Honeywell Bull

BIOMEDICAL (BMD) STATISTICAL PROGRAMS REFERENCE MANUAL

SERIES 600/6000

MANAGEMENT SCIENCE

APPLICATIONS

Ref.: 19.60.802 A, Rev. 0

Honeywell Bull

BIOMEDICAL (BMD) STATISTICAL PROGRAMS REFERENCE MANUAL

SERIES 600/6000

MANAGEMENT SCIENCE

SUBJECT:

Summarized Descriptions of Biomedical (BMD) Statistical Programs and Implementation Procedures for the Series 6000 Systems.

SPECIAL INSTRUCTIONS:

This edition supersedes CPB-1183A, dated October 1969, and Addendum CPB-1183A-1, dated March 31, 1971. The new order number is consistent with the overall Honeywell publications numbering system. Change bars in the margins indicate new and changed information; asterisks denote deletions.

The Series 600/6000 Biomedical (BMD) Statistical Programs are cataloged under Honeywell Program Library Number CD600D3.007.

DATE:

June 1973

ORDER NUMBER:

BP82, Rev. 0 (Supersedes CPB-1183A)

19 60 802.A, Rev.0

PREFACE

This manual describes a series of computer programs for performing statistical analyses and solving a variety of data processing problems in management science. Although these programs were developed for the biomedical (BMD) field, they have a broad range of applications in extensive analyses of large amounts of recorded data. These descriptions are addressed to users of BMD programs and summarize the functions of each program, the type of output obtained, and limitations in the parameters and elements allowed in the input data.

BMD programs fall into two general categories: description and tabulation and statistical analyses. Section I includes the programs in the description and tabulation category that perform the data screening to make the problem data more acceptable to the other BMD programs. Sections II through VII summarize the programs in the other category, that of performing statistical analyses of several types, including multivariate, regression, time series, variance, and contingency table statistical analyses. Section VIII contains a series of later-released programs to supplement those in the statistical analysis classes covered in Sections II through VII.

These BMD programs are an adaptation of the original biomedical programs prepared by the staff of the Health Sciences Computing Facility, School of Medicine, University of California at Los Angeles (UCLA). Complete detailed descriptions of these BMD programs are published in a reference manual by UCLA. The Series 6000 BMD user should have the following manual available for reference and as a guide to preparing input data: BMD, Biomedical Computer Programs.

Permission to use portions of the original material was obtained from Dr. W. J. Dixon, Editor of the manual, available from the University of California Press, Berkeley, CA. 94720.

Certain Series 6000 system software functions are used during copying and executing of the BMD programs, as noted in the text and described in the following reference manuals:

- Series 600/6000 Macro Assembler Program, Order Number BN86
- Series 600/6000 General Comprehensive Operating Supervisor, Order Number BR43
- Series 600/6000 FORTRAN, Order Number BJ67
- Series 600/6000 System Startup and Operation, Order Number DA06
- Series 600/6000 Service Routines, Order Number DA97
- Series 600/6000 Control Cards Reference Manual, Order Number BS19

The Series 6000 BMD programs are a coded system characterized by a highly structured format for ease of implementation plus considerable input/output flexibility to accommodate varying requirements. These programs are supported by comprehensive documentation and training; periodic program maintenance is furnished for the current version of the programs, provided they are not modified by the user.

File No.: 16M3, 17M3

CONTENTS

CONTENTS (cont)

TABLES

SECTION I

INTRODUCTION TO BMD PROGRAMS

SOLVING RESEARCH PROBLEMS

Many problems in research require extensive analyses of large amounts of recorded data. Thus, the data handling processes during research should be made both automatic and quick to achieve solutions in minimum time. The electronic computer provides the researcher with the appropriate tools to accomplish the required types of research analyses.

This requirement has led to the development of a number of generalized computer programs to perform the data handling and statistical analyses. These programs were originally written to solve problems using biomedical type data. They have thus become known as BMD programs, because their source is the School of Medicine, University of California at Los Angeles (UCLA).

These BMD programs are sufficiently general to be applied to a broad range of management science applications, such as the following types of statistical analyses: multivariate, regression, time series, variance, and contingency table analysis. Another category of BMD programs edits and screens the basic input data to be more acceptable to the other categories of programs; they also can provide special tabulations, counts, etc.

The BMD programs are extensively documented in a reference manual published by UCLA, as listed in the Preface. The BMD user should have this manual available for a more complete understanding of the programs' functions and also for preparation of the input data cards.

The solution of many problems may require that several BMD programs be executed in succession. For example, an analysis may begin with a data screening program (from the description and tabulation category) to make the statistical problem data more acceptable to the later programs. The solution then proceeds to various types of analyses, each based on the findings of the preceding analysis.

ADAPTATION TO SERIES 6000

The primary aim of adapting these BMD programs to run on Series 6000 information systems is to retain both the input and output in formats virtually identical to those of the original UCLA BMD programs. The original programs are written in the FORTRAN II language, except for certain subroutines that are coded in the FORTRAN Assembly Program (FAP) language. The adaptation to the Series 6000 systems consists of three major processes:

1. Translating the portions written in FORTRAN II into the Series 6000 FORTRAN language.

2. Rewriting the routines coded in the FAP language into the Series 6000 macro assembly language.

3. Reprogramming to eliminate certain features of the original programs that depend on nonstandard system features unique to the UCLA BMD systems.

BMD PROGRAM USAGE

The preparation of the basic research data for computer analysis differs greatly from problem to problem. Generally, the data involved covers either many variables for each case or many observations on fewer variables.

The usual computer system requires that input data be reduced to a set of symbols that can be read and interpreted by the system's input devices. The input data for the Series 6000 BMD programs is essentially the same as that described for the UCLA BMD programs. The output data may differ from that illustrated in the UCLA BMD programs reference manual because the Series 6000 systems software system output (SYSOUT) routines allow the output to be printed online. The UCLA programs usually write the output data onto a magnetic tape file, which is later listed on peripheral printing equipment.

Data Cards

The input cards for BMD programs to be processed on a Series 6000 system are punched in the exact formats described in the UCLA BMD reference manual listed in the Preface. Most BMD programs require that transgeneration codes be specified by the user when preparing input data for a problem. A list of these codes appears in Section III-B of the Introduction in the UCLA BMD reference manual.

System Cards

Two sections of the Introduction in the UCLA BMD reference manual do not apply to the programs adapted for processing on the Series 6000 systems. They are Section IV, "Preparation of System Control Cards", and Section V, "Program Operation". The information on control cards and setup of BMD programs for processing on Series 6000 systems appears later in this section.

PROGRAM CATEGORIES

Nearly 50 original BMD programs have been adapted to Series 6000 systems. They are classed in seven categories; the programs are summarized in this section. Each program is described in detail in the UCLA BMD reference manual.

Class D - Description and Tabulation

The class D BMD programs provide editing and screening of a problem's basic information to make the statistical data acceptable to the more elaborate analyses in the other classes of BMD programs, such as those in classes M, R, and V. The class D programs allow the user to examine the original data after it is punched into cards and detect problems such as coding or keypunch errors, gross input observation errors, and cases that are inappropriate for analysis.

Class D programs can also provide a broad variety of "byproduct" type information along with the main analytical solution. This includes cross tabulations, frequency counts, correlation coefficients, standard deviations, means and extremes for each variable, specified subsets of the data, and graphical output of the data. The programs in Class D have the following numbers and names:

BMD01D	Simple Data Description
BMD02D	Correlation with Transgeneration
BMD03D	Correlation with Item Deletion
BMD04D	Alphanumeric Frequency Count
BMD05D	General Plot including Histogram
BMD06D	Description of Strata
BMD07D	Description of Strata with Histograms
BMD08D	Cross-Tabulation with Variable Stacking
BMD09D	Cross-Tabulation, Incomplete Data
BMD10D	Data Patterns for Dichotomies
BMD11D	Data Patterns for Polychotomies

Class M - Multivariate Analysis

Class M programs provide various analyses on multivariate observations, that is, several variables observed for each case. The analyses use the information on the interrelationships among these variables in the form of correlation coefficients of each variable with every other variable. The programs base their analyses on data which are not categorized by case. They provide analyses of the effectiveness of the discrimination on the input observations. Programs in this class are:

BMD01M	Principal Component Analysis
BMD02M	Regression on Principal Components
BMD03M	Factor Analysis
BMD04M	Discriminant Analysis for Two Groups
BMD05M	Discriminant Analysis for Several Groups
BMD06M	Canonical Analysis
BMD07M	Stepwise Discriminant Analysis

Class R - Regression Analysis

Class R programs provide a relationship between a dependent variable and one or more independent variables. The relationship is estimated by least squares methods for linear relationships in programs 1R, 2R, and 3R and for polynomials in 5R. In program 4R, a trigonometric function is used, and in program 6R the exponential function $\alpha + B_p{}^x$ is used. Programs in this class are:

BMD01R	Simple Linear Regression
BMD02R	Stepwise Regression
BMD03R	Multiple Regression with Case Combinations
BMD04R	Periodic Regression and Harmonic Analysis
BMD05R	Polynomial Regression
BMD06R	Asymptotic Regression

Class S - Special Programs

There are two Class S special programs. The first can be used to construct contingency tables for one or more specified sets of intervals on each variable. Chi-Square, the contingency coefficient, and a maximum likelihood ratio are computed for each table and for tables collapsed on the bases of small cell expected values. The second program provides a generalized editing process (transgeneration) for preparing a modified set of punched cards (or card images on magnetic tape) for input to other BMD programs. The programs in this class are as follows:

BMD02S	Contingency Table Analysis
BMD09S	Transgeneration

Class T - Time Series Analysis

Two programs comprise the Class T for time series analysis. The first program measures the frequency, amplitude, and phase corresponding to a variety of records. The second program computes the autovariance, power spectrum, crossvariance, cross spectrum, transfer function, and coherence function of time series. These programs are:

BMD01T	Amplitude and Phase Analysis
BMD02T	Autocovariance and Power Spectral Analysis

Class V - Variance Analysis

Programs in Class V provide a simpler solution for various deviations from the standard simple analysis of variance programs. Programs in this class are:

BMD01V	Analysis of Variance for One-Way Design
BMD02V	Analysis of Variance for Factorial Design
BMD03V	Analysis of Covariance for Factorial Design
BMD04V	Analysis of Covariance with Multiple Covariates
BMD05V	General Linear Hypothesis
BMD06V	General Linear Hypothesis with Contrasts
BMD07V	Multiple Range Tests
BMD08V	Analysis of Variance for Hierarchical Design with Equal Cell Sizes

Class X - Supplement to Other BMD Programs

The Class X BMD programs is comprised of a collection of later releases in the six categories of programs just summarized. Their functions are described in Section VIII of this manual. The names of these programs in each category are as follows:

Class D -	BMDX70	t Program
	BMDX84	Asymmetric Correlation with Missing Data
Class M -	BMDX74	Identification of Outliers
	BMDX75	Canonical Analysis
Class R -	BMDX85	Nonlinear Least Squares
Class S -	BMDX76	Life Tables and Survival Rates
	BMDX77	Transgeneration
Class T -	BMDX68	Multiple Time Series Spectral Analysis
	BMDX92	Time Series Spectrum Estimation
Class V -	BMDX63	Multivariate General Linear Hypothesis
	BMDX64	General Linear Hypothesis
	BMDX69	Multivariate Analysis of Variance and Covariance

Descriptions of the foregoing BMD programs adapted for processing on the Series 6000 systems appear in Sections II through VIII of this manual.

BMD PROGRAM TAPES

The BMD programs are supplied by the Honeywell Program Library on two magnetic tapes. One tape contains the source programs in input media conversion (IMCV) format suitable for conversion by the Series 6000 input program. The usage of an IMCV tape is described in the Series 6000 System Startup and Operation manual listed in the Preface. Users can modify these source programs to suit their specific requirements.

The other tape contains the BMD programs in object program format as a loader input (R*) file. The programs can be loaded from this tape to form a BMD user library. Both tapes contain all of the BMD programs listed under "Program Categories" and in Table 1-1. This table lists the pertinent facts that the BMD user must know to set up the programs for execution, such as program number and the corresponding job number (SNUMB), file codes, and the amount of main memory required to execute the program. The programs are not necessarily stored on the tapes in the sequence listed in this table.

IMCV Tape Copying

The IMCV magnetic tape contains approximately 40,000 card images. The usual procedure is to make a copy of the programs from the library-supplied IMCV tape on another tape, thus preserving the original program tape. The copying can be done with a program called into execution by the control card deck listed in Table 1-2. This deck causes the system to copy the IMCV tape and print a listing of the program data with the change numbers of both the input and output tapes on an execution report. This function is described in "Change Numbering a Tape" in Section VI of the Series 6000 Service Routines manual listed in the Preface.

Execution from IMCV Tape

The BMD programs on the IMCV tape are capable of executing the test data for each program described in the UCLA BMD reference manual. The standard procedures for running an IMCV tape on the Series 6000 systems are used, as described in "Using an IMCV Tape" in Section VI of the Series 6000 System Startup and Operation manual. Each sample program is executed under control of the central system's console typewriter via messages.

The operator must enter a job number for the desired BMD program in the message format START␣ssssss, where sssss is the job number (SNUMB) of from 00010 through 00480. These numbers are listed in the "Job Number" column of Table 1-1. The SNUMB entry is followed by a user identity card image in the format $ IDENT 6000, BMDxxx, where xxx is the last three characters of the desired BMD program number.

Table 1-1. Tape Requirements of BMD Programs

Program Number	Job (SNUMB) Number	File Codes	Memory Required
BMD01D	00010	- -	19K
BMD02D	00020	01	39K
BMD03D	00030	- -	38K
BMD04D	00040	- -	38K
BMD05D	00050	- -	35K
BMD06D	00060	- -	35K
BMD07D	00070	- -	39K
BMD08D	00080	- -	36K
BMD09D	00090	- -	37K
BMD10D	00100	- -	35K
BMD11D	00110	- -	33K
BMD01M	00120	- -	34K
BMD02M	00130	- -	34K
BMD03M	00140	03	35K
BMD04M	00150	- -	34K
BMD05M	00160	- -	38K
BMD06M	00170	02, 03	26K
BMD07M	00180	01, 02	41K
BMD01R	00190	- -	31K
BMD02R	00200	01, 02	30K
BMD03R	00210	02, 03	31K
BMD04R	00220	- -	39K
BMD05R	00230	- -	27K
BMD06R	00240	01	37K
BMD02S	00250	03	36K
BMD09S	00260	02	25K
BMD01T	00270	- -	34K
BMD02T	00280	- -	38K
BMD01V	00290	- -	38K
BMD02V	00300	02, 03	37K

Table 1-1 (cont). Tape Requirements of BMD Programs

Program Number	Job (SNUMB) Number	File Codes	Memory Required
BMD03V	00310	02	36K
BMD04V	00320	02, 03	27K
BMD05V	00330	02	39K
BMD06V	00340	02, 03	28K
BMD07V	00350	01	35K
BMD08V	00360	01, 02	33K
BMDX63	00370	- -	31K
BMDX64	00380	01	31K
BMDX68	00390	- -	16K
BMDX69	00400	- -	34K
BMDX70	00410	01	32K
BMDX74	00420	01	30K
BMDX75	00430	01, 02	27K
BMDX76	00440	01, 02, 04	32K
BMDX77	00450	01	30K
BMDX84	00460	01	34K
BMDX85	00470	01	39K
BMDX92	00480	01	38K

Table 1-2. Card Deck for Copying IMCV Tape

Card Columns ⟶

1	8	16	
$	SNUMB	-----	
$	IDENT	-----	(local user identity)
$	PROGRAM	SCED	
$	TAPE	IN, X1DD, , nnnn, , BMDTAPE	(nnnn is the tape serial number)
$	TAPE	OT, X2DD, , , , NEWTAPE	
$	TAPE	P*, X3S	
$	BREAK		
$	CONVER		
$	TAPE	IN, X3R	
$	PRINT	OT	
$	ENDJOB		
***EOF			

Loader Input (R*) File

The loader input (R*) file tape as supplied by the Honeywell Program Library contains the object card deck images for all of the BMD programs. Each program has a number of subroutines and these are in the same sequence on the tape in which they are called during execution of a program. The R* tape has been created with the Series 6000 file edit system.

The R* tape can be used for execution of BMD programs, although the user may find it more convenient to store these programs on a file as a BMD user library. The programs can be loaded on a permfile using the standard Series 6000 permfile creation and loading procedures. Programs are executed with the deck of system control cards listed in Table 1-3. The function of each card and the data in certain fields to be supplied by the user is indicated under "Description". Additional information on control cards is published in the Control Cards reference manual listed in the Preface.

Table 1-3. Card Deck for Program Execution From R* Tape

Card Columns → 1	8	16	Description
$	SNUMB	-----	Job number for the BMD program (Table 1-1).
$	IDENT	-----	Local user identity.
$	OPTION	FORTRAN	Calls the FORTRAN language processor.
$	LIBRARY	BD	Informs the system general loader that a user's library is used; can be on tape, disk, or drum, depending on system configuration. If a file is used, 90 links of storage are required.
$	USE	BMDxxx	Calls the BMD program from the user's library, where xxx are the last three characters in BMD program numb
$	ENTRY	BMDxxx	Specifies the point in program at which execution begins. For simplicity, the program name is used for the label in user's library and the entry point name.
$	EXECUTE		Causes execution of program to begin.
$	LIMITS	25, nnK, xxxx	Specifies extent of system resources to be allocated to the run, such as: First field specifies 25 one-hundredths of an hour as maximum processor time. nnK specifies maximum amount of main memory required for program; must be at least the amount given in Table 1-1. xxxx is the amount of shared memory.
$	TAPE	BD, D1DD, , nnnn, , BMD-R*	Specifies the R* tape on which the BMD programs are stored, where nnnn is the tape serial number. Other $ TAPE cards, or $ FILE cards, can be used to specify other files required during execution of a program, such as for library, alternate input, alternate output, or working storage ("scratch") files. See the Control Cards reference manual listed in the Preface for additional information.
.			Insert data cards for BMD problem here.
.			
.			
$	ENDJOB		Signals the end of the control card deck for this program.

LISTING PROGRAM

A simple FORTRAN program can be prepared to print listings and punch decks of output cards for selected BMD programs. It is intended as an aid to becoming familiar with the preparation of input for the user's own BMD programs. The user prepares the deck of system control and source cards shown in Table 1-4 and assembles them into the order shown in the list. This deck is preceded by the usual $ SNUMB and $ IDENT cards containing the job number and user-assigned identity information.

FILE USAGE

The files used with the BMD programs are usually stored on magnetic tape, but the user can store them on other media (such as disk or drum storage) if desired. Since the BMD programs operate in a FORTRAN environment, the files are specified according to the standards given in the FORTRAN manual listed in the Preface.

With the exception of the library file, the file codes (see "File Designation" in Section V of the FORTRAN reference manual) contain two numeric characters. File code 05 is reserved for standard system input and file code 06 is reserved for standard system output.

Nearly every BMD program is capable of reading input data from previously generated files. One program (BMD04D) can be requested to write its output via an alternate file. The input file options, if any, and the scratch files with their logical numbers are listed in the "File Codes" column in Table 1-1.

The amount of main memory required for each BMD program must be specified on the $ LIMITS card. This figure is given in the "Memory Required" column in Table 1-1. The amount of memory required may change as updated Series 6000 system software is released.

Table 1-4. Sample Deck for Listing and Punching IMCV Tape

```
$       OPTION  FORTRAN
$       FORTY
        SUBROUTINE BMDPRG
*       PROGRAM TO LIST AND PUNCH PROGRAMS FROM BMD IMCV TAPE
        DIMENSION IA(14)
        DATA IB/6H      /
        DATA ID/6H IDENT/
        CALL FLGEOF(10,N)
        REWIND 10
        READ(5,10) IC
    5   READ(10,10) IA
        IF(N.NE.0) GO TO 40
   10   FORMAT(3A6,A2,10A6)
        IF(IA(2).NE.ID) GO TO 5
   20   IF(IA(5).NE.IC) GO TO 5
   30   WRITE(6,70) IA
        WRITE(43,10) IA
        READ(10,10) IA
        IF(N.NE.0) GO TO 40
        IF(IA(2).NE.ID) GO TO 30
        READ(5,10) IC
        WRITE(6,60)
        IF(IC.EQ.IB) STOP
        GO TO 20
   40   WRITE(6,50)
        STOP
   50   FORMAT(1H1/////35HOEOF ENCOUNTERED READING INPUT TAPE)
   60   FORMAT(1H1)
   70   FORMAT(1H 3A6,A2,10A6)
        END
$       EXECUTE
$       LIMITS  05,8K,,5000
$       TAPE    10,X1DU,,8995,,BMD
BMD04D

$       ENDJOB
***EOF
```

SECTION II

CLASS D – DESCRIPTION AND TABULATION

BMD01D – SAMPLE DATA DESCRIPTION

General Description

a. This program computes simple averages and measures of dispersion of variables, omitting those values which the user specifies for exclusion from the computations.

Methods for specifying the exclusion of certain special values from the computations are given below.

Method Number	Method
0	Set all blanks equal to 0; these and all other numbers will enter computations.
1	Not implemented.
2	Not implemented.
3	Pre-specified special values not counted; blanks set equal to 0 and entered with all other numbers into computations.

b. Output for this program includes:

(1) Means
(2) Standard deviations
(3) Standard errors of means
(4) Maximum values
(5) Minimum values
(6) Ranges
(7) Sample sizes (see the four methods listed above).

c. Limitations per problem:

(1) p, number of variables ($p \leq 999$)
(2) n, number of cases ($n \leq 99{,}999$)
(3) k, number of Variable Format Cards ($1 \leq k \leq 10$)
(4) c, number of special values specified for methods 2 or 3 ($0 < c \leq 8$)
(5) t, number of Transgeneration Cards ($0 \leq t \leq 100$)
(6) q, number of variables added after transgeneration ($-998 \leq q \leq 999$)

d. The program allows transgeneration of the input data. Codes 1-16 and 40 of the transgeneration list may be used.

BMD02D – CORRELATION WITH TRANSGENERATION (Boolean Selection of Cases)

General Description

a. This program computes simple correlation coefficients, averages and measures of dispersion on entering variables and/or transgenerated variables from selected cases whose values for specified variables have a precise logical relationship in agreement with a specified Boolean expression.

b. Output from this program includes:

(1) Sums
(2) Means
(3) Cross-product deviations
(4) Standard deviations
(5) Variance--covariance matrix
(6) Correlation matrix

Optional output includes:

(7) One-page cross-tabulation plots of any two variables, automatically scaled to 50 (vertical) by 100 (horizontal) character spaces or units.

c. Limitations per problem:

(1) p, number of original variables ($2 \leq p \leq 150$)
(2) n, number of original cases ($2 \leq n \leq 99,999$)
(3) j, number of Plot Selection Cards ($0 \leq j \leq 99$)
(4) q, number of variables added to the original set after transgeneration ($-148 \leq q \leq 148$)
(5) b, number of Case Selection Cards ($0 \leq b \leq 9$)
(6) m, number of Transgeneration Cards ($0 \leq m \leq 150$)
(7) k, number of Variable Format Cards ($1 \leq k \leq 10$)

d. The program allows transgeneration of the input data. Codes 01, 02, ..., 16 and 41 of the transgeneration list may be used.

e. A special feature of this program is the selection of cases from the input data by specifying a Boolean expression. A case is accepted if it is in agreement with the expression; otherwise the case is skipped. The expression consists of variables and constants involving relationships of equality or inequality written in a logical form using the operations AND and OR.

BMD03D – CORRELATION WITH ITEM DELETION

General Description

a. This program computes a simple correlation matrix, omitting values of variables that the user specifies to be deleted. Blanks as well as other codes may be specified for deletion, e.g., to indicate missing values.

b. Output from this program includes:

(1) Correlation matrix printed--punched cards or BCD tape output optional.
(2) Number of pairs of observations used in computing each correlation coefficient.

c. Limitations per problem:

(1) p, number of variables ($1 \leq p \leq 90$)
(2) n, number of cases ($2 \leq n \leq 99,999$)
(3) c, number of deletion codes (for each variable) ($0 \leq c \leq 10$)
(4) k, number of Variable Format Cards ($1 \leq k \leq 10$)
(5) d, number of Deletion Code Cards ($0 \leq d \leq 90$)
(6) m, number of Transgeneration Cards ($0 \leq m \leq 99$)

d. The program allows transgeneration of the input data. Codes 01-10 of the transgeneration list may be used.

BMD04D – ALPHANUMERIC FREQUENCY COUNT (One Column Data)

General Description

a. This program computes frequencies of legal characters on one-column data. Any numeric, alphabetic, or special character is counted. The input may be on cards, or on magnetic tape in BCD mode.

b. The output for this program consists of a frequency count, by column

c. Limitations per problem:

(1) n, number of cases ($1 \leq n \leq 99,999$)
(2) p, number of columns (variables) ($1 \leq p \leq 400$)
(3) k, number of Variable Format Cards ($1 \leq k \leq 10$)

BMD05D – GENERAL PLOT INCLUDING HISTOGRAM

General Description

a. This program provides a method by which graphs and histograms can be produced.

b. Output for this program includes:

(1) GRAPHS. Two methods of plotting are available:

(a) The first method gives a one page graph which has 50 units vertically and 100 units horizontally. The points are automatically scaled to conform to these dimensions, and a scale is printed both horizontally and vertically. The points (data cards) need be in no special order.

(b) The second method gives a multiple-page graph with as many units vertically as there are values of the base variable. The values of the base variable (data cards) must be ordered and consecutive. The base variable is not scaled. The cross variables are scaled by the computer to conform to a horizontal dimension of 100 units.

(2) HISTOGRAMS

A one-page histogram can be produced, with a maximum of 34 intervals. The width of the interval must be specified; however, if the specified width would result in more than 34 intervals, the program will print comments to this effect and will compute a new width which will give exactly 34 intervals. Scales are printed on the vertical and horizontal axes.

c. Limitations per problem:

(1) p, number of original variables ($1 \leq p \leq 500$)
(2) n, number of cases ($2 \leq n \leq 20000$)
(3) q, number of variables added to the original set after transgeneration ($-499 \leq q \leq 499$)
(4) p+q, total number of variables ($1 \leq p+q \leq 500$)
(5) (p+q)n, total number of data ($2 \leq (p+q)n \leq 20000$)
(6) m, number of Transgeneration Cards ($0 \leq m \leq 999$)
(7) k, number of Variable Format Cards ($1 \leq k \leq 10$)

d. This program allows transgeneration. Codes 01, 02, ..., 14 of the transgeneration list may be used.

BMD06D – DESCRIPTION OF STRATA

General Description

a. Cases are separated into groups based on specified intervals of one variable, the conditioning variable. For these selected groups, computations are performed for any other variables which are designated as the conditioned variables.

b. Output for this program includes:

(1) Frequencies, means, variances, standard deviations, and standard errors of means for a specified set of variables conditioned on one variable within the set.

(2) Correlation coefficients for the specified set of variables.

c. Limitations per problem:

(1) p, number of original variables ($2 \leq p \leq 30$)

(2) n, sample size or number of cases ($3 \leq n \leq 700$)

(3) m, number of Transgeneration Cards ($0 \leq m \leq 99$)

(4) q, number of variables added to original set after transgeneration ($-9 \leq q \leq 28$) ($2 \leq (p+q) \leq 30$)

(5) l, number of Selection Cards ($1 \leq l \leq 99$)

(6) k, number of Variable Format Cards ($1 \leq k \leq 10$)

d. This program allows transgeneration. Codes 01 through 14 of the transgeneration list may be used.

BMD07D — DESCRIPTION OF STRATA WITH HISTOGRAMS

General Description

a. This program groups the data into a specified number of groups based on the order of entry of the data or into groups whose values for a base variable are within intervals established by specified cut points. For these groups, histograms are printed for each variable. The number of classes or categories of the histograms may be specified or they may be computed by the program. Means, standard deviations, and correlation coefficients are computed for each group; means and standard deviations are computed also for the combined groups of a variable. Special values may be specified for all variables, except the base variable, to exclude certain values or codes from computations.

b. Output from this program includes:

(1) Input data after transgeneration.

(2) Input data after ordering from high to low on the specified base variable.

(3) Histograms for each variable showing the frequencies of distribution of c classes over the g groups.

(4) Correlation matrices for each group.

(5) Means and standards deviations.

(6) Tabulations of special values.

c. Limitations per problem:

(1) p, number of variables ($1 \leq p \leq 100$)

(2) n, number of observations ($1 \leq n \leq 9999$)

Note: In no problem may the product of n and p+q, n(p+q) be greater than 19,000.

(3) g, number of groups ($0 \leq g \leq 10$)

(4) c, number of classes ($5 \leq c \leq 30$)

(5) M_i, number of special values for variable i($0 \leq M_i \leq 5$)

(6) q, number of variables added in transgeneration ($0 \leq q \leq 99$)

(7) k, number of Variable Format Cards ($1 \leq k \leq 10$)

(8) TG, number of Transgeneration Cards ($0 \leq TG \leq 99$)

(9) T, number of group cut points ($1 \leq T \leq 9$)

d. The program allows transgeneration of the input data. Codes 01-23 of the transgeneration list may be used.

BMD08D — CROSS-TABULATION WITH VARIABLE STACKING

General Description

a. This program computes two-way frequency tables of data input. Frequency tables are computed from specified ranges of the original variables, variables after transgeneration, stacked variables, or combinations of these. Data input may be positive or negative integers only. The program will not accept data input card fields which have a punched decimal point.

b. Output from this program includes:

(1) Frequency tables of all combinations of the variables or only those specified by the user.

(2) Chi-square values and degrees of freedom for each table.

(3) Means, standard deviations, and correlation coefficients for each pair of variables.

c. Limitations per problem:

(1) p, number of original variables ($2 \leq p \leq 100$)

(2) N, number of cases ($2 \leq N \leq 1500$)

(3) q, number of variables added to the original set after transgeneration or stacking ($-98 \leq q \leq 98$), ($p+q \leq 100$)

(4) $(p+q)\,N$, total data input ($4 \leq (p+q)\,N \leq 19000$)

(5) m, number of Transgeneration Cards ($0 \leq m \leq 99$)

(6) k, number of Variable Format Cards ($1 \leq k \leq 10$)

(7) s, number of Selection Cards ($0 \leq s \leq 99$)

(8) The range of each variable to be cross-tabulated is specified with the following restrictions:

$$1 \leq \text{Max.X} - \text{Min.X} \leq 34$$
$$1 \leq \text{Max.Y} - \text{Min.Y} \leq 99$$

where X and Y are the abscissa and ordinate respectively. The program generates a continuous range for each variable from which a frequency matrix is computed on any two variables. Designation of abscissa or ordinate is arbitrary within the above restrictions.

(9) Only those rows and columns which have non-zero entries are printed unless the user indicates on the Problem Card that rows and columns which have zero entries are to be printed.

(10) All values outside the specified range for each variable are listed in the output under the heading, VALUES NOT ENTERED, if the number of such values is less than 50. Otherwise, only the number is printed.

(11) The maximum frequency of each point when cross-tabulated is 999. If any frequency should be greater than 999, it will be set equal to 999.

d. This program allows transgeneration. Codes 08...13,15,16,40, and 41 of the transgeneration list may be used.

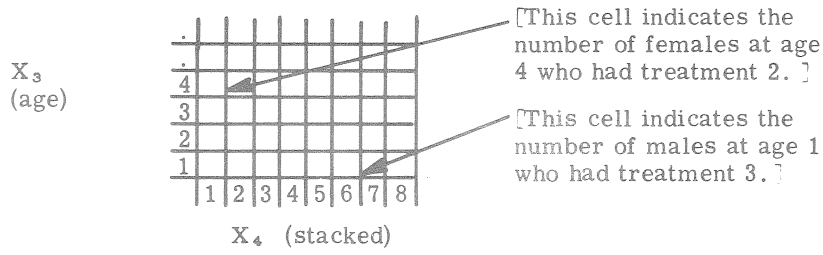
e. A special feature of this program is the Variable Stacker. In addition to the above codes, four stacker codes are available. Codes 82,83,84 and 85 stack a nested classification of several variables into a single variable. Suppose, for example, the user wishes a frequency table of males who had treatment 1,2,3 or 4 and females who had treatment 1,2,3 or 4 cross-tabulated with age. Let X_1 be the treatment variable having values (1,2,3,4) X_2 be the sex variable having values (0,1) and X_3 be the age variable having values (1,2,3, ...,99). Code 82 stacks X_1 into X_2 generating X_4 which has values (1,2,3,4,5,6,7,8) as illustrated in the table below.

									Variable
Sex	0				1				X_2
Treatment	1	2	3	4	1	2	3	4	X_1
Output Code	1	2	3	4	5	6	7	8	X_4

The user may then add the usual type of heading material, such as:

Female				Male			
T_1	T_2	T_3	T_4	T_1	T_2	T_3	T_4

The frequency table when X_4 is cross-tabulated with X_3:



Five levels of stacking are permissible.

BMD09D – CROSS-TABULATION, INCOMPLETE DATA

General Description

a. This program performs cross-tabulations of input data, excluding from cross-tabulation specified special values or codes used to designate missing values. The program differentiates between blanks and zeros in testing for special values. Different special values or codes may be used for each variable. Specified missing values excluded from cross-tabulation are identified in the output. Special codes -999 and -999.00999 are printed in the output to designate a missing value which occurs for a transgenerated variable as a result of the transgeneration of a variable containing a missing-value code for that case.

b. Output for this program includes:

(1) Input data matrix
(2) Cross-tabulations of selected variables
(3) Identification of cases containing missing-value codes
(4) Correlation coefficient.

c. Limitations per problem:

(1) n, sample size or number of cases ($2 \leq n \leq 2000$)

(2) p, number of original variables ($1 \leq p \leq 100$)

(3) m, number of Transgeneration Cards ($0 \leq m \leq 999$)

(4) q, number of variables added after transgeneration ($-99 \leq q \leq 99$)

(5) If printing of the identification number in the data matrix is desired $\boxed{n(p+q) \leq 15000}$ If it is not desired $\boxed{n(p+q) \leq 17000}$

(6) S, number of Selection Cards ($1 \leq S \leq 99$)

(7) k, number of Variable Format Cards ($1 \leq k \leq 10$)

(8) V_i, number of missing-value codes for the i^{th} variable ($0 \leq V_i \leq 10$)

d. This program allows transgeneration. Codes 01 through 14 of the transgeneration list may be used.

e. The program computes the correlation coefficient of the two selected variables.

BMD10D – DATA PATTERNS FOR DICHOTOMIES

General Description

a. This program finds frequencies and patterns of any one particular specified code in the input data. Frequent use will be for a code representing missing values. The program prints 0's to designate the specified code, or missing values, and 1's to designate all other values. A data matrix of 0's and 1's is printed. Frequencies of the specified code, or missing values, are computed; and the cases having the specified code are identified by item numbers which correspond to the order in which the cards appear in the data input deck. If desired, patterns of data may be obtained also after eliminating each variable in turn. Thus, patterns are available for the p + 1 different choices of p variables.

b. Output for this program includes:

(1) Patterns of data (0's for the specified code, or missing values; 1's for all other values) tabulated by numbers of missing values, and item numbers to identify cases.

(2) Data matrix of 0's and 1's.

c. Limitations per problem:

(1)	p,	number of variables	$(1 \leq p \leq 30)$
(2)	n,	number of items or cases	$(1 \leq n \leq 700)$
(3)	k,	number of Variable Format Cards	$(1 \leq k \leq 10)$

BMD11D – DATA PATTERNS FOR POLYCHOTOMIES

General Description

a. This program prints patterns of one-column data and item numbers or case numbers to identify cases having these data patterns. If desired, original data may be recoded by the program before the data patterns are printed.

b. Output for this program includes:

(1) Data patterns for original codes or, for recoded data, frequencies of patterns, and item numbers to identify cases having these patterns.

(2) Data patterns for original or recoded data and identification codes for all cases printed in the order in which they appear in the data input deck.

c. Limitations per problem:

(1) p, number of variables $(1 \leq p \leq 25)$
(2) n, number of items or cases $(1 \leq n \leq 700)$
(3) k, number of Variable Format Cards $(1 \leq k \leq 10)$

SECTION III

CLASS M – MULTIVARIATE ANALYSIS

BMD01M – PRINCIPAL COMPONENT ANALYSIS

General Description

a. This program computes the principal components of standardized data and separately rank orders each standardized case by the size of each principal component.

b. Output from this program includes:

(1) Correlation coefficients

(2) Eigenvalues, including cumulative proportion of total variance

(3) Eigenvectors (principal components of standardized data)

(4) Rank order of each standardized case, ordered by size of each principal component separately.

c. Limitations per problem:

(1) p, number of original variables $(2 \leq p \leq 25)$
(2) n, number of cases $(3 \leq n \leq 400)$
(3) q, number of variables added to the original set after transgeneration $(-23 \leq q \leq 23)$
(4) p+q, total number of variables $(2 \leq p+q \leq 25)$

d. This program allows transgeneration. Codes 01, 02, ..., 14 from the transgeneration list may be used.

BMD02M – REGRESSION ON PRINCIPAL COMPONENTS

General Description

a. This program computes the principal components of standardized data and separately rank orders each standardized case by the size of each principal component. Each dependent variable is regressed on the first, first two, first three, and all principal components when each component is expressed in terms of standardized data.

b. Output from this program includes:

(1) Correlation coefficients.

(2) Eigenvalues, including cumulative proportion of total variance.

(3) Eigenvectors (principal components of standardized data).

(4) Rank order of each standardized case ordered by size of each principal component separately.

(5) Regression coefficients.

(6) Reduction in sum of squares of the residuals due to using principal components.

c. Limitations per problem:

(1) p, number of independent variables ($1 \leq p \leq 25$)
(2) d, number of dependent variables ($1 \leq d \leq 20$)
(3) t, total number of original variables ($2 \leq t \leq 45$)
(4) n, number of cases ($3 \leq n \leq 200$)

d. This program allows transgeneration. Codes 01, 02, 03, ..., 14 may be used.

BMD03M – GENERAL FACTOR ANALYSIS

General Description

a. This program performs a principal component solution and an orthogonal rotation of the factor matrix.

Data input to this program may be in the form of raw data, a correlation matrix, or a factor matrix. Data input may be read in from punched cards, BCD tape, or binary tape.

Communalities are estimated from the squared multiple correlation coefficients or the maximum absolute row values, or they may be specified by the user and will replace the set of r_{ii} in the main diagonal of the correlation matrix.

b. Output from this program includes:

(1) Means and standard deviations
(2) Correlation matrix
(3) Eigenvalues including cumulative proportions of total variance
(4) Eigenvectors
(5) Factor matrix
(6) Factor check matrix
(7) Orthogonal rotated factor matrix
(8) Original and successive variances
(9) Check on communalities
(10) Factor scores printed or on tape

(1)-(10) are included if input is raw data
(1)-(9) are included if input is raw data
(2)-(9) are included if input is a correlation matrix.
(5), (7)-(9) are included if input is a factor matrix.

c. Limitations per problem:

(1) p, number of variables ($2 \leq p \leq 80$)
(2) n, number of cases ($p \leq n \leq 9999$)
(3) m, maximum number of factors (to be rotated) ($2 \leq m \leq p$)
(4) k, number of Variable Format Card(s) ($1 \leq k \leq 6$)

d. The number of factors to be rotated is determined by the smaller of the following two numbers:

(1) the number of eigenvalues which are greater than c ($c \geq 0$), where c is specified by the user, or

(2) the number of factors specified by the user.

BMD04M – DISCRIMINATE ANALYSIS FOR TWO GROUPS

General Description

a. This program computes a linear function of p variables measured on each individual of two groups. This function can serve as an index for discrimination between the groups. It is determined from the criterion of "best" in that the difference between the mean indices for the two groups divided by a pooled standard deviation of the indices is maximized.

b. Output from this program includes:

(1) Table of means of the variables by group and the mean differences.
(2) Matrix of the sum of the cross-products of the deviations from the means.
(3) Inverse of the matrix in (2).
(4) Discriminant function coefficients.
(5) Mahalanobis' D^2 and associated F-Statistic.
(6) Table of the mean variance and standard deviation of $Z_{i\alpha}$ for i = 1,2.
(7) Table of $Z_{i\alpha}$ arranged in order of algebraic size.

c. Limitations per problem:

(1) n_i, sample size in i^{th} group $(p+q \leq n_i \leq 300)$

(2) p, number of original variables $(2 \leq p \leq 25)$

(3) t, number of Transgeneration Cards $(0 \leq t \leq 99)$

(4) q, number of variables added to the original set of variables $(-9 \leq q \leq 23)$, $2 \leq (p+q) \leq 25$

(5) l, number of Selection Cards $(1 \leq 99)$

(6) k, number of Variable Format Cards $(1 \leq k \leq 10)$

d. This program allows transgeneration. Codes 01 through 14 of the transgeneration list may be used.

BMD05M – DISCRIMINANT ANALYSIS FOR SEVERAL GROUPS

General Description

a. This program directs the computation of a set of linear functions for the purpose of classifying an individual into one of several groups. The input data consist of a set of observations for each of the classification groups; each observation consists of the values of a set of variables, and each observation contains a value for each of the variables.

The group assignment procedure followed is derived from a model of a multivariate normal distribution of observations within groups such that the covariance matrix is the same for all groups. An individual is classified into the group for which the estimated probability density is largest. The equivalent computational procedure followed evaluates the computed linear function corresponding to each of the groups and assigns an individual to the group for which the value is largest.

The hypothesis, that group means are the same, is tested.

b. Output for this program includes:

(1) Mean scores
(2) Matrix of cross-products of deviation from means
(3) Dispersion matrix
(4) Inverse of dispersion matrix
(5) D-square statistic
(6) Coefficients and constants
(7) Evaluation of classification function for each case
(8) Classification

c. Limitations per problem:

(1) g, number of groups ($2 \leq g \leq 5$)

(2) p, number of original variables ($g \leq p \leq 25$)

(3) n_i, sample size or number of cases in any one group ($n_i \leq 175$)

(4) t, number of Transgeneration Cards ($0 \leq t \leq 99$)

(5) q, number of variables added to the original set after transgeneration ($-9 - q \leq 25$) $g \leq (p+q) \leq 25$

(6) k, number of Variable Format Cards ($1 \leq k \leq 10$)

d. This program allows transgeneration. Codes 01 through 14 of the transgeneration list may be used.

BMD06M – CANONICAL ANALYSIS

General Description

a. This program computes the canonical correlations between two sets of variables.

b. Output from this program includes:

(1) Data input (optional)
(2) Simple correlation matrix
(3) Standard deviations
(4) Canonical correlations
(5) Canonical coefficients for standardized variables.
(6) Check matrix.

c. Limitations per problem:

(1) t, total number of entering variables ($t \leq 100$)
(2) p, number of variables of the first set, after transgeneration, if any ($p \leq 35$)
(3) q, number of variables of the second set, after transgeneration, if any ($p \leq q \leq 35$)
(4) n, number of cases ($n \leq 9999$)
(5) k, number of Variable Format Card(s) ($1 \leq k \leq 10$)
(6) m, number of Transgeneration Cards ($0 \leq m \leq 99$)

d. This program allows transgeneration of the data input. Codes 01 to 17 and 41 of the transgeneration list may be used.

BMD07M – STEPWISE DISCRIMINANT ANALYSIS

General Description

a. This program performs a multiple discriminant analysis in a stepwise manner. At each step one variable is entered into the set of discriminating variables. The variable entered is selected by the first of the following equivalent criteria:

(1) The variable with the largest F value (see computational procedure)

(2) The variable which when partialed on the previously entered variables has the highest multiple correlation with the groups

(3) The variable which gives the greatest decrease in the ratio of within to total generalized variances

A variable is deleted if its F value becomes too low. The program also computes canonical correlations and coefficients for canonical variables. It plots the first two canonical variables to give an optimal two-dimensional picture of the dispersion.

b. The output consists of:

(1) Group means and standard deviations

(2) Within groups covariance matrix

(3) Within groups correlation matrix

(4) At each step:

(a) Variables included and F to remove

(b) Variables not included and F to enter

(c) U statistic and approximate F statistic to test equality of group means

(d) Matrix of F staistics to test the equality of means between each pair of groups

(5) At certain specified steps and after the last step:

(a) Discriminant functions

(b) Classification matrix

(6) For each case:

(a) The posterior probability of coming from each group

(b) Square of the Mahalanobis distance from each group

(7) Summary table. For each step of the procedure the following is tabulated:

(a) Variable entered or removed

(b) F value to enter or remove

(c) Number of variables included

(d) U statistic

(8) Eigenvalues, canonical correlation, and coefficients of canonical variables

(9) Plot of the first canonical variable against the second

c. Limitations per problem:

(1) p, number of variables ($1 \leq p \leq 80$)

(2) t, total number of groups ($2 \leq t \leq 80$)

(3) j, number of Variable Format Card(s) ($1 \leq j \leq 16$)

SECTION IV

CLASS R — REGRESSION ANALYSIS

BMD01R — SIMPLE LINEAR REGRESSION (One Way Analysis of a Covariance)

General Description

a. This program performs simple linear regression analysis on single or combined treatment groups with unequal sample sizes. (The words "treatment groups" are used here to describe categories.) The "within" cross-products sums and coefficients are computed; thus, analysis-of-covariance information is also provided in the output.

b. Output from this program includes:

(1) Sum of squares and products for treatment means, within, and total with degrees of freedom.

(2) Deviations about regression for within, total, and adjusted means with degrees of freedom.

(3) Regression coefficients for treatment means, within, and total.

(4) F ratios for treatment means, adjusted means, and within coefficients with degrees of freedom.

(5) Bioassay information (optional).

c. Limitations per problem:

(1) t, number of treatment groups ($1 \leq t \leq 999$)
(2) n_i, number of cases in the i^{th} treatment group ($1 \leq n_i \leq 999$)
(3) s, number of subset specifications ($0 \leq s \leq 500$)
(4) c, number of combinations of subsets ($0 \leq c \leq 99$)
(5) k, number of Variable Format Cards ($1 \leq k \leq 10$)

d. This program allows transgeneration of either the dependent variable or independent variable or of both. Codes 1-10 from the transgeneration list may be used.

e. Subsets and combinations of subsets can be selected from the input data as illustrated in the following example:

X = Pre-treatment measurement (or independent variable)
Y = Post-treatment measurement (or dependent variable)

Group A	Group B	Group C	Group D	etc.
$Y_1 X_1$	$Y_1 X_1$	$Y_1 X_1$	$Y X_1$	.
. .	. .	. .	. .	.
. .	. .	. .	. .	.
. .	. .	. .	. .	.
$Y_1 X_a$	$Y_b X_b$	$Y_1 X_c$	$Y_2 X_d$	.

Subset Specification

Subset Number	Group(s)
1	A
2	B, C, D
3	E, F, G
4	G
.	.
.	.
.	.

An analysis-of-variance table can be computed for each subset or one table for the combined subsets. If any additional tables are desired for combinations of these subsets, they are specified as follows:

Combination of Subsets

Combination Number	Subsets Included
1	1, 2, 3
2	1, 4
3	.
.	.
.	.
.	.

BMD02R - STEPWISE REGRESSION

General Description

a. This program computes a sequence of multiple linear regression equations in a stepwise manner. At each step one variable is added to the regression equation. The variable added is the one which makes the greatest reduction in the error sum of squares. Equivalently it is the variable which has highest partial correlation with the dependent variable partialed on the variables which have already been added; and equivalently it is the variable which, if it were added, would have the highest F value. In addition, variables can be forced into the regression equation. Non-forced variables are automatically removed when their F values become too low. Regression equations with or without the regression intercept may be selected.

b. Output from this program includes:

(1) At each step:
- (a) Multiple R
- (b) Standard error of estimate
- (c) Analysis-of-variance table
- (d) For variables in the equation:
  1. Regression coefficient
  2. Standard error
  3. F to remove
- (e) For variables not in the equation:
  1. Tolerance
  2. Partial correlation coefficient
  3. F to enter

(2) Optional output prior to performing regression:
- (f) Means and standard deviations
- (g) Covariance matrix
- (h) Correlation matrix

(3) Optional output after performing regression:
- (i) List of residuals
- (j) Plots of residuals vs. input variables
- (k) Summary table

c. Limitations per problem:

(1) p, number of original variables ($2 \leq p \leq 80$)
(2) q, number of variables added by transgeneration ($-9 \leq q \leq 78$)
(3) p+q, total number of variables ($2 \leq p+q \leq 80$)
(4) s, number of Sub-problem Cards ($1 \leq s \leq 99$)
(5) k, number of Variable Format Cards ($1 \leq k \leq 10$)
(6) i, number of variables to be plotted ($0 \leq i \leq 30$)
(7) n, number of cases ($1 \leq n \leq 9999$)
(8) m, number of Transgeneration Cards ($0 \leq m \leq 99$)

d. This program allows transgeneration of the variables. Codes 01-17 and 20-24 of the transgeneration list may be used.

Order of Cards in Job Deck

Cards indicated by letters enclosed in parentheses are optional. All other cards must be included in the order shown.

a. System Cards [Introduction, IV]

b. Problem Card

(c.) Transgeneration Card(s) [Introduction, III-B]

(d.) Labels Card(s) [Introduction, III-A]

e. F-type Variable Format Cards [Introduction, III-C]

(f.) DATA INPUT Cards [Introduction, II-C]
(Place data input deck here if data input is from cards.)

g. Sub-problem Card(s)

(h.) Control-Delete Card(s)

(i.) Index-Plot Card(s)

BMD03R – MULTIPLE REGRESSION WITH CASE COMBINATIONS

General Description

a. This program performs multiple regression and correlation analyses on the data within each selection of subsamples from the same population. A selection may be any specified set of subsamples. For example, if subsamples are classified according to some method of collecting data, we might wish to analyze as follows:

Selection 1	Subsamples 1, 2, 3, 4, 5, 6
Selection 2	Subsamples 1, 2, 3
Selection 3	Subsamples 4, 5, 6
Selection 4	Subsamples 1, 3, 5
...	

As illustrated above, any set of subsamples arranged in blocks, namely subsample 1 data together, subsample 2 data together, etc., may be combined.

b. Output for this program includes:

(1) Sums and sums of squares

(2) Cross-products of deviations

(3) Correlation matrix

(4) Inverse of correlation matrix

(5) Means and standard deviations

(6) Regression, coefficients, their standard errors and t-values

(7) Sums of squares and mean squares due to regression and deviation about regression, with degrees of freedom and F-value

(8) Sums of squares due to regression for each variable

(9) Standard error of estimate

(10) Intercept

(11) Partial correlation coefficients

(12) Multiple correlation coefficient, R and R^2

For each selection and for each subproblem, the following can be obtained:

(13) Table of residuals

(14) Analysis of extreme residuals

c. Limitations:

Limitations per problem:

(1) M, number of subsamples (1 ≤ M 28)
(2) p, original number of variables (2 ≤ p ≤ 50)
(3) N, total sample size of all subsamples combined (N ≤ 99,999)
(4) k, number of Variable Format Cards (1 ≤ k ≤ 10)
(5) S, number of Selection Cards (1 · S ≤ 99)

Limitations per selection:

(1) m, number of subsamples selected (1 ≤ m ≤ 28)

(2) p, original number of variables (2 ≤ p ≤ 50)

(3) q, number of variables added to the original set after transgeneration (-9 · q ≤ 48), (2 ≤ p+q ≤ 50)

(4) n, total sample size of subsamples selected (p+q ≤ n ≤ N)

(5) t, number of Transgeneration Cards (0 · t ≤ 50)

(6) r, number of Replacement-Deletion Cards (0 ≤ r ≤ 99)

Limitations per replacement-deletion:

(1) d, number of variables deleted (d ≤ 28)

Additional Limitation:

This program may give erroneous results when the correlation matrix is singular or near singular. There are no special routines in the program which test or make adjustments for singularity of the correlation matrix before or after computing the inverse matrix.

d. Within each selection of subsamples, the transgeneration feature may be used. Codes 01 through 17 of the transgeneration list are allowed.

e. Within each selection, Replacement and Deletion Cards may be used in multiple regression and correlation analyses to replace the dependent variable and to delete independent variables.

BMD04R — PERICDIC REGRESSION AND HARMONIC ANALYSIS

General Description

a. This program performs periodic or harmonic regression analysis using the regression function of the form.

$$Y_t = a_0 + \sum_{i=1}^{n} [\, a_i \cos(2\pi it/k) + b_i \sin(2\pi it/k) \,]$$

Periodic regressions are computed for successive values of n up to a maximum specified by the user.

b. Output from this program includes:

(1) Table of observed values, predicted values, and their differences for each regression

(2) Regression coefficients

(3) For the first harmonic, amplitude, phase angle, range, time of maximum and minimum.

(4) Printed plots of the observed and predicted values for each regression

(5) An analysis-of-variance table for each regression provided there are two or more replicates

(6) An analysis-of-covariance table (for each regression) for each set of covariate data

c. Limitations per problem:

(1) n, number of harmonics ($1 \leq n \leq 9$)

(2) k, number of time values ($2n+1 \leq k \leq 400$)

(3) m, number of replications (for analysis of variance) ($2 \leq m \leq 19$)
m, number of replications (no analysis of variance) ($1 \leq m \leq 19$)

(4) c, number of covariate data sets ($0 \leq c \leq 99$)

(5) l, number of Variable Format Card(s) ($1 \leq l \leq 10$)

d. This program allows special transgeneration of either the variate or covariate data set or both. Codes 01 to 10 of the transgeneration list may be used.

e. The number of harmonics, time values, and replications for the analysis of covariance is the same as that specified for the dependent variable. Only one covariate is allowed but as many as 99 covariate data sets may be processed for each variate (dependent variable) data set.

BMD05R – POLYNOMIAL REGRESSION

General Description

a. This program computes polynomial regressions of the form:

$$Y = \alpha + \beta_1 X + \beta_2 X^2 + \ldots . + \beta_k X^k + e \ (k^{th} \text{ degree})$$

where k is some positive integer.

b. Output for this program includes:

(1) For data input, after transgeneration:
  (a) Means
  (b) Correlation coefficient.

(2) For each successive degree of polynomial regression:
  (a) Intercept and regression coefficients
  (b) Standard errors of regression coefficients
  (c) Analysis-of-variance table.

(3) For the final degree of polynomial regression:
  (a) Analysis-of-variance table
  (b) Table of residuals
  (c) Plot of the observed values and the value predicted from the regression equation.

c. Limitations per problem:

(1) k, final degree polynomial regression ($k \leq 10$)
(2) n, number of cases ($k+1 \leq n \leq 500$)
(3) m, number of Variable Format Card(s) ($1 \leq m \leq 10$)

d. This program allows transgenerations of the dependent variable, if desired. Codes 01 through 10 may be used.

BMD06R — ASYMPTOTIC REGRESSION (Logistic, Gompertz, etc.)

General Description

a. This program performs asymptotic regression analysis using the modified exponential function of the form

$$Y = \alpha + \beta \rho^{x}$$

By making a transgeneration (or successive transgenerations) on the dependent variable Y, other equations can be used. For example:

The Gompertz Curve	Transgeneration Code(s)[1]
$Y_1 = e^{(\alpha + \beta \rho^{x})}$	17
$Y_2 = 10^{(\alpha + \beta \rho^{x})}$	3
The Logistic (or Growth) Curve	
$Y_3 = \frac{1}{\alpha + \beta \rho^{x}}$	7
$Y_4 = \frac{1}{1 + 10^{(\alpha + \beta \rho^{x})}}$	7,8,3
$Y_5 = \frac{c}{1 + e^{(\alpha + \beta \rho^{x})}}$	7,9,8,17

It is noted, however, that the computational procedure fits by least squares the regression equation $Y' = \alpha + \beta \rho^{x}$ where Y' is the transformed Y. Hence, the curve $Y = \alpha + \beta \rho^{x}$ is obtained by not performing a transgeneration. This is indicated by specifying transgeneration code 00.

b. Output from this program includes:

(1) Estimates of parameters, their standard errors and variance-covariance matrix.
(2) Analysis-of-variance table.
(3) Tables and plots of the predicted and observed values for each curve.

c. Limitations per problem:

(1) k, number of values of the independent variable ($4 \leq k \leq 999$)

(2) m_i, number of values of the dependent variable corresponding to the i^{th} value of the independent variable ($1 \leq m_i \leq 999$)

(3) N, total number of values of the dependent variable

$$(k \leq N \leq 5000) \quad \left(N = \sum_{i=1}^{k} m_i\right)$$

(4) f, number of regression fits per problem ($1 \leq f \leq 99$)

(5) t, number of Variable Format Cards ($1 \leq t \leq 10$)

[1] Codes 00, 01, ..., 10 and 17 of the transgeneration list may be used independently or in combination as shown here.

d. Codes 00-10, 17 of the transgeneration list may be used.

e. The program will sort the data in ascending order of the values of the independent variable. Unless presorting has been done, it is necessary that sorting be specified on the Problem Card.

f. If the values of the independent variable advance in equal steps, the program can rescale these values to 0, 1, 2, ..., k-1. If the values of the independent variable do not advance in equal steps, they will be rescaled from 0 to k-1 according to their relative values.

   Rescaling is highly recommended for computational accuracy when the maximum value of the independent variable is greater than 2K.

SECTION V

CLASS S — SPECIAL PROGRAMS

BMD02S — CONTINGENCY TABLE ANALYSIS

General Description

a. This program computes two-way frequency and percentage tables, chi-squares, contingency coefficients, and maximum likelihood ratios. Each variable may be categorized in several different ways; each way is referred to as a categorization.

b. Output for this program includes the following for the original tables and for the collapsed tables:

(1) Count of rejects

(2) Frequency tables

(3) Row, column, and/or table percentages (optional)

(4) Chi-square and degrees of freedom

(5) Contingency coefficient

(6) The quantity $-2 \log \lambda$, where λ is the maximum likelihood ratio.

c. Limitations per problem:

(1) k, number of intervals in each categorization ($k \leq 22$)

(2) g_i, number of different categorizations for the i^{th} variable ($g_i \leq 10$)

(3) s, total number of different categorizations formed from all variables ($s < 700$) Included in the limitation of 700 are those variables for which no Interval Cards are prepared, i.e., those for which the program assumes boundaries with unit intervals. (See Section 3-c of the UCLA BMD reference manual.)

(4) v, number of Variable Format Cards ($1 \leq v \leq 9$)

(5) The maximum frequency for a cell in the tables is 9,999. Any frequency greater than this number is set equal to 9,999.

(6) q, the total number of intervals for all variables ($q \leq 4000$)

BMD09S — TRANSGENERATION

General Description

a. This program performs selected transgenerations on specified variables in the data. Any of the transgeneration codes may be selected. (See list of codes in Section III, Introduction, of the UCLA BMD manual.) Input to the program may be read from punched cards, or from magnetic tape in BCD code or binary code.

b. Output from this program includes:

(1) List of specified transgenerations

(2) Data transgenerated as specified

(3) List of violations of transgeneration restrictions

(4) List of all variables before and after transgeneration for the first and last cases

Printout specified:

All of the above output items appear on the printout.

Punched card output specified:

Data are transgenerated and punched on cards. All other output items on the above list appear on a printout.

Tape output, BCD or binary specified:

Data are transgenerated and written on tape. All other output items on the above list appear on a printout.

Any combination of printout, punched cards, and tape may be specified, except that both BCD and binary may not be specified.

c. Limitations per problem:

(1) n, number of cases ($1 \leq n \leq 130{,}000$)

(2) p, number of original variables ($1 \leq p \leq 999$)

(3) q, number of variables added after transgeneration ($1 \leq p+q \leq 999$) ($-998 \leq q \leq 998$)

(4) m, number of variables desired for punched card output ($0 \leq m \leq 999$)

(5) t, total number of Transgeneration Cards ($0 \leq t \leq 999$)

(6) s, number of Transgeneration Cards specifying transgeneration code 40 or 42 ($0 \leq s \leq 50$)

(7) k, number of Variable Format Cards for input ($1 \leq k \leq 10$)

(8) j, number of Variable Format Cards for printout of output ($0 \leq j \leq 10$)

(9) i, number of Variable Format Cards for punched output ($0 \leq i \leq 10$)

(10) h, number of Variable Format Cards for BCD tape output ($0 \leq h \leq 10$)

SECTION VI

CLASS T – TIME SERIES ANALYSIS

BMD01T – AMPLITUDE AND PHASE ANALYSIS

General Description

a. This program computes the amplitude and phase of moderately wide-band noise and noise contaminated by extraneous noise. The amplitude and phase are determined from a pair of finite moving averages on the sample noise.

A generalized Tukey filter with a variable number of triangles and resolutions is used. The user must specify the number of triangles and the number of resolutions, and indicate where the main lobe of the filter is to be centered.

b. Output from this program includes:

(1) The discrete values of the filter, its graph, its band width, and the leakage in percent.

(2) The coefficients of cos ($j\omega$) in $C_{m,N,\omega_0}(\omega)$ and the coefficients of sin ($j\omega$) in $S_{m,N,\omega_0}(W)$. (See Computational Procedure.)

(3) Frequency, amplitude and phase.

(4) Original data input.

c. Limitations per problem:

(1) n, number of points in the series ($50 \leq n \leq 5000$)
(2) N, number of triangles making up the filter ($1 \leq N \leq 999$)
(3) m, number of resolutions ($m \leq 4000$)
(4) l, number of points where the filter is evaluated ($50 \leq l \leq 999$)
(5) ω_0, frequency at which the filter is centered ($0 \leq \omega \leq \pi$)
(6) k, number of Variable Format Card(s) ($1 \leq k \leq 10$)

d. This program allows transgeneration of the data input. Codes 01 to 05, 07 to 10, and 17 of the transgeneration list may be used.

e. If desired, the second and following problems can use the same filter (i.e., the same coefficients of cos $j\omega$ and sin $j\omega$) as the first problem.

BMD02T – AUTOCOVARIANCE AND POWER SPECTRAL ANALYSIS

General Description

a. This program computes the autocovariance, power spectrum, cross-covariance, cross-spectrum, transfer function, and coherence function of time series.

b. Output from this program includes:

(1) Input data printed and plotted.

(2) Autocovariance printed and plotted.

(3) Power spectral estimates (power spectrum) printed and plotted.

The estimates are checked for negative powers. If any occur, they are flagged, set equal to zero for future calculations, and set equal to the maximum value of the estimates for plotting purposes.

(4) Cross-covariance of two time series printed and plotted.

(5) Cross-spectrum of two time series printed and plotted.

(6) Phase shift between two time series printed and plotted.

(7) Coherence function of two time series printed and plotted.

This function is checked to see if any values are greater than 1.1. If so, they are flagged as being erroneous due to sampling error and set equal to the maximum value of the remaining functions for plotting purposes.

(8) Transfer function of two time series printed and plotted.

If any power spectral estimate was negative or zero, the corresponding transfer function of that series and the coherence function at that lag cannot be computed. These are so indicated on the output and, for plotting purposes, are set equal to the maximum value of the remaining function values.

c. Limitations per problem:

(1) s, number of series ($1 \leq s \leq 20$)

(2) n, number of discrete data points per series ($1 \leq n \leq 1000$)

Note: In no problem may the product of n and s (n x s) be greater than 17000.

(3) m, number of lags ($1 \leq m \leq 199$)

(4) q, number of selection cards ($1 \leq q < 99$)

(5) k, number of Variable Format Cards ($1 \cdot k < 10$)

d. The program allows transformation of the input data. Codes 01 to 05, 07 to 10, and 17 of the transgeneration list may be used.

e. A simple prewhitening transformation in the form of a moving linear combination $W_t = X_{t+1} - CX_t$ is optional. C is a constant supplied by the user ($|C| < 1.0$). The final spectra are then recolored. Prewhitening cannot be used when the cross-spectrum is to be obtained.

f. The program allows detrending of the input series as an option.

SECTION VII

CLASS V – VARIANCE ANALYSIS

BMD01V – ANALYSIS OF VARIANCE FOR ONE-WAY DESIGN

General Description

a. This program computes an analysis-of-variance table for one variable of classification, with unequal group sample sizes. Optionally data may be read in from an alternate input tape, either in BCD or in binary mode. Rewinding of the tape before read-in is also optional.

b. Output for this program includes:

(1) Optional listing of the group or treatment means and standard deviations

(2) An analysis-of-variance table including:
(a) Within Groups, Between Groups, and Total Sums of Squares
(b) Within Groups, Between Groups, and Total Degrees of Freedom
(c) Within Groups and Between Groups Mean Squares
(d) F Ratio (for $H_o : \mu_1 = \mu_2 \ldots = \mu_k$).

c. Limitations per problem:

(1) k, the number of different groups or categories ($2 \leq k \leq 5000$)

(2) n_i, the number of observations or cases (sample size) for the i^{th} treatment group or category ($1 \leq n_i \leq 20000$)

(3) N, the total number of observations in all groups or categories combined,

$$N = \sum_{i=1}^{k} n_i, \quad (N \leq 100,000,000)$$

(4) m, the number of Special Transgeneration Cards ($0 \leq m \leq 9$)

d. This program allows transgeneration. Codes 01-10 of the transgeneration list may be used.

BMD02V – ANALYSIS OF VARIANCE FOR FACTORIAL DESIGN

General Description

a. This program computes an analysis of variance for a factorial design.

b. Output for this program includes:

(1) Analysis-of-variance table and the grand mean.

(2) A breakdown of the sums of squares into orthogonal polynomial components for as many as four main effects and all of their first order interactions.

(3) Main effects and first order interactions for the variables specified in (2).

(4) Cell and marginal means.

c. Limitations per problem:

(1) W, number of variables or ways ($W \leq 8$)

(2) R, number of replicates ($R \leq 999$)

(3) L_i, number of categories or levels of any one variable ($L_i \leq 999$) and ($L_1 \times L_2 \times L_3 \times \ldots \times L_w \leq 18{,}000$)

(4) k, number of Variable Format Cards ($1 \leq k \leq 5$)

d. The program can perform transgenerations of input data, if desired, according to the codes specified on one Special Transgeneration Card. Codes 01 through 10 of the transgeneration list may be used.

BMD03V – ANALYSIS OF COVARIANCE FOR FACTORIAL DESIGN

General Description

a. This program computes a full factorial analysis of covariance.

b. Output for this program includes:

(1) The total covariance matrix and a breakdown into its full factorial design components.
(2) The components in (1) adjusted by residuals.
(3) Inverses of the covariate parts of the adjusted components.
(4) Regression coefficients.
(5) t-values, an F-statistic, and the residual mean square.
(6) An analysis-of-variance table for the factorial components of the design.

c. Limitations per problem:

(1) v, number of analysis-of-variance classifications ($1 \leq v \leq 6$)

(2) p, number of covariates ($1 \leq p \leq 8$)

(3) R, number of replicates ($R \leq 999$)

(4) L_i, number of categories or levels of any one analysis-of-variance classification ($L_i \leq 999$) and ($L_1 \times L_2 \times L_3 \times \ldots \times L_v \leq 1500$)

(5) m, number of Transgeneration Cards ($m \leq 64$)

(6) q, number of covariates generated ($-7 \leq q \leq 7$) and [$0 < (p+q) \leq 8$]

(7) k, number of Variable Format Cards ($1 \leq k \leq 5$)

d. This program allows transgeneration of the input data. Codes 01-14 of the transgeneration list may be used.

BMD04V — ANALYSIS OF COVARIANCE WITH MULTIPLE COVARIATES

General Description

a. This program is designed to compute analysis-of-covariance information for one analysis-of-variance variable with multiple covariates and unequal treatment group sizes. Cases may be specified by the user as being in certain treatment groups, or cases may be placed in treatment groups by the program in accordance with a specified Boolean expression.

b. Output from this program includes:

(1) List of case numbers, data input, and group designation (optional).

(2) Variable means for each treatment group.

(3) Sums of products matrices for Total, Treatment, and Error.

(4) The inverses of the covariate matrices for Total, Treatment, and Error.

(5) Analysis-of-covariance table with degrees of freedom, sums of squares, mean squares, and F ratio.

(6) Tables of regression coefficients, their standard errors and computed t-values with and without adjustment for groups.

(7) Table of adjusted means and their standard errors.

c. Limitations per problem:

(1) v, number of variables read in. However, no more than 36 variables (including the analysis-of-variance variable) may enter the computation, i.e., $v+q \leq 36$. Also see limitations (2) and (6) below. $(3 \leq v \leq 99)$

(2) p, number of covariates $(1 \leq p \leq 35)$

(3) t, number of treatment groups $(2 \leq t \leq 99)$

(4) n_i, sample size of the i^{th} treatment group $(1 \leq n_i \leq 999)$

(5) N, total number of cases $[(t+p+1) \leq N \leq 99{,}999]$

(6) q, number of variables added to the original set after transgeneration $(-96 \leq q \leq 32)$

(7) m, number of Transgeneration Cards $(0 \leq m \leq 99)$

(8) g, number of Group Definition Cards $(0 \leq g \leq 99)$

(9) k, number of Variable Format Cards $(1 \leq k \leq 10)$

The above limitations are built into the program and are checked for each problem. If violations occur, a message to that effect is printed, and the entire job is skipped by the system.

d. This program allows transgeneration. Codes 1, 2, ..., 16 and 41 of the transgeneration list may be used.

e. One set of data input may be processed an indefinite number of times. The user may select various combinations of variables by specifying different variable formats. For each variable selection of the data input, different transgeneration options may be used since data do not remain transformed from one variable selection to the next.

f. This program allows for the placement of cases into particular treatment groups in accordance with a Boolean expression on the Group Definition Card. If the case under consideration satisfies the Boolean expression given on the Group Definition Card, it is assigned to the treatment group specified on the cards. If the case does not satisfy the Boolean expression, the case is skipped.

The Boolean expression contains the following:

Relationships	GT	(greater than)
	GE	(greater than or equal to)
	LT	(less than)
	LE	(less than or equal to)
	EQ	(equal to)
	NE	(not equal to)
Operations	A	(and)
	O	(or)

For example,

6 21GT3.000A05EQ1.000O07LE68.50

states that a case is in treatment group 6 if variable 21 is greater than 3.000 and variable 05 equals 1.000 or variable 07 is less than or equal to 68.50.

The Boolean expression should be prepared carefully so that there is no possibility of a case satisfying the expression for more than one group.

g. Information from this program includes that usually computed from standard multiple regression analysis. By an arbitrary partitioning of the data input into at least two treatment groups and ignoring the "within" output, the program may be used to perform multiple regression analysis.

BMD05V – GENERAL LINEAR HYPOTHESIS

General Description

a. This program performs the calculations required for a general linear hypothesis model. The independent variables are of two general types:

(1) Variables used to specify the analysis-of-variance classifications.
(2) Variables used as covariates.

By use of these variables, the program can be used for balanced or unbalanced analysis-of-variance or covariance designs and missing-value problems.

b. The output of this program includes:

(1) Means and standard deviations of the dependent variable and means of the covariates.

(2) Sums of squares explained by hypotheses.

(3) Estimates of regression coefficients.

(4) Residual sums of squares.

(5) F-tests and degrees of freedom.

(6) Accuracy of coefficients.

c. Limitations per problem:

(1) p, number of variables used to specify analysis-of-variance design ($1 \leq p \leq 60$)
(2) q, number of covariates ($1 \leq p+q \leq 60$)
(3) d, number of sets of Design Cards ($1 \leq d \leq 999$)
(4) R_i, number of replicates for the i^{th} set of Design Cards ($1 \leq R_i \leq 99$)
(5) H, number of Hypothesis Cards ($1 \leq H \leq 57$)
(6) m, number of Transgeneration Cards ($0 \leq m \leq 60$)
(7) k, number of Variable Format Cards ($1 \leq k \leq 5$)

d. The program can perform transgenerations of the input data. Codes 01, 02, ..., 14 of the transgeneration list may be used.

BMD06V – GENERAL LINEAR HYPOTHESIS WITH CONTRASTS

General Description

a. This program is similar to BMD05V in that it is designed to estimate and test the statistical significance of the parameters which occur in the general linear hypothesis model. This program is more general in that it can test the statistical significance of any real valued linear function of the parameters. It is less general in that higher order simultaneous tests cannot be made. Thus this program can test the hypothesis $\beta_1 = \beta_2 = 0$.

b. Output for this program includes:

(1) Data input
(2) Cell means
(3) Sums of cross-products and inverted cross-products matrix
(4) Table of residuals
(5) Regression coefficients and their standard deviations
(6) Standard errors of estimate
(7) Coefficients, estimate, and standard error of estimate for each contrast.

c. Limitations per problem:

(1) p, number of design variables ($1 \leq p \leq 60$)
(2) q, number of covariates ($0 \leq q \leq 60$)
(3) l, total number of variables, $l = p+q$, ($1 \leq l \leq 60$)
(4) n, number of observations of the dependent variable in the experiment ($1 \leq n \cdot 9999$)
(5) m, number of Transgeneration Cards ($0 \leq m \leq 99$)
(6) c, number of Contrast Cards ($1 \leq c \leq 99$)
(7) k, number of Variable Format Cards ($1 \leq k \leq 10$).

d. The program allows transgeneration of the input data. Codes 01 through 14 of the transgeneration list may be used.

BMD07V — MULTIPLE RANGE TESTS

General Description

a. This program computes an analysis-of-variance table for one variable of classification, with unequal group sizes. The treatment group or category means are ranked in increasing order; and a multiple range test, using significant ranges input by the user, is performed on the ranked means. A maximum of nine range tests may be done on the ranked means.

b. Output for this program includes:

(1) An analysis-of-variance table including:

(a) Within Groups, Between Groups, and Total Sums of Squares
(b) Within Groups, Between Groups, and Total Degrees of Freedom
(c) Within Groups and Between Groups Mean Squares
(d) F Ratio (for $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$)

(2) A listing of the treatment group means along with their labels, sample sizes, and ranks, both in their original order and ranked order.

(3) A listing of all homogeneous subsets (subsets of means not differing significantly) for each set of Range Cards.

c. Limitations per problem:

(1) k, the number of different groups or categories ($2 \leq k \leq 100$)

(2) n_i, the number of replications or cases (sample size) for the i^{th} treatment group or category ($1 \leq n_i \leq 20000$)

(3) N, the total number of observations in all groups or categories combined. Note that

$$N = \sum_{i=1}^{k} n_i \leq 2,000,000$$

(4) m, the number of Special Transgeneration Cards ($0 \leq m \leq 9$)

(5) j, the number of sets of Range Cards ($1 \leq j \leq 9$)

d. This program allows transgeneration. Codes 1-10 of the transgeneration list may be used.

BMD08V – ANALYSIS OF VARIANCE FOR HIERARCHICAL DESIGN WITH EQUAL CELL SIZE

General Description

a. This program performs analysis of variance for any hierarchical design with equal cell sizes. This includes the nested, partially nested and partially crossed, and fully crossed designs. The model is specified by indicating the nesting relationships of the indices. (See Design Card in the BMD08V program description in the UCLA BMD Reference Manual.) Separate analyses may be performed on several dependent variables simultaneously. Each analysis of variance table includes an expected mean square in terms of the population variance components. For this calculation each index may be specified as fixed or random from a finite or infinite population.

b. The output consists of:

(1) Cell means (e.g., $x_{i.k}$) and cell deviations (e.g., $x_{i.k} - x_{i..} - x_{..k} + x_{...}$, $x_{i..} - x_{...}$, etc.) for specified terms.

(2) Analysis of variance table including source, sum of squares, degrees of freedom, mean square and expected mean square.

c. Limitations per problem:

(1) n, number of analysis of variance indices ($n \leq 10$)

(2) c, number of components in the analysis of variance table ($c \leq 100$)

(3) q, number of dependent variables

(4) m, number of means which must be stored. This may be as large as half the amount of data, depending on the design and includes at least all the means which are to be printed. Care is taken to store as few means as possible.

NOTE: $(cq + m \leq 15000)$

SECTION VIII

CLASS X – SUPPLEMENT OF GENERAL BMD PROGRAMS

CLASS D - DESCRIPTION AND TABULATION

BMDX70 t Program

This program computes t-statistics and associated probability levels for the equality of the means of two groups based on pooled and separate variance estimates.

BMDX84 Asymmetric Correlation with Missing Data

This program is designed to compute large correlation matrices or subsets of large correlation matrices from data with missing values. This program is similar to BMD03D shown on Page 2-3.

CLASS M - MULTIVARIATE ANALYSIS

BMDX74 Identification of Outliers

This program screens multivariate data for outliers by computing the Mahalanobis distance of each case from the center of the distribution of the remaining cases.

BMDX75 Canonical Analysis

This program computes canonical correlations, canonical coefficients and canonical variables. This program is similar to BMD06M shown on Page 3-6.

CLASS R - REGRESSION ANALYSIS

BMDX85 Nonlinear Least Squares

This program obtains a least squares fit

$$y = f(x_1, \ldots, x_t; \theta_t, \ldots, \theta_p) + e$$

of a user specified function f to data values $x_1, \ldots x_t$, y by means of stepwise Gauss-Newton iterations on the parameters $\theta_1, \ldots, \theta_p$.

CLASS S - SPECIAL PROGRAMS

BMDX76 Life Tables and Survival Rate

Starting with survival data for each case or a summary table, this program computes interval and cumulative survival rates together with standard errors and effective sample sizes.

BMDX77 Transgeneration

This program performs transgeneration by means of user supplied FORTRAN statements. For standard use only the arithmetic statements are required, although the entire FORTRAN language is available.

CLASS T - TIME SERIES ANALYSIS

BMDX68 Multiple Time Series Spectral Analysis

Beginning with a sequence of cross-spectral matrices, normally generated by BMDX92, this program estimates multiple coherence functions and frequency response functions between a set of input and a set of output series.

BMDX92 Time Series Spectrum Estimation

This program estimates auto-spectra, cross-spectra, and coherences for stationary time series.

CLASS V - ANALYSIS OF VARIANCE

BMDX63 Multivariate General Linear Hypothesis

This program computes likelihood ratio statistics and approximate F-statistics to test multivariate hypotheses of the form $ABC' = D$ where B is a matrix of regression coefficients and A, C, and D are matrices specified by the user.

BMDX64 General Linear Hypothesis

This is a univariate general linear hypothesis routine. Dummy variables specifying the analysis of variance part of the design are generated by the program. This program is similar to BMD05V shown on Page 7-6.

BMDX69 Multivariate Analysis of Variance and Covariance

This program will perform a Model I univariate or multivariate analysis of variance or covariance for any complete balanced design with equal cell sizes.

BMDX63 MULTIVARIATE GENERAL LINEAR HYPOTHESIS

General Description

a. This program performs a multiple regression where the dependent variable is a vector. It computes U-statistics and approximate F-statistics to test hypotheses of the form ABC' = D where B is a matrix of regression coefficients and where A, C, and D are matrices specified by the user. Estimates of $\Gamma = ABC' - D$ and the covariance matrix of its estimator are also obtained. With proper specification it can be used to carry out balanced or unbalanced multivariate analyses of variance and covariance.

b. Output from this program includes:

(1) Cross-product matrix $(X, Y)'(X, Y)$
(2) Regression coefficients, $B = (X'Y)^{-1}X'Y$ and residual cross product matrix $R = Y'Y - B'X'Y$
(3) For each hypothesis, A, C, D, ABC'-D, $A(X'X)^{-1}A'$ and CRC' matrices are printed.
(4) For each hypothesis, the hypothesis sum of products matrix, U-statistic, F-statistic, and degrees of freedom are printed.

c. Limitations per problem:

With p independent variables and q dependent variables, the following restriction must be satisfied for each hypothesis being tested.

$$(p+q)^2 + [r, q]\, p + [p, q]\, r + [r, s]\, q + qs < 9000$$

where r is the number of rows in A, s is the number of rows in C, and $[x, y]$ denotes the larger of x and y. In any case, if $(p+q)<55$, the inequality is satisfied. No transgenerations are available.

BMDX64 GENERAL LINEAR HYPOTHESIS

General Description

a. The purpose of this program is to estimate parameters and test hypotheses concerning a general linear model. Dummy variables specifying the analysis of variance part of the model are generated by the program and tested automatically. Additional hypotheses stated in the form of linear restrictions on the regression coefficients may be tested. The program does not require balanced or even full rank models. It is written in FORTRAN and uses double-precision arithmetic.

b. Output from this program includes:

(1) Regression coefficients under each hypothesis and under no hypothesis.

(2) A tolerance factor for each matrix inversion required in (a) which may be used as an indication of the accuracy of the results (see Computational Procedure).

(3) Analysis of variance table.

(4) The following printed for each cell of the design:

(i) Dummy variables generated by the program.

(ii) Predicted value of the dependent variable evaluated at the sample means of the covariates. These correspond to adjusted cell means if a full model is used.

c. Creation of dummy variables:

Dummy variables specifying the analysis of variance part of the design matrix can be created in one of two ways; totally automatically through the use of Dummy Variable Cards or semi-automatically through the use of Design Cards and Dummy Variable Cards. The use of Design Cards allows the user to specify generators (orthogonal polynomials, e.g.) from which the dummy variables, specified by the Dummy Variable Cards, are created.

d. Three methods of data input:

(1) Cell index values are read as part of each case record. When this method is used the records may be in any order.

(2) Cell index values are read from an Index Card which precedes the case records for the corresponding cell. This is especially useful when the cell index values are not readily available from the records.

(3) Design variable values for a cell are read from a Design Card which immediately precedes the case records for the corresponding cell. In this case index values are not used in the generation of dummy variables.

(4) $t + q + \max(h, 10) \leq 90$

where t = number of dummy analysis of variance variables generated
q = number of covariates
h = number of restrictions in the largest additional hypothesis

(5) $p < 10$ where p = number of analysis of variance indices

(6) Alternate input tape cannot be 1 if $TAPE is used for intermediate storage

BMDX68 MULTIPLE TIME SERIES SPECTRAL ANALYSIS

General Description

a. Beginning with a sequence of complex valued matrices of spectra and cross-spectra, normally generated by BMDX92, this program estimates multiple coherences and frequency response matrices corresponding to a set of input and output series. Confidence bands are obtained for the gain and phase of each component of the frequency response matrices. Also obtained are estimates of the spectra and cross-spectra of the output series partialed on the input series (herein called the conditional spectral matrix).

b. Output from this program (all optional) includes for each frequency:

(1) Initial matrix of spectra and cross-spectra
(2) Conditional spectral matrix
(3) Pairwise coherences
(4) Multiple coherences for the output series and tolerances for the input series
(5) Frequency response matrix
(6) Gain and phase for each component of (5)
(7) Confidence bands for each gain and phase
(8) Inverse of the cross-spectral matrix for the input series
(9) Matrix of standard errors of the frequency response coefficients

c. Limitations per problem:

(1) p, number of original series ($2 \leq p \leq 10$)

(2) n, number of frequencies (input matrices) ($1 \leq n \leq 999$)

(3) q, number of input series ($1 \leq q \leq p-1$)

(4) m, number of output series ($1 \leq m \leq p-q$)

(5) k, number of Variable Format Cards ($1 \leq k \leq 10$)

BMDX69 MULTIVARIATE ANALYSIS OF VARIANCE AND COVARIANCE

General Description

a. This program will perform Model I univariate or multivariate analysis of variance or covariance for any hierarchical design with equal cell sizes. This includes nested, partially nested and partially crossed, and fully crossed designs. The design is specified by indicating the nesting relationships of the indices. Several analyses may be performed for each problem by specifying different dependent variables or covariates on separate Subproblem Cards. Univariate analysis of variance should be performed with BMD08V.

b. Output from this program includes:

(1) Covariance matrix for each analysis of variance component
(2) Cell means for each variable
(3) For each univariate analysis of covariance:

(i) Regression coefficients under each hypothesis

(ii) Analysis of variance table including source, sum of squares, mean square, degrees of freedom, and F-statistic for each analysis of variance component of the model and each covariate

(iii) Adjusted cell means

(4) For each multivariate analysis of variance or covariance the following is tabulated for each analysis of variance component and each covariate:

(i) Generalized variance

(ii) U-statistic and degrees of freedom

(iii) Approximate F-statistic and degrees of freedom

If regression coefficients or adjusted cell means are needed in the multivariate case, each dependent variable may be run separately as a univariate problem by specifying additional Subproblem Cards.

c. Limitations per problem:

(1) $n \leq 10$

(2) $cp(p+1)/2+m \leq 15000$

(3) $c \leq 100$

where: n = number of analysis of variance indices
c = number of components in the analysis of variance part of the model
p = total number of variables
m = number of means which must be stored. This may be as large as half the amount of data, depending on the design, and includes at least the means to be printed.

BMDX70 t-PROGRAM

General Description

a. This program computes t-statistics and associated probability levels for the equality of the means of two groups based on pooled and separate variance estimates. An F-statistic and associated probability level for the equality of group variances is also computed. Groups are defined by means of a cut point for a category variable. Several dependent variables may be analyzed concurrently. Each problem may contain from one to twenty subproblems. Each subproblem is defined through Boolean selection of cases. Transgenerations are available and data specified as "missing" will be deleted for that variable. Paired comparison t-rations may be obtained through transgeneration.

b. Output from this program includes:

(1) F-ratio of variance
(2) t-value (based on pooled variance estimate)
(3) t-value (based on separate variance estimate)
(4) Two-tailed probability levels for each t and for the F
(5) Means
(6) Standard deviations
(7) Standard error of the means
(8) Number of observations included in computation of items 5 and 7

c. Limitations per problem:

(1) p, number of original variables ($1 \leq p \leq 199$)
(Note: only variables 1-100 can be analyzed; the remaining variables are available for transgeneration.)
(2) n, number of cases ($1 \leq n \leq 32000$)
(3) q, number of variables added to the original set after transgeneration ($-198 \leq q \leq 99$)
(4) p+q, total number of variables output ($1 \leq p+q \leq 100$)
(5) m, number of Transgeneration Cards ($0 \leq m \leq 100$)
(6) D, number of Missing Value Cards ($0 \leq D \leq 100$)
(7) S, number of subproblems ($1 \leq S \leq 20$)
(8) b, number of Case Selection Cards per subproblem ($1 \leq b \leq 2$)
(9) K, number of Variable Format Cards ($1 \leq K \leq 10$)
(10) t, alternate input tape cannot be equal to 1 if $TAPE is used for intermediate storage

BMDX74 IDENTIFICATION OF OUTLIERS

General Description

a. This program screens multivariate data for outliers by computing the Mahalanobis distance of each case from the center of the distribution of the remaining cases. If the probability of the F-statistic corresponding to the greatest distance is smaller than a specified value, the case involved is removed and the process is repeated until all probabilities are sufficiently large.

b. Output from this program includes:

(1) Means and standard deviations before and after deleting cases.
(2) For each case removed, the following is tabulated:

Case number
F-value
Probability
D-square
Input data

(3) Plot of the first principal component against the second with the removed cases identified.

c. Limitations per problem:

(1) $p^2 + 7p + 3n + np \leq 15000$.

(2) $n \geq p + 2$

(3) Alternate input tape $\neq 1$, if intermediate file is a $TAPE.

BMDX75 CANONICAL ANALYSIS

General Description

a. This program computes canonical correlations, canonical coefficients and canonical variables corresponding to two sets of input variables. The canonical variables may be placed on an alternate output tape for further processing. The computations are based on covariance or correlation matrices computed either about the mean or about the origin.

b. Output from this program includes:

(1) Means and standard deviations
(2) Canonical correlations
(3) Canonical coefficients
(4) Canonical variables evaluated for each case (optional)
(5) Original data (optional)

c. Limitations per problem:

(1) Total number of variables ($t \leq 99$)
(2) Total number of input or output Variable Format Cards ($f \leq 9$)

BMDX76 LIFE TABLE AND SURVIVAL RATE

General Description

a. This program provides for the single data group the proportion surviving, survival rate, standard error and effective sample size for every period; and a plot of the cumulative proportion surviving. For two independent groups it provides survival rates, standard errors, effective sample sizes, and values of t-tests by periods; and a plot of the survival rate of one group against the other on a combined plot. It also provides a combined plot of all the group survival rates in a problem and Boolean selection of cases.

b. Five forms of input acceptable to the program:

Input				
Survival Table	Form A	Form B	Form C	Form D
Period	Date of diagnosis	Date of diagnosis	Date of diagnosis	Date of diagnosis
Number alive at beginning of period	Date of death		Time from diagnosis to death	
Number who died during period	Data last observed for those alive		Time from diagnosis to last observation for those alive	
Number lost to followup during period		Date last observed		Time from diagnosis to last observation
Number withdrawn alive during period		Dead or alive status		Dead or alive status

c. Output from this program includes:

(1) number of cases alive at the beginning of the interval
(2) number of cases who died during the interval
(3) number of cases lost to observation or withdrawn alive during the interval
(4) effective number exposed to the risk of dying
(5) the proportion dying
(6) the proportion surviving
(7) the survival rate
(8) the standard error
(9) the effective sample size

d. Optional output includes:

(1) TABLE I. Listing of cases: A six digit alphanumeric ID is used to identify any individual case. If one or more dates are outside the range of the study, an error message will appear under that case.

(2) TABLE II. Survival data for single period cohorts: Cases are grouped by their period of entry and followed to the end of the study.

(3) TABLE III. Single cohort survival rates and their standard errors at date of report: For every single cohort the number of cases diagnosed during specified cohort period (number of patients dead, lost to followup and withdrawn alive at date of report); cohort survival rate and standard error of survival rate; and total number of cases diagnosed in the study (total number of patients dead, lost to followup and withdrawn alive during the study).

(4) TABLE IV. Computation for successively reduced number of periods: Combination of cohorts illustrating reduction in standard error with increase in cohort size.

e. Plot of cumulative proportion surviving: A plot of the cumulative proportion surviving against time. The unit of time is the same used for the input data.

f. Limitations per problem: (for all forms)

(1) p = number of time periods $1 \leq p \leq 50$
(2) v = number of variables $5 \leq v \leq 25$
(3) n = number of cases $1 \leq n \leq 9999$
(4) k = number of Variable Format Cards per subproblem $0 \leq k \leq 5$

BMDX77 TRANSGENERATION

General Description

a. The purpose of this program is to preprocess data intended for further processing by other BMD programs. The program performs transgeneration by means of user supplied FORTRAN statements. For standard use only the arithmetic statements are required, although the entire FORTRAN language is available. Means and standard deviations of the original variables are precomputed if the user indicates that they are used in the transgeneration. Case selection and variable selections are readily specified.

b. Output consists of the transgenerated data on the specified BCD output tape. Additional output may be obtained by print or write statements. The number of cases for each problem, up to a maximum of 16 problems, are written on tape unit 1 under a 1615 format.

BMDX84 ASYMMETRICAL CORRELATION WITH MISSING DATA

General Description

a. The program is designed to compute large correlation matrices or subsets of large correlation matrices from data with missing values. Arbitrary sets of rows and columns may be selected and the results restricted to lie between specified diagonals. The entire matrix and single rows, columns and diagonals are special cases. Blanks identify missing data. Correlations are computed using all possible complete pairs of data. An effort has been made to minimize the computing required under the missing data option.

b. Output from this program includes:

(1) Specified correlations
(2) Number of pairs of observations used to compute each correlation
(3) Means and standard deviations
(4) Number of observations used to compute each mean and standard deviation

c. Limitations per problem:

$g \leq 20000$

where

$g = 3p + 4n_c$ if missing data are not specified;

$g = 3p + 10n_c$ if missing data are specified;

p = number of variables after transgeneration;

n_c = number of columns per subproblem.

In any case, any problem with $p \leq 1500$ can be handled.

BMDX85 NONLINEAR LEAST SQUARES

General Description

a. This program obtains a weighted least squares fit

$$y = f(x_1, \ldots, x_t; \theta_1, \ldots, \theta_p) + e$$

of a user specified function f to data values $x_1, \ldots, x_t$, y by means of step-wise Gauss-Newton iterations on the parameters $\theta_1, \ldots, \theta_p$. Within each iteration parameters are selected at a given step is the one which, differentially at least, makes the greatest reduction in the error sum of squares. If necessary to avoid singularity problems, only a subset of the parameters may be modified in a given iteration. In addition, parameters with boundary values whose modification in a given iteration would lead to boundary violations are not modified. In effect the iteration is performed "on the boundary" when it appears that the best fit occurs outside the parameter range.

b. Output from this program includes:

(1) Parameter values and residual mean square after each iteration

(2) Asymptotic standard deviations and correlations for the estimated parameters after the last iteration

(3) For each case after the last iteration

(a) Original data

(b) Value of the function

(c) Residual

(d) Standard deviation of predicted y

c. Limitations per problem:

(1) Total number of data, $n(t + 1) \leq 15000$

(2) Number of parameters, $p \leq 100$

(3) Number of Variable Format Cards, $f \leq 9$

where n is the number of cases and t is the number of independent variables.

BMDX92 TIME SERIES SPECTRUM ESTIMATION

General Description

a. This program estimates auto-spectra, cross-spectra, and coherences for stationary time series. Each series is decomposed into frequency components by means of a finite Fourier transform and the required estimates are obtained by summing products of the transformed series. Linear trend is removed from each series before transformation. If desired the series may be prefiltered and decimated before detrending. Matrices of cross-spectra in complex form for input into BMDX68 may be obtained. The use of the fast transform algorithm (see UCLA manual) makes it possible to resolve a large number of frequency bands. Two-sided spectra are used throughout. This leads to estimates which are one half those obtained from programs using one-sided spectra.

b. Output from this program includes:

(1) Graph of the frequency response function of the prefilter (optional)
(2) Graph of the frequency window of the power spectrum estimates (optional)
(3) Power spectrum for each series
(4) Graphs of the power against frequency (optional)
(5) Graphs of the logarithm of the power against frequency (optional)
(6) Amplitude, phase, and coherence spectra for each pair of series
(7) Matrices of cross-spectra in complex form in format for input to BMDX68 (optional)
(8) Filtered data (optional)
(9) UCLA Brain Research Institute (BRI) binary output tape for MERGE (optional)

c. Notation and definition of terms:

Scan: one sampling from each of the series. At time t this is

$$(x_{t1}, x_{t2}, \ldots, x_{tp})$$

Prefilter: a moving linear function of the series. The result of applying a prefilter with weights a_n to the series x_t at time t would be

$$\sum_{n=-w}^{w} a_n x_{t+n}$$

d decimation ratio

If the series are to be prefiltered and decimated, every d^{th} point of the filter output is used.

P number of series to be analyzed

N' number of initial data scans

w half length of the prefilter

s' initial sampling rate (scans per unit time)

f_n' initial Nyquist frequency, $f_n' = s'/2$

N number of data scans analyzed

$$N = \begin{cases} \left[\frac{N' - (2w+1)}{d}\right] + 1 & \text{if prefiltering is done (brackets denote the integer part)} \\ N' & \text{if prefiltering is not done} \end{cases}$$

m smallest power of 2 greater than or equal to N ($m = 2^k \leq N \ 2^{k-1}$, $k = 4, 5, 6, \ldots$)

s sampling rate

$$s = \begin{cases} s'/d & \text{if prefiltering is done} \\ s' & \text{if prefiltering is not done} \end{cases}$$

f_n Nyquist frequency, the highest frequency which can be analyzed, $f_n = s/2$

b number of frequency bands resolved

b must be a power of 2 (i.e., 4, 8, etc.) and satisfy the relations $r \leq b \leq m/4$.

d. Limitations on the size of each problem:

The total amount of space, W, the program will require for any problem must be less than 17000. W is computed in the following manner. Let

q = number of Variable Format Cards.

w = half length of the prefilter if prefiltering is done; 0 if prefiltering is not done.

p_2 = p if p is even; p + 1 if p is odd.

h = the smallest power of 2 greater than or equal to 4w.

$$j = \begin{cases} 2 & \text{if graphs of both power and logarithm of power for auto-spectra are desired.} \\ 1 & \text{if graphs of either power or logarithm of power for auto-spectra are desired.} \\ 0 & \text{if neither graphs of power nor logarithm of power for auto-spectra.} \end{cases}$$

Then W is obtained through the following computations.

$$a_1 = \begin{cases} \dfrac{((2p-1)(b+1)-m)^2}{4(b+1)} + 2p(b+1) + 1 & \text{if } (qp-1)(b+1) > m \\ 2p(b+1) & \text{if } (2p-1)(b+1) \leq m \end{cases}$$

$$a_2 = \max\{a_1,\ 2w + 1 + 18q + p\}$$

$$a_3 = \max\{mp_2,\ 2h\}$$

$$a_4 = (b+1)p(p+1) + \max\{j(b+1),\ 2p\}$$

$$W = \max\{a_2 + a_3,\ a_4\}$$

For example, the program can handle problems with 14 channels and 1000 scans or 6 channels and 2000 scans. An error message is printed if the problem size limitations are exceeded.

Honeywell Bull

HONEYWELL INFORMATION SYSTEMS

BP 82, Rev. 0