

UNIVERSITY OF QUEENSLAND

Computer Centre

NEWSLETTER - SPECIAL EDITION

Date : 30 September 1971  
Authorization : Director of the Computer Centre

BMD - BIOMEDICAL COMPUTER PROGRAMS

On MONDAY 4 OCTOBER, several BMD programs will be released for general use.

1. INTRODUCTION

Many problems require extensive analysis of large amounts of recorded data. As far as possible, the data handling processes should be made automatic and rapid and the Biomedical Computer Program (BMD) Series has been written for that purpose. These programs are provided in a 'package' form that enables users to effect their desired computations with simple coded instructions. The data is usually presented in a standard form, so that a wide variety of analyses may be carried out by specifying different parameters to various programs.

A number of these statistical programs, which were developed at the Health Sciences Computing Facility, UCLA, are now available on the PDP-10. The programs are batch-oriented but can be used from a terminal if great care is taken with data formats.

2. WRITE-UPS

The original write-ups published in the BMD Manual, UCLA Press, 1968 have had to be modified slightly to conform with certain PDP-10 restrictions and to allow input from files as well as from card decks. These write-ups each rely in some measure on the introductory descriptive material in the early chapters of the BMD manual. No attempt has been made here to provide a complete write-up for each separate program. Instead the user has been referred to the appropriate section in the BMD Manual and given the necessary changes required for the program to run on the PDP-10.

3. BMD MANUAL

The BMD manual 'Biomedical Computer Programs' is edited by W.J. Dixon. It is the second in a series of University of California Publications in Automatic Computation and is published by the University of California Press. There is a copy of the manual in the Engineering Library, and copies also exist at the Medical School in the following libraries - Social and Preventive Medicine, Microbiology, and Clinical Sciences.



#### 4. RELEASE AND RUNNING OF PROGRAMS

BMD programs will be progressively converted and released, according to demand from users. Initially the following programs are available:

BMD01D	Simple Data Description
BMD02D	Correlation with Transgeneration
BMD08D	Cross Tabulation with Variable Stacking
BMD01M	Principle Component Analysis
BMD04M	Discriminant Analysis for Two Groups
BMD07M	Stepwise Discriminant Analysis
BMD02R	Stepwise Regression
BMD01V	One Way Analysis of Variance
BMD04V	Analysis of Covariance with Multiple Covariates

The programs are invoked by a single command following the .JOB card. The programs are in the \$BMD library and each has a filename derived from the last three identifying characters. For example, BMD01D has the filename D01, BMD04M is M04, BMD02R is R02, etc.

The command required to initiate BMD02R would therefore be

. \$BMD.R02

An up to date list of BMD programs available will be published from time to time in the Bulletin.

#### 5. PROGRAM CLASSIFICATION

The BMD programs are classified as Type 4 programs, which means that they are unsupported by the Computer Centre and are made available on a 'user beware' basis. (See Computer Centre Bulletin Vol 4 No 4, p 42). Although the Computer Centre is making these write-ups available, no knowledge of the workings of each program is claimed by the Computer Centre personnel.

The programs are not guaranteed to work and, in general, no personal assistance will be given if a user has difficulty with a program.

However, the test data listed in the BMD Manual has been run with each program that has been released, and the results agree to within rounding error.



## BMD01D

### SIMPLE DATA DESCRIPTION

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

For a write-up of this program, consult 'Biomedical Computer Programs' edited by W.J. Dixon, pages 42-48, with the following exceptions:

(a) Blanks

Blanks are interpreted as zeros so methods 1 and 2 for specifying exclusion of blanks are not available.

(b) The Problem Card - PROBLM

- (i) Cols 69-70 Not used.
- (ii) Cols 73-74 Output device number, 10 for output to a disk file, FOR10; otherwise, leave blank for output to the job output device.
- (iii) Cols 75-79 Input file name if data is on disk; otherwise, leave blank for input from the job input device.



CORRELATION WITH TRANSGENERATION

(Boolean Selection of Cases)

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.

For a write-up of this program, consult 'Biomedical Computer Programs' edited by W.J. Dixon, pages 49-59, with the following exceptions:

(a) Limitations per problem

- (i)  $p$ , the number of original variables ( $2 \leq p \leq 100$ )
- (ii)  $g$ , the number of variables added to the original set after transgeneration ( $-98 \leq q \leq 98$ )
- (iii)  $k$ , the number of Variable Format Cards ( $1 \leq k \leq 8$ )

(b) The Problem Card - PROBLM

- (i) Cols 13-15 Number of original variables ( $2 \leq p \leq 100$ )
- (ii) Cols 23-26  $+q$   $q$  variables added to the original set after transgeneration ( $2 \leq p + q \leq 100$ )
- (iii) Cols 33-34 Not used
- (iv) Cols 69-70 Not used
- (v) Cols 71-72 Number of Variable Format Cards ( $1 \leq k \leq 8$ )
- (vi) Cols 73-74 Output device number, 10 for output to a disk file, FOR10; otherwise, leave blank for output to the job output device.
- (vii) Cols 75-79 Input file name if data is on disk; otherwise, leave blank for input from the job input device.

(c) Plot Selection Card - PLOTSL

- (i) Cols 10-11 Number of variables to be cross-plotted with this base variable ( $\leq 10$ ).



CROSS-TABULATION WITH VARIABLE STACKING

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 fields that have a punched decimal point.

For a write-up of this program, consult 'Biomedical Computer Programs' edited by W.J. Dixon, pages 109-119, with the following exceptions:

(a) Limitations per problem

- (i) The limitation on the maximum value of each data item is  $\pm 1 \times 10^{38}$  approximately.

(b) The Problem Card - PROBLM

- |       |            |  |
|-------|------------|--|
| (i)   | Cols 23-25 | <u>Note:</u> $N(p+q) \leq 8,000$ , $p+q \leq 100$  |
| (ii)  | Cols 73-74 | Output device number, 10 for output to a disk file, FOR10; otherwise, leave blank for output to the job output device. |
| (iii) | Cols 75-79 | Input filename if data is on disk; otherwise, leave blank for input from the job input device.                         |



PRINCIPAL COMPONENT ANALYSIS

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

For a write-up of this program, consult 'Biomedical Computer Programs' edited by W.J. Dixon, pages 150-158, with the following exceptions:

(a) Limitations per problem

- (i) n, number of cases ( $3 \leq n \leq 220$ )

(b) The Problem Card - PROBLM

- (i) Cols 15-17 Number of cases ( $3 \leq n \leq 220$ )
- (ii) Cols 71-72 Number of Variable Format Card(s) ( $1 \leq k \leq 8$ )
- (iii) Cols 73-74 Output device number, 1Ø for output to a disk file, FOR1Ø; otherwise, leave blank for output to the job output device.
- (iv) Cols 75-79 Input filename if data is on disk; otherwise, leave blank for input from the job input device.



DISCRIMINANT ANALYSIS FOR TWO GROUPS

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.

For a write-up of this program, consult 'Biomedical Computer Programs' edited by W.J. Dixon, pages 185-195, with the following exceptions:

(a) The Problem Card - PROBLM

- (i) Cols 21-22 Not used.
- (ii) Cols 73-74 Output device number, 10 for output to a disk file, FOR10; otherwise, leave blank for output to the job output device.
- (iii) Cols 75-79 Input file name if data is on disk; otherwise, leave blank for input from the job input device.



BMDO4V

ANALYSIS OF COVARIANCE WITH MULTIPLE COVARIATES

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.

For a write-up of this program, consult 'Biomedical Computer Programs' edited by W.J. Dixon, pages 525-542, with the following exceptions:

(a) The Problem Card - PROBLM

- |       |            |  |
|-------|------------|--|
| (i)   | Cols 69-70 | Not used   |
| (ii)  | Cols 73-74 | Output device number, 10 for output to a disk file, FOR10; otherwise, leave blank for output to the job output device. |
| (iii) | Cols 75-79 | Input filename if data is on disk; otherwise leave blank for input from the job input device.                          |



## BMD01V

### ANALYSIS OF VARIANCE FOR ONE-WAY DESIGN

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 a file.

For a write-up of this program, consult 'Biomedical Computer Programs' edited by W.J. Dixon, pages 488-494, with the following exceptions:

(a) Limitations of the problem size

- (i)  $n_i$ , the number of observations or cases (sample size) for the  $i$ th treatment group or category ( $1 \leq n_i \leq 16,000$ )
- (ii)  $N$ , the total number of observations in all groups or categories combined,

$$N = \sum_i n_i, \quad (N \leq 80,000,000)$$

(b) The problem card - PROBLM

- (i) Cols 64-70 Ignored, all input is in ASCII
- (ii) Cols 73-74 Output device number, 10 for output to a disk file, FOR10; otherwise, leave blank for output to the job output device.
- (iii) Cols 75-70 Input file name if data is on disk; otherwise, leave blank for input from the job input device.



## BMD02R

### STEPWISE REGRESSION

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 that makes the greatest reduction in the error sum of squares. Equivalently it is the variable that has highest partial correlation with the dependent variable partialled on the variables that 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.

For a write-up of this program, consult 'Biomedical Computer Programs' edited by W.J. Dixon, pages 233-257d, with the following exceptions:

(a) The Problem Card - PROBLM

- (i) Cols 73-74      Output device number, 10 for output to a disk file, FOR10; otherwise, leave blank for output to the job output device.
- (ii) Cols 75-79      Input file name if data is on disk; otherwise, leave blank for input from the job input device.



STEPWISE DISCRIMINANT ANALYSIS

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 partialled 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.

For a write-up of this program, consult 'Biomedical Computer Programs' edited by W.J. Dixon, pages 214a-214t, with the following exceptions:

(a) The Problem Card - PROBLM

- |       |            |  |
|-------|------------|--|
| (i)   | Cols 41-42 | Not used   |
| (ii)  | Col 43     | Not used   |
| (iii) | Cols 73-74 | Output device number, 10 for output to a disk file, FOR10; otherwise, leave blank for output to the job output device. |
| (iv)  | Cols 75-79 | Input file name if data is on disk; otherwise, leave blank for input from the job input device.                        |