNNF : Vol.4, 104
DOFNNV : Vol.4, 98
DOHNLV : Vol.4, 123
DOHNNF : Vol.4, 117
DOHNNV : Vol.4, 111
DOIEF2 : Vol.4, 134
DOIEV1 : Vol.4, 137
DOLNLV : Vol.4, 129
DOPDH2 : Vol.4, 140
DOPDH3 : Vol.4, 147
DOSNNF : Vol.4, 91
DOSNNV : Vol.4, 84
DPDAPN : Vol.4, 307
DPDOPL : Vol.4, 304
DPGOPL : Vol.4, 316
DPLOPL : Vol.4, 310
DQFODX : Vol.4, 162

DQMOGX : Vol.4, 165
DQMOHX : Vol.4, 168
DQMOJX : Vol.4, 171
DSMGON : Vol.5, 304
DSMGPA : Vol.5, 308
DSSTA1 : Vol.5, 290
DSSTA2 : Vol.5, 293
DSSTPT : Vol.5, 300
DSSTRA : Vol.5, 297
DXA005 : Vol.1, 40

GAM1HH : SMP Functions(*) , 40
GAM1HM : SMP Functions, 36
GAM1MH : SMP Functions, 32
GAM1MM : SMP Functions, 28
GAN1HH : SMP Functions, 53
GAN1HM : SMP Functions, 50
GAN1MH : SMP Functions, 47
GAN1MM : SMP Functions, 44
GBHESL : SMP Functions, 131
GBHEUD : SMP Functions, 135
GBHFSL : SMP Functions, 125
GBHFUD : SMP Functions, 129
GBHPSL : SMP Functions, 111
GBHPUD : SMP Functions, 116
GBHRSL : SMP Functions, 118
GBHRUD : SMP Functions, 123
GCGJAA : SMP Functions, 262
GCGJAN : SMP Functions, 266
GCGKAA : SMP Functions, 268
GCGKAN : SMP Functions, 273
GCGRAA : SMP Functions, 254
GCGRAN : SMP Functions, 259
GCHEAA : SMP Functions, 214
GCHEAN : SMP Functions, 218
GCHESN : SMP Functions, 225
GCHESS : SMP Functions, 220
GCHRAA : SMP Functions, 200
GCHRAN : SMP Functions, 204
GCHRSN : SMP Functions, 211
GCHRSS : SMP Functions, 206
GFC2BF : SMP Functions, 324
GFC2FB : SMP Functions, 321
GFC3BF : SMP Functions, 351
GFC3FB : SMP Functions, 347
GFCMBF : SMP Functions, 295
GFCMFB : SMP Functions, 291

HAM1HH : SMP Functions, 40
HAM1HM : SMP Functions, 36

(*) DMP Functions: Distributed Memory Parallel Functions
(*) SMP Functions: Shared Memory Parallel Functions

HAM1MH : SMP Functions, 32
HAM1MM : SMP Functions, 28
HAN1HH : SMP Functions, 53
HAN1HM : SMP Functions, 50
HAN1MH : SMP Functions, 47
HAN1MM : SMP Functions, 44
HBGMLC : SMP Functions, 87
HBGMLU : SMP Functions, 85
HBGMSL : SMP Functions, 81
HBGMSM : SMP Functions, 76
HBGNLC : SMP Functions, 97
HBGNLU : SMP Functions, 95
HBGNSL : SMP Functions, 92
HBGNSM : SMP Functions, 89
HBHESL : SMP Functions, 131
HBHEUD : SMP Functions, 135
HBHFSL : SMP Functions, 125
HBHFUD : SMP Functions, 129
HBHPSL : SMP Functions, 111
HBHPUD : SMP Functions, 116
HBHRSL : SMP Functions, 118
HBHRUD : SMP Functions, 123
HCGJAA : SMP Functions, 262
HCGJAN : SMP Functions, 266
HCGKAA : SMP Functions, 268
HCGKAN : SMP Functions, 273
HCGRAA : SMP Functions, 254
HCGRAN : SMP Functions, 259
HCHEAA : SMP Functions, 214
HCHEAN : SMP Functions, 218
HCHESN : SMP Functions, 225
HCHESS : SMP Functions, 220
HCHRAA : SMP Functions, 200
HCHRAN : SMP Functions, 204
HCHRSN : SMP Functions, 211
HCHRSS : SMP Functions, 206
HFC2BF : SMP Functions, 324
HFC2FB : SMP Functions, 321
HFC3BF : SMP Functions, 351
HFC3FB : SMP Functions, 347
HFCMBF : SMP Functions, 295
HFCMFB : SMP Functions, 291

IIIERF : Vol.5, 243

JIIERF : Vol.5, 243

PAM1MM : SMP Functions, 16
PAM1MT : SMP Functions, 19
PAM1MU : SMP Functions, 13
PAM1TM : SMP Functions, 22
PAM1TT : SMP Functions, 25
PBSNSL : SMP Functions, 105
PBSNUD : SMP Functions, 109

PBSPSL : SMP Functions, 99
PBSPUD : SMP Functions, 103
PCGJAA : SMP Functions, 242
PCGJAN : SMP Functions, 246
PCGKAA : SMP Functions, 248
PCGKAN : SMP Functions, 252
PCGSAA : SMP Functions, 227
PCGSAN : SMP Functions, 232
PCGSSN : SMP Functions, 239
PCGSSS : SMP Functions, 234
PCSMAA : SMP Functions, 189
PCSMAN : SMP Functions, 192
PCSMSN : SMP Functions, 198
PCSMSS : SMP Functions, 194
PFC2BF : SMP Functions, 316
PFC2FB : SMP Functions, 312
PFC3BF : SMP Functions, 342
PFC3FB : SMP Functions, 338
PFCMBF : SMP Functions, 285
PFCMFB : SMP Functions, 281
PFCN2D : SMP Functions, 366
PFCN3D : SMP Functions, 373
PFCR2D : SMP Functions, 382
PFCR3D : SMP Functions, 389
PFPS2D : SMP Functions, 399
PFPS3D : SMP Functions, 406
PFR2BF : SMP Functions, 333
PFR2FB : SMP Functions, 329
PFR3BF : SMP Functions, 360
PFR3FB : SMP Functions, 356
PFRMBF : SMP Functions, 305
PFRMFB : SMP Functions, 301
PSSTA1 : SMP Functions, 422
PSSTA2 : SMP Functions, 425
PXE010 : SMP Functions, 148
PXE020 : SMP Functions, 156
PXE030 : SMP Functions, 164
PXE040 : SMP Functions, 172

QAM1MM : SMP Functions, 16
QAM1MT : SMP Functions, 19
QAM1MU : SMP Functions, 13
QAM1TM : SMP Functions, 22
QAM1TT : SMP Functions, 25
QBGMLC : SMP Functions, 74
QBGMLU : SMP Functions, 72
QBGMSL : SMP Functions, 68
QBGMSM : SMP Functions, 64
QBSNSL : SMP Functions, 105
QBSNUD : SMP Functions, 109
QBSPSL : SMP Functions, 99
QBSPUD : SMP Functions, 103
QCGJAA : SMP Functions, 242
QCGJAN : SMP Functions, 246
QCGKAA : SMP Functions, 248
QCGKAN : SMP Functions, 252
QCGSAA : SMP Functions, 227
QCGSAN : SMP Functions, 232
QCGSSN : SMP Functions, 239
QCGSSS : SMP Functions, 234
QCSMAA : SMP Functions, 189
QCSMAN : SMP Functions, 192
QCSMSN : SMP Functions, 198
QCSMSS : SMP Functions, 194
QFC2BF : SMP Functions, 316
QFC2FB : SMP Functions, 312
QFC3BF : SMP Functions, 342
QFC3FB : SMP Functions, 338
QFCMBF : SMP Functions, 285
QFCMFB : SMP Functions, 281
QFCN2D : SMP Functions, 366
QFCN3D : SMP Functions, 373
QFCR2D : SMP Functions, 382
QFCR3D : SMP Functions, 389
QFPS2D : SMP Functions, 399
QFPS3D : SMP Functions, 406
QFR2BF : SMP Functions, 333
QFR2FB : SMP Functions, 329
QFR3BF : SMP Functions, 360
QFR3FB : SMP Functions, 356
QFRMBF : SMP Functions, 305
QFRMFB : SMP Functions, 301
QSSTA1 : SMP Functions, 422
QSSTA2 : SMP Functions, 425
QXE010 : SMP Functions, 148
QXE020 : SMP Functions, 156
QXE030 : SMP Functions, 164
QXE040 : SMP Functions, 172

R1CDBN : Vol.6, 72
R1CDBT : Vol.6, 114
R1CDCC : Vol.6, 147
R1CDCH : Vol.6, 75
R1CDEX : Vol.6, 132
R1CDFB : Vol.6, 100
R1CDGM : Vol.6, 107
R1CDGU : Vol.6, 135
R1CDIB : Vol.6, 117
R1CDIC : Vol.6, 78
R1CDIF : Vol.6, 104
R1CDIG : Vol.6, 111
R1CDIN : Vol.6, 69
R1CDIS : Vol.6, 97
R1CDIT : Vol.6, 91
R1CDIX : Vol.6, 85
R1CDLD : Vol.6, 138
R1CDLG : Vol.6, 144
R1CDLN : Vol.6, 141

R1CDNC : Vol.6, 81
R1CDNO : Vol.6, 66
R1CDNT : Vol.6, 94
R1CDPA : Vol.6, 126
R1CDTB : Vol.6, 88
R1CDTR : Vol.6, 123
R1CDUF : Vol.6, 120
R1CDWE : Vol.6, 129
R1DDBP : Vol.6, 150
R1DDGO : Vol.6, 154
R1DDHG : Vol.6, 159
R1DDHN : Vol.6, 162
R1DDPO : Vol.6, 157
R2BA1T : Vol.6, 173
R2BA2S : Vol.6, 179
R2BAGM : Vol.6, 191
R2BAHM : Vol.6, 199
R2BAMO : Vol.6, 195
R2BAMS : Vol.6, 186
R2BASM : Vol.6, 203
R2CCMA : Vol.6, 225
R2CCMT : Vol.6, 220
R2CCPR : Vol.6, 231
R2VCGR : Vol.6, 212
R2VCMT : Vol.6, 207
R3IECD : Vol.6, 307
R3IEME : Vol.6, 293
R3IERA : Vol.6, 290
R3IESR : Vol.6, 311
R3IESU : Vol.6, 297
R3IETC : Vol.6, 304
R3IEVA : Vol.6, 301
R3TSCD : Vol.6, 347
R3TSME : Vol.6, 326
R3TSRA : Vol.6, 317
R3TSRD : Vol.6, 321
R3TSSR : Vol.6, 350
R3TSSU : Vol.6, 331
R3TSTC : Vol.6, 342
R3TSVA : Vol.6, 338
R41WR1 : Vol.6, 363
R42WR1 : Vol.6, 383
R42WRM : Vol.6, 375
R42WRN : Vol.6, 369
R4BI01 : Vol.6, 438
R4GL01 : Vol.6, 434
R4MU01 : Vol.6, 415
R4MWRF : Vol.6, 391
R4MWRM : Vol.6, 402
R4RB01 : Vol.6, 430
R5CHEF : Vol.6, 447
R5CHMD : Vol.6, 456
R5CHMN : Vol.6, 453
R5CHTT : Vol.6, 450
R5TEMH : Vol.6, 466
R5TESG : Vol.6, 459
R5TESP : Vol.6, 470
R5TEWL : Vol.6, 462
R6CLAN : Vol.6, 518
R6CLDA : Vol.6, 523
R6CLDS : Vol.6, 513
R6CPCC : Vol.6, 482
R6CPSC : Vol.6, 484
R6CVAN : Vol.6, 495
R6CVSC : Vol.6, 498
R6DAFN : Vol.6, 503
R6DASC : Vol.6, 507
R6FALD : Vol.6, 489
R6FAVR : Vol.6, 491
RABMCS : Vol.1, 12
RABMEL : Vol.1, 15
RAM1AD : Vol.1, 47
RAM1MM : Vol.1, 64
RAM1MS : Vol.1, 56
RAM1MT : Vol.1, 67
RAM1MU : Vol.1, 53
RAM1SB : Vol.1, 50
RAM1TM : Vol.1, 70
RAM1TP : Vol.1, 109
RAM1TT : Vol.1, 73
RAM1VM : Vol.1, 100
RAM3TP : Vol.1, 111
RAM3VM : Vol.1, 103
RAM4VM : Vol.1, 106
RAMT1M : Vol.1, 59
RAMVJ1 : Vol.1, 114
RAMVJ3 : Vol.1, 118
RAMVJ4 : Vol.1, 122
RARGJM : Vol.1, 26
RARSJD : Vol.1, 21
RASBCS : Vol.1, 17
RASBEL : Vol.1, 19
RATM1M : Vol.1, 61
RBBDDI : Vol.2, 231
RBBDLC : Vol.2, 227
RBBDLS : Vol.2, 229
RBBDLU : Vol.2, 225
RBBDLX : Vol.2, 233
RBBDSL : Vol.2, 220
RBBPDI : Vol.2, 247
RBBPLS : Vol.2, 245
RBBPLX : Vol.2, 249
RBBPSL : Vol.2, 237
RBBPUC : Vol.2, 243
RBBPUU : Vol.2, 241
RBGMDI : Vol.2, 48
RBGMLC : Vol.2, 41
RBGMLS : Vol.2, 43

RBGMLU : Vol.2, 39
RBGMLX : Vol.2, 50
RBGMMS : Vol.2, 45
RBGMSL : Vol.2, 35
RBGMSM : Vol.2, 31
RBPDDI : Vol.2, 106
RBPDLS : Vol.2, 104
RBPDLX : Vol.2, 108
RBPDSL : Vol.2, 96
RBPDUC : Vol.2, 102
RBPDUU : Vol.2, 100
RBSMDI : Vol.2, 140
RBSMLS : Vol.2, 135
RBSMLX : Vol.2, 142
RBSMMS : Vol.2, 137
RBSMSL : Vol.2, 127
RBSMUC : Vol.2, 133
RBSMUD : Vol.2, 131
RBSNLS : Vol.2, 150
RBSNSL : Vol.2, 144
RBSNUD : Vol.2, 148
RBSPDI : Vol.2, 123
RBSPLS : Vol.2, 118
RBSPLX : Vol.2, 125
RBSPMS : Vol.2, 120
RBSPSL : Vol.2, 110
RBSPUC : Vol.2, 116
RBSPUD : Vol.2, 114
RBTDSL : Vol.2, 251
RBTLCO : Vol.2, 291
RBTLDI : Vol.2, 293
RBTLSL : Vol.2, 288
RBTOSL : Vol.2, 271
RBTPSL : Vol.2, 254
RBTSSL : Vol.2, 274
RBTUCO : Vol.2, 284
RBTUDI : Vol.2, 286
RBTUSL : Vol.2, 281
RBVMSL : Vol.2, 277
RCGBFF : Vol.1, 337
RCGEAA : Vol.1, 148
RCGEAN : Vol.1, 153
RCGGAA : Vol.1, 273
RCGGAN : Vol.1, 278
RCGJAA : Vol.1, 299
RCGJAN : Vol.1, 303
RCGKAA : Vol.1, 305
RCGKAN : Vol.1, 309
RCGNAA : Vol.1, 155
RCGNAN : Vol.1, 158
RCGSAA : Vol.1, 280
RCGSAN : Vol.1, 284
RCGSEE : Vol.1, 292
RCGSEN : Vol.1, 297
RCGSSN : Vol.1, 290
RCGSSS : Vol.1, 286
RCSBAA : Vol.1, 222
RCSBAN : Vol.1, 225
RCSBFF : Vol.1, 233
RCSBSN : Vol.1, 231
RCSBSS : Vol.1, 227
RCSJSS : Vol.1, 260
RCSMAA : Vol.1, 171
RCSMAN : Vol.1, 174
RCSMEE : Vol.1, 182
RCSMEN : Vol.1, 186
RCSMSN : Vol.1, 180
RCSMSS : Vol.1, 176
RCSRSS : Vol.1, 254
RCSTAA : Vol.1, 237
RCSTAN : Vol.1, 240
RCSTEE : Vol.1, 248
RCSTEN : Vol.1, 252
RCSTSN : Vol.1, 246
RCSTSS : Vol.1, 242
RFASMA : Vol.6, 256
RFC1BF : Vol.3, 49
RFC1FB : Vol.3, 45
RFC2BF : Vol.3, 108
RFC2FB : Vol.3, 104
RFC3BF : Vol.3, 135
RFC3FB : Vol.3, 131
RFCMBF : Vol.3, 77
RFCMFB : Vol.3, 73
RFCN1D : Vol.3, 161
RFCN2D : Vol.3, 170
RFCN3D : Vol.3, 177
RFCR1D : Vol.3, 187
RFCR2D : Vol.3, 196
RFCR3D : Vol.3, 203
RFCRCS : Vol.6, 254
RFCRCZ : Vol.6, 252
RFCRSC : Vol.6, 250
RFCVCS : Vol.6, 246
RFCVSC : Vol.6, 243
RFDPED : Vol.6, 262
RFDPES : Vol.6, 260
RFDPET : Vol.6, 265
RFLAGE : Vol.3, 245
RFLARA : Vol.3, 240
RFPS1D : Vol.3, 213
RFPS2D : Vol.3, 221
RFPS3D : Vol.3, 228
RFR1BF : Vol.3, 67
RFR1FB : Vol.3, 63
RFR2BF : Vol.3, 126
RFR2FB : Vol.3, 122
RFR3BF : Vol.3, 155

RFR3FB : Vol.3, 150
RFRMBF : Vol.3, 98
RFRMFB : Vol.3, 93
RFWTFF : Vol.3, 272
RFWTFT : Vol.3, 274
RFWTH1 : Vol.3, 249
RFWTH2 : Vol.3, 258
RFWTHI : Vol.3, 264
RFWTHR : Vol.3, 251
RFWTHS : Vol.3, 254
RFWTHT : Vol.3, 261
RFWTMF : Vol.3, 268
RFWTMT : Vol.3, 270
RGICBP : Vol.4, 447
RGICBS : Vol.4, 467
RGICCM : Vol.4, 422
RGICCN : Vol.4, 425
RGICCO : Vol.4, 417
RGICCP : Vol.4, 408
RGICCQ : Vol.4, 410
RGICCR : Vol.4, 412
RGICCS : Vol.4, 414
RGICCT : Vol.4, 419
RGIDBY : Vol.4, 451
RGIDCY : Vol.4, 431
RGIDMC : Vol.4, 391
RGIDPC : Vol.4, 382
RGIDSC : Vol.4, 386
RGIDYB : Vol.4, 439
RGIIBZ : Vol.4, 453
RGIICZ : Vol.4, 433
RGIIMC : Vol.4, 404
RGIIPC : Vol.4, 396
RGIISC : Vol.4, 399
RGIIZB : Vol.4, 444
RGISBX : Vol.4, 449
RGISCX : Vol.4, 429
RGISI1 : Vol.4, 470
RGISI2 : Vol.4, 474
RGISI3 : Vol.4, 482
RGISMC : Vol.4, 377
RGISPC : Vol.4, 369
RGISPO : Vol.4, 455
RGISPR : Vol.4, 458
RGISS1 : Vol.4, 487
RGISS2 : Vol.4, 491
RGISS3 : Vol.4, 498
RGISSC : Vol.4, 372
RGISSO : Vol.4, 461
RGISSR : Vol.4, 464
RGISXB : Vol.4, 435
RH2INT : Vol.4, 263
RHBDFS : Vol.4, 233
RHBSFC : Vol.4, 236
RHEMNH : Vol.4, 239
RHEMNI : Vol.4, 253
RHEMNL : Vol.4, 199
RHNANL : Vol.4, 230
RHNEFL : Vol.4, 209
RHNENH : Vol.4, 246
RHNENL : Vol.4, 221
RHNFML : Vol.4, 279
RHNFNM : Vol.4, 270
RHNIFL : Vol.4, 213
RHNINH : Vol.4, 249
RHNINI : Vol.4, 259
RHNINL : Vol.4, 226
RHNOFH : Vol.4, 242
RHNOFI : Vol.4, 256
RHNOFL : Vol.4, 205
RHNPNL : Vol.4, 217
RHNRML : Vol.4, 274
RHNRNM : Vol.4, 266
RHNSNL : Vol.4, 202
RIBAID : Vol.5, 166
RIBAIX : Vol.5, 162
RIBBEI : Vol.5, 148
RIBBER : Vol.5, 146
RIBBID : Vol.5, 168
RIBBIX : Vol.5, 164
RIBIMX : Vol.5, 121
RIBINX : Vol.5, 117
RIBJMX : Vol.5, 86
RIBJNX : Vol.5, 81
RIBKEI : Vol.5, 152
RIBKER : Vol.5, 150
RIBKMX : Vol.5, 124
RIBKNX : Vol.5, 119
RIBSIN : Vol.5, 138
RIBSJN : Vol.5, 132
RIBSKN : Vol.5, 140
RIBSYN : Vol.5, 135
RIBYMX : Vol.5, 89
RIBYNX : Vol.5, 83
RIEII1 : Vol.5, 192
RIEII2 : Vol.5, 194
RIEII3 : Vol.5, 196
RIEII4 : Vol.5, 198
RIGIG1 : Vol.5, 175
RIGIG2 : Vol.5, 177
RIICOS : Vol.5, 225
RIIERF : Vol.5, 241
RIISIN : Vol.5, 223
RILEG1 : Vol.5, 245
RILEG2 : Vol.5, 248
RIMTCE : Vol.5, 265
RIMTSE : Vol.5, 268
RIOPC2 : Vol.5, 261

RIOPCH : Vol.5, 259
RIOPGL : Vol.5, 263
RIOPHE : Vol.5, 257
RIOPLA : Vol.5, 255
RIOPLE : Vol.5, 250
RIXEPS : Vol.5, 283
RIZBS0 : Vol.5, 96
RIZBS1 : Vol.5, 98
RIZBSL : Vol.5, 105
RIZBSN : Vol.5, 100
RIZBYN : Vol.5, 103
RIZGLW : Vol.5, 252
RJTEBI : Vol.6, 49
RJTECC : Vol.6, 31
RJTEEX : Vol.6, 28
RJTEGM : Vol.6, 42
RJTEGU : Vol.6, 34
RJTELG : Vol.6, 45
RJTENG : Vol.6, 52
RJTENO : Vol.6, 24
RJTEPO : Vol.6, 55
RJTEUN : Vol.6, 19
RJTEWE : Vol.6, 38
RKFNCS : Vol.4, 67
RKHNCS : Vol.4, 73
RKINCT : Vol.4, 52
RKMNCN : Vol.4, 78
RKSNCA : Vol.4, 46
RKSNCS : Vol.4, 41
RKSSCA : Vol.4, 61
RLARHA : Vol.5, 342
RLNRDS : Vol.5, 348
RLNRIS : Vol.5, 352
RLNRSA : Vol.5, 358
RLNRSS : Vol.5, 355
RLSRDS : Vol.5, 364
RLSRIS : Vol.5, 370
RMCLAF : Vol.5, 436
RMCLCP : Vol.5, 459
RMCLMC : Vol.5, 454
RMCLMZ : Vol.5, 447
RMCLSN : Vol.5, 430
RMCLTP : Vol.5, 465
RMCQAZ : Vol.5, 481
RMCQLM : Vol.5, 476
RMCQSN : Vol.5, 471
RMCUSN : Vol.5, 427
RMSP11 : Vol.5, 500
RMSP1M : Vol.5, 493
RMSPMM : Vol.5, 497
RMSQPM : Vol.5, 487
RMUMQG : Vol.5, 418
RMUMQN : Vol.5, 414
RMUSSN : Vol.5, 422
RMUUSN : Vol.5, 411
RNCBPO : Vol.4, 345
RNDAAO : Vol.4, 319
RNDANL : Vol.4, 328
RNDAPO : Vol.4, 324
RNGAPL : Vol.4, 340
RNLNMA : Vol.6, 550
RNLNRG : Vol.6, 537
RNLNRR : Vol.6, 543
RNNLGF : Vol.6, 560
RNRAPL : Vol.4, 334
ROFNNF : Vol.4, 104
ROFNNV : Vol.4, 98
ROHNLV : Vol.4, 123
ROHNNF : Vol.4, 117
ROHNNV : Vol.4, 111
ROIEF2 : Vol.4, 134
ROIEV1 : Vol.4, 137
ROLNLV : Vol.4, 129
ROPDH2 : Vol.4, 140
ROPDH3 : Vol.4, 147
ROSNNF : Vol.4, 91
ROSNNV : Vol.4, 84
RPDAPN : Vol.4, 307
RPDOPL : Vol.4, 304
RPGOPL : Vol.4, 316
RPLOPL : Vol.4, 310
RQFODX : Vol.4, 162
RQMOGX : Vol.4, 165
RQMOHX : Vol.4, 168
RQMOJX : Vol.4, 171
RSMGON : Vol.5, 304
RSMGPA : Vol.5, 308
RSSTA1 : Vol.5, 290
RSSTA2 : Vol.5, 293
RSSTPT : Vol.5, 300
RSSTRA : Vol.5, 297
RXA005 : Vol.1, 40

VIBH0X : Vol.5, 154
VIBH1X : Vol.5, 156
VIBHY0 : Vol.5, 158
VIBHY1 : Vol.5, 160
VIBI0X : Vol.5, 107
VIBI1X : Vol.5, 112
VIBJ0X : Vol.5, 71
VIBJ1X : Vol.5, 76
VIBK0X : Vol.5, 109
VIBK1X : Vol.5, 114
VIBY0X : Vol.5, 73
VIBY1X : Vol.5, 78
VIDBEY : Vol.5, 273
VIECI1 : Vol.5, 188
VIECI2 : Vol.5, 190

VIEJAC : Vol.5, 200
VIEJEP : Vol.5, 211
VIEJTE : Vol.5, 213
VIEJZT : Vol.5, 209
VIENMQ : Vol.5, 203
VIEPAI : Vol.5, 215
VIERFC : Vol.5, 239
VIERRF : Vol.5, 237
VIETHE : Vol.5, 206
VIGAMX : Vol.5, 170
VIGBET : Vol.5, 185
VIGDIG : Vol.5, 183
VIGLGX : Vol.5, 173
VIICNC : Vol.5, 235
VIICND : Vol.5, 233
VIIDAW : Vol.5, 231
VIIEXP : Vol.5, 218
VIIFCO : Vol.5, 229
VIIFSI : Vol.5, 227
VIILOG : Vol.5, 221
VINPLG : Vol.5, 275
VIXSLA : Vol.5, 278
VIXSPS : Vol.5, 271
VIXZTA : Vol.5, 280

WBTCLS : Vol.2, 267
WBTCSL : Vol.2, 263
WBTDLS : Vol.2, 260
WBTDSL : Vol.2, 257
WIBH0X : Vol.5, 154
WIBH1X : Vol.5, 156
WIBHY0 : Vol.5, 158
WIBHY1 : Vol.5, 160
WIBI0X : Vol.5, 107
WIBI1X : Vol.5, 112
WIBJ0X : Vol.5, 71
WIBJ1X : Vol.5, 76
WIBK0X : Vol.5, 109
WIBK1X : Vol.5, 114
WIBY0X : Vol.5, 73
WIBY1X : Vol.5, 78
WIDBEY : Vol.5, 273
WIECI1 : Vol.5, 188
WIECI2 : Vol.5, 190
WIEJAC : Vol.5, 200
WIEJEP : Vol.5, 211
WIEJTE : Vol.5, 213
WIEJZT : Vol.5, 209
WIENMQ : Vol.5, 203
WIEPAI : Vol.5, 215
WIERFC : Vol.5, 239
WIERRF : Vol.5, 237
WIETHE : Vol.5, 206
WIGAMX : Vol.5, 170
WIGBET : Vol.5, 185
WIGDIG : Vol.5, 183
WIGLGX : Vol.5, 173
WIICNC : Vol.5, 235
WIICND : Vol.5, 233
WIIDAW : Vol.5, 231
WIIEXP : Vol.5, 218
WIIFCO : Vol.5, 229
WIIFSI : Vol.5, 227
WIILOG : Vol.5, 221
WINPLG : Vol.5, 275
WIXSLA : Vol.5, 278
WIXSPS : Vol.5, 271
WIXZTA : Vol.5, 280

ZAM1HH : Vol.1, 85
ZAM1HM : Vol.1, 82
ZAM1MH : Vol.1, 79
ZAM1MM : Vol.1, 76
ZAN1HH : Vol.1, 97
ZAN1HM : Vol.1, 94
ZAN1MH : Vol.1, 91
ZAN1MM : Vol.1, 88
ZANVJ1 : Vol.1, 126
ZARGJM : Vol.1, 37
ZARSJD : Vol.1, 32
ZBGMDI : Vol.2, 72
ZBGMLC : Vol.2, 64
ZBGMLS : Vol.2, 66
ZBGMLU : Vol.2, 62
ZBGMLX : Vol.2, 74
ZBGMMS : Vol.2, 68
ZBGMSL : Vol.2, 58
ZBGMSM : Vol.2, 54
ZBGNDI : Vol.2, 92
ZBGNLC : Vol.2, 84
ZBGNLS : Vol.2, 86
ZBGNLU : Vol.2, 82
ZBGNLX : Vol.2, 94
ZBGNMS : Vol.2, 88
ZBGNSL : Vol.2, 79
ZBGNSM : Vol.2, 76
ZBHEDI : Vol.2, 216
ZBHELS : Vol.2, 211
ZBHELX : Vol.2, 218
ZBHEMS : Vol.2, 213
ZBHESL : Vol.2, 203
ZBHEUC : Vol.2, 209
ZBHEUD : Vol.2, 207
ZBHFDI : Vol.2, 199
ZBHFLS : Vol.2, 194
ZBHFLX : Vol.2, 201
ZBHFMS : Vol.2, 196
ZBHFSL : Vol.2, 186

ZBHFUC : Vol.2, 192
ZBHFUD : Vol.2, 190
ZBHPDI : Vol.2, 165
ZBHPLS : Vol.2, 160
ZBHPLX : Vol.2, 167
ZBHPMS : Vol.2, 162
ZBHPSL : Vol.2, 152
ZBHPUC : Vol.2, 158
ZBHPUD : Vol.2, 156
ZBHRDI : Vol.2, 182
ZBHRLS : Vol.2, 177
ZBHRLX : Vol.2, 184
ZBHRMS : Vol.2, 179
ZBHRSL : Vol.2, 169
ZBHRUC : Vol.2, 175
ZBHRUD : Vol.2, 173
ZCGEAA : Vol.1, 160
ZCGEAN : Vol.1, 164
ZCGHAA : Vol.1, 318
ZCGHAN : Vol.1, 323
ZCGJAA : Vol.1, 325
ZCGJAN : Vol.1, 329
ZCGKAA : Vol.1, 331
ZCGKAN : Vol.1, 335
ZCGNAA : Vol.1, 166
ZCGNAN : Vol.1, 169
ZCGRAA : Vol.1, 311
ZCGRAN : Vol.1, 316
ZCHEAA : Vol.1, 205
ZCHEAN : Vol.1, 208
ZCHEEE : Vol.1, 216
ZCHEEN : Vol.1, 220
ZCHESN : Vol.1, 214
ZCHESS : Vol.1, 210
ZCHJSS : Vol.1, 267
ZCHRAA : Vol.1, 188
ZCHRAN : Vol.1, 191
ZCHREE : Vol.1, 199
ZCHREN : Vol.1, 203
ZCHRSN : Vol.1, 197
ZCHRSS : Vol.1, 193
ZFC1BF : Vol.3, 58
ZFC1FB : Vol.3, 54
ZFC2BF : Vol.3, 117
ZFC2FB : Vol.3, 113
ZFC3BF : Vol.3, 145
ZFC3FB : Vol.3, 141
ZFCMBF : Vol.3, 87
ZFCMFB : Vol.3, 83
ZIBH1N : Vol.5, 142
ZIBH2N : Vol.5, 144
ZIBINZ : Vol.5, 127
ZIBJNZ : Vol.5, 92
ZIBKNZ : Vol.5, 129
ZIBYNZ : Vol.5, 94
ZIGAMZ : Vol.5, 179
ZIGLGZ : Vol.5, 181
ZLACHA : Vol.5, 345
ZLNCIS : Vol.5, 361