USAFA TR 91-19

OVERVIEW OF TWO COMMON STATISTICAL COMPUTER PACKAGES: SAS AND BMDP

MAJOR KAYE A. deRUIZ
DEPARTMENT OF MATHEMATICAL SCIENCES

DEAN OF THE FACULTY
UNITED STATES AIR FORCE ACADEMY
COLORADO SPRINGS, CO 80840

USAFA-TR-91-19

Technical Review by Major Joe Grisham
Department of Behavioral Science and Leadership
USAFA Academy, Colorado 80840

Technical Review by Captain Jim Rutledge
Department of Mathematical Sciences
USAF Academy, Colorado 80840

Editorial Review by Lt Col Donald C. Anderson
Department of English
USAF Academy, Colorado 80840

This research report entitled "Overview of Two Common Statistical Computer Packages: SAS and BMDP" is presented as a competent treatment of the subject, worthy of publication. The United States Air Force Academy vouches for the quality of the research, without necessarily endorsing the opinions and conclusions of the author.

This report has been cleared for open publication and public release by the appropriate Office of Information in accordance with AFM 190-1, AFR 12-30, and AFR 80-3. This report may have unlimited distribution.

Robert K. Morrow Jr.

ROBERT K. MORROW JR., Lt Col, USAF
Director of Research

23 Dec 91
Dated

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
December 1991

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
Overview of Two Statistical Packages: SAS and BMDP

5. FUNDING NUMBERS

6. AUTHOR(S)
Major Kaye A. de Ruiz

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Mathematical Sciences
United States Air Force Academy, Colorado 80840

8. PERFORMING ORGANIZATION REPORT NUMBER
USAFA TR 91-19

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report provides a brief introduction to two statistical computer packages, SAS and BMDP. The overall approach used by these packages to input data, perform statistical analyses, etc., is discussed. A common data set is provided and used to give examples of programming steps and the corresponding output. Examples of simple statistical computations such as descriptive statistics, linear regression, and analysis of variance, are all presented along with a reference for more advanced techniques. This paper is not meant to replace the associated manuals, but to provide a beginner with enough information to read the manuals easily.

14. SUBJECT TERMS
SAS, BMDP, Linear Regression, Analysis of Variance

15. NUMBER OF PAGES
17

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18

OVERVIEW OF TWO COMMON STATISTICAL COMPUTER PACKAGES: SAS AND BMDP

by

Major Kaye A. deRuiz, USAF

Department of Mathematical Sciences
United States Air Force Academy
Colorado Springs, Colorado 80840

TABLE OF CONTENTS

Page

I. Introduction....................1

II. The Data Set....................1

III. SAS....................3

IV. BMDP....................10

V. Program Reference Guide....................14

VI. References....................15

TABLE OF FIGURES

Figure

III-1 SAS Scatter Plot....................4

III-2 SAS Linear Regression Model 1....................5

III-3 SAS Linear Regression Model 2....................6

III-4 SAS ANOVA Without Interaction....................8

III-5 SAS ANOVA With Interaction....................9

IV-6 BMDP Histogram....................11

IV-7 BMDP Regression....................12

IV-8 BMDP Predicted Values and Residuals....................12

IV-9 BMDP ANOVA....................13

I. INTRODUCTION:

In statistics courses at USAFA, MYSTAT is used to solve problems in linear regression and analysis of variance because it works well , is easy to use, and is inexpensive. There are other statistical packages that are more commonly used for large data sets, which are usually more powerful and used in conjunction with a mainframe computer. In recent years, three statistical programs have become well-known throughout civilian institutions: SAS, BMDP, and SPSS. At the Air Force Academy, SPSS is used in social science and economics classes since it is widely used in that realm. This pamphlet introduces you to SAS and BMDP so that you can perform simple statistical computations using these packages. This information does not replace the manuals for these packages, but should enable you to use the manuals more adeptly.

II. THE DATA SET:

To compare the use of the two packages, we will employ a common data set taken from "Minitab Student Handbook" (1976 edition).

Students in an introductory statistics course participated in an in-class experiment. The students took their own pulse rate. They were then asked to flip a coin. If their coin came up heads, they were to run in place for one minute. Then everyone took their own pulse again. The pulse rates and some other data are given on the next page.

VARIABLE	DESCRIPTION
1	First pulse rate
2	Second pulse rate
3	1 = Ran in place; 2 = Did not run in place
4	1 = Smokes regularly; 2 = Does not smoke regularly
5	1 = Male; 2 = Female
6	Height in inches
7	Weight in pounds
8	Usual level of physical activity: 1 = Slight; 2 = Moderate; 3 = Heavy

VARIABLE								VARIABLE							
1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
64	88	1	2	1	66.00	140	2	68	64	2	2	1	69.50	150	3
58	70	1	2	1	72.00	145	2	82	84	2	1	1	73.00	180	2
62	76	1	1	1	73.50	160	3	64	62	2	2	1	75.00	160	3
66	78	1	1	1	73.00	190	1	58	58	2	2	1	66.00	135	3
64	80	1	2	1	69.00	155	2	54	50	2	2	1	69.00	160	2
74	84	1	2	1	73.00	165	1	70	62	2	1	1	66.00	130	2
84	84	1	2	1	72.00	150	3	62	68	2	1	1	73.00	155	2
68	72	1	2	1	74.00	190	2	48	54	2	1	1	68.00	150	0
62	75	1	2	1	72.00	195	2	76	76	2	2	1	74.00	148	3
76	118	1	2	1	71.00	138	2	88	84	2	2	1	73.50	155	2
90	94	1	1	1	74.00	160	1	70	70	2	2	1	70.00	150	2
80	96	1	2	1	72.00	155	2	90	88	2	1	1	67.00	140	2
92	84	1	1	1	70.00	153	2	78	76	2	2	1	72.00	180	3
68	76	1	2	1	67.00	145	2	70	66	2	1	1	75.00	190	2
60	76	1	2	1	71.00	170	3	90	90	2	2	1	68.00	145	1
62	58	1	2	1	72.00	175	3	92	94	2	1	1	69.00	150	2
66	82	1	1	1	69.00	175	2	60	70	2	1	1	71.50	164	2
70	72	1	1	1	73.00	170	3	72	70	2	2	1	71.00	140	2
68	76	1	1	1	74.00	180	2	68	68	2	2	1	72.00	142	3
72	80	1	2	1	66.00	135	3	84	84	2	2	1	69.00	136	2
70	106	1	2	1	71.00	170	2	74	76	2	2	1	67.00	123	2
74	76	1	2	1	70.00	157	2	68	66	2	2	1	68.00	155	2
66	102	1	2	1	70.00	130	2	84	84	2	2	2	66.00	130	2
70	94	1	1	1	75.00	185	2	61	70	2	2	2	65.50	120	2
96	140	1	2	2	61.00	140	2	64	60	2	2	2	66.00	130	3
62	100	1	2	2	66.00	120	2	94	92	2	1	2	62.00	131	2
76	104	1	1	2	68.00	130	2	60	66	2	2	2	62.00	120	2
82	100	1	2	2	68.00	138	2	72	70	2	2	2	63.00	118	2
100	115	1	1	2	63.00	121	2	58	56	2	2	2	67.00	125	2
68	112	1	2	2	70.00	125	2	88	74	2	1	2	65.00	135	2
96	116	1	2	2	68.00	116	2	66	72	2	2	2	66.00	125	2
78	118	1	2	2	69.00	145	2	84	80	2	2	2	65.00	118	1
88	110	1	1	2	69.00	150	2	62	66	2	2	2	65.00	122	3
62	98	1	1	2	62.75	112	2	66	76	2	2	2	65.00	115	2
80	128	1	2	2	68.00	125	2	80	74	2	2	2	64.00	102	2
62	62	2	2	1	74.00	190	1	78	78	2	2	2	67.00	115	2
60	62	2	2	1	71.00	155	2	68	68	2	2	2	69.00	150	2
72	74	2	1	1	69.00	170	2	72	68	2	2	2	68.00	110	2
62	66	2	2	1	70.00	155	2	82	80	2	2	2	63.00	116	1
76	76	2	2	1	72.00	215	3	76	76	2	1	2	62.00	108	3
68	66	2	1	1	67.00	150	2	87	84	2	2	2	63.00	95	3
54	56	2	1	1	69.00	145	2	90	92	2	1	2	64.00	125	1
74	70	2	2	1	73.00	155	3	78	80	2	2	2	68.00	133	1
74	74	2	2	1	73.00	155	2	68	68	2	2	2	62.00	110	2
68	68	2	2	1	71.00	150	3	86	84	2	2	2	67.00	150	3
72	74	2	1	1	68.00	155	3	76	76	2	2	2	61.75	108	2

III. SAS:

The SAS computer program was originally developed for a mainframe computer. More recently, the package has been adjusted to operate on personal computers as well. The SAS program has two main parts--data and procedural. The data portion is a collection of data with variable names, while the procedural (PROC) steps are used to analyze the data. A simple SAS job might look like the following:

```
DATA PULSE;

     INPUT PULSE1   PULSE2   HEIGHT   WEIGHT;
     CARDS;

62   76   73.5   160
64   84   73     165

: (continue typing in all data to be used)

76   76   26   108

;

PROC   PLOT   DATA  = PULSE;

    PLOT   PULSE1 * HEIGHT;
```

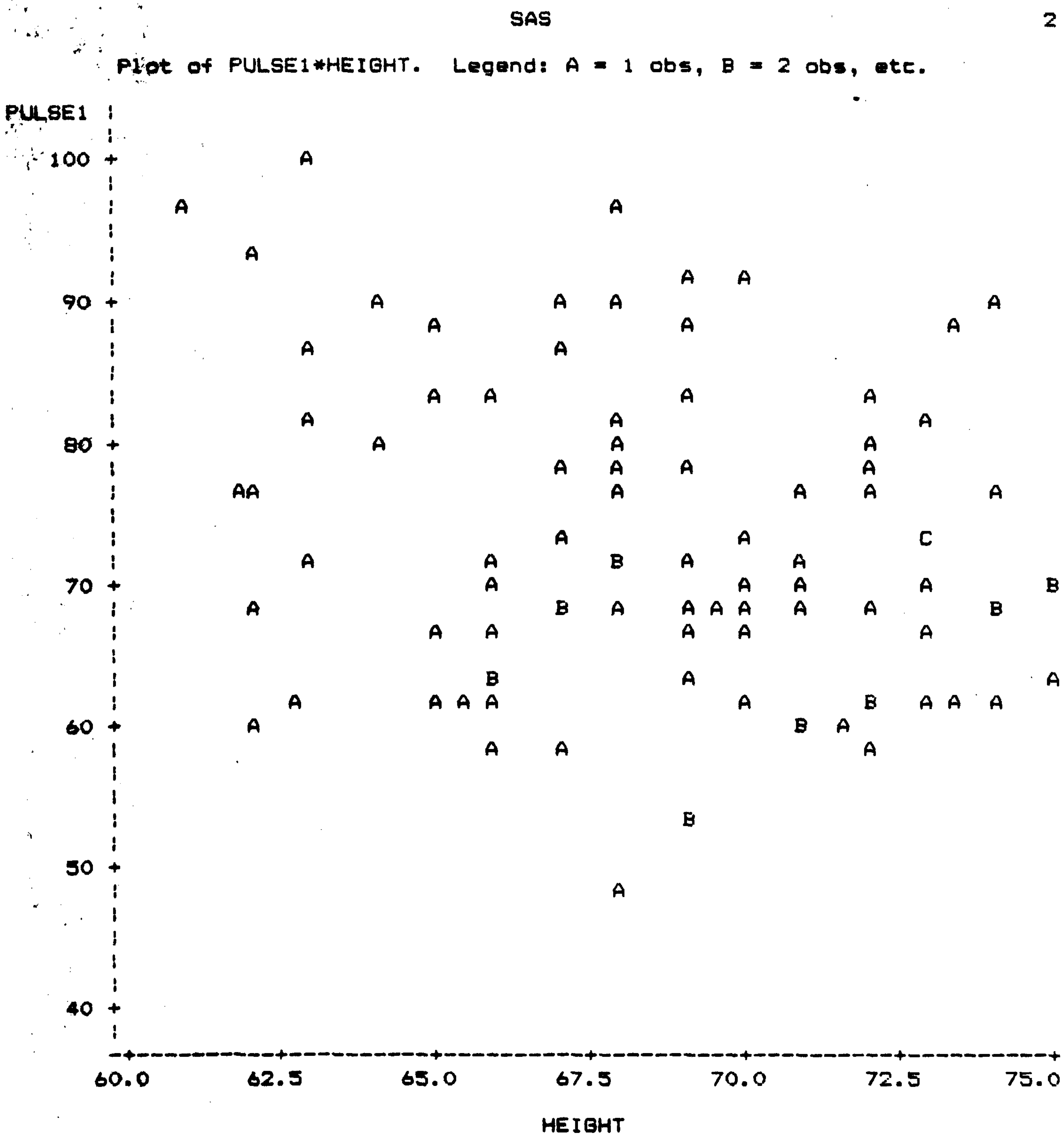
The above steps would lead to the plotting of the data points using PULSE1 and the corresponding height. PULSE was chosen as the name of the data set we are using since, with many data sets, we need to identify in PROC precisely which set we want to analyze. Note that every SAS statement (excluding the data lines) ends with a semicolon. Indentation is not required in SAS, but it makes it easier to find the DATA and PROC steps.

The CARDS statement tells SAS that the data lines follow. There are other ways to input data; for example, large data sets may be read directly from ASCI computer files. However, since inputting data directly is the most common method, we will not discuss alternative methods here. Other methods are discussed in the reference manuals accompanying SAS.

The PROC statement specifies which procedure and data set you will use. The range of analyses that can be accomplished on SAS are very extensive. PROC statements are used to perform simple analysis of variance (ANOVA) and linear regression as well as more complicated analyses such as factor analysis, discriminant analysis, and stepwise regression.

Figure 1 on the next page shows the results of the plot requested in the above program.

Figure III-1 SAS Scatter Plot



The SAS Statistics Manual discusses all of the procedures available and gives many examples of how to use each type of procedure. Many of the terms used, such as "PRESS residuals," may not be familiar to you, and there are certain tests we cannot cover in this course. However, if you continue in statistics or in an area that utilizes statistics, you will find many of the methods quite useful.

By using the pulse data in SAS, the following steps would lead to linear regression. The output is also included in Figures 2 and 3 on the next page.

```
DATA PULSE;

INPUT  PULSE1  RAN  WEIGHT  ACTIVITY;

CARDS;
(Data lines)

;

PROC  REG  DATA = PULSE

     MODEL1   PULSE1 = WEIGHT ACTIVITY;

     MODEL2   PULSE1 = RAN WEIGHT ACTIVITY;
```

The MODEL statement in PROC REG contains the name of the dependent variables you want to include in the model. (See Figures 2 and 3.)

Figure III-2 SAS Linear Regression Model 1

odel: MODEL1
ependent Variable: PULSE1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	604.46914	302.23457	2.585	0.0811
Error	89	10407.40043	116.93708		
C Total	91	11011.86957			

Root MSE	10.81375	R-square	0.0549
Dep Mean	72.84783	Adj R-sq	0.0337
C.V.	14.84429		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	87.875383	7.13929274	12.309	0.0001
WEIGHT	1	-0.093625	0.04775285	-1.961	0.0531
ACTIVITY	1	-0.615189	0.52732534	-1.167	0.2465

Figure III-3 SAS Linear Regression Model 2

Model: MODEL2
Dependent Variable: PULSE1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	741.07539	247.02513	2.117	0.1038
Error	88	10270.79418	116.71357		
C Total	91	11011.86957			

Root MSE	10.80341	R-square	0.0673
Dep Mean	72.84783	Adj R-sq	0.0355
C.V.	14.83010		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	93.926983	9.06427323	10.362	0.0001
RAN	1	-2.590996	2.39492590	-1.082	0.2823
WEIGHT	1	-0.105279	0.04890820	-2.153	0.0341
ACTIVITY	1	-0.685248	0.53078632	-1.291	0.2001

To perform Analysis of Variance (ANOVA), you use PROC ANOVA. You can do a simple one-way ANOVA or an n-way analysis of variance for balanced data. The program steps for the one-way ANOVA follow:

```
DATA PULSE;

     INPUT   PULSE2   RAN   SMOKES;

     CARDS;
     (Data lines)

     ;

PROC ANOVA DATA = PULSE;

     CLASS   RAN;

     MODEL   PULSE2 = RAN;
```

For an ANOVA involving more than one factor, list all classification factors in the CLASS statement. You then use the MODEL statement to list the terms you want in the model. Interaction terms are given by listing the variables with an * between them (i.e., SMOKES*RAN).

Steps for two-way ANOVA with and without interaction terms are given below, with Figures 4 and 5 on the following pages containing the results of these examples.

EXAMPLES:

I. Two-way ANOVA without interaction:

```
DATA   PULSE;
   INPUT   PULSE1   PULSE2   RAN   SMOKES;
   CARDS;
   (Data lines)
;
PROC   ANOVA   DATA = PULSE;
   CLASS   RAN   SMOKES;
   MODEL   PULSE2 = RAN   SMOKES;
```

II. Two-way ANOVA with interaction:

```
DATA   PULSE;
   INPUT   PULSE1   PULSE2   RAN   SMOKES;
   CARDS;
   (Data lines)
;
PROC   ANOVA   DATA = PULSE;
   CLASS   SMOKES   RAN;
   MODEL   PULSE2 = SMOKES   RAN   SMOKES*RAN;
```

Analysis of Variance Procedure
Class Level Information

Class	Levels	Values
RAN	2	1 2
SMOKES	2	1 2

Number of observations in data set = 92

Analysis of Variance Procedure

Dependent Variable: PULSE2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	8759.67437	4379.83718	22.06	0.0001
Error	89	17673.93433	198.58353		
Corrected Total	91	26433.60870			

R-Square	C.V.	Root MSE	PULSE2 Mean
0.331384	17.62933	14.0920	79.934783

Source	DF	Anova SS	Mean Square	F Value	Pr > F
RAN	1	8697.40719	8697.40719	43.80	0.0001
SMOKES	1	62.26718	62.26718	0.31	0.5769

Figure III-4 SAS ANOVA Without Interaction

Analysis of Variance Procedure
Class Level Information

Class	Levels	Values
RAN	2	1 2
SMOKES	2	1 2

Number of observations in data set = 92

Analysis of Variance Procedure

Dependent Variable: PULSE2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	8871.71262	2957.23754	14.82	0.0001
Error	88	17561.89608	199.56700		
Corrected Total	91	26433.60870			

	R-Square	C.V.	Root MSE	PULSE2 Mean
	0.335622	17.67293	14.1268	79.934783

Source	DF	Anova SS	Mean Square	F Value	Pr > F
RAN	1	8697.40719	8697.40719	43.58	0.0001
SMOKES	1	62.26718	62.26718	0.31	0.5779
RAN*SMOKES	1	112.03825	112.03825	0.56	0.4557

Figure III-5 SAS ANOVA With Interaction

IV. BMDP:

This statistical computer package is popular in social sciences and education. It can be quite easy and convenient to use for complex statistical analyses. BMDP is different from MYSTAT, SAS, or SPSS in that it is a collection of programs that each perform a different kind of analysis. SAS and MYSTAT, as you have seen, are single programs that perform many kinds of analyses. When using BMDP, you must know which programs you want to use.

A BMDP program consists of paragraphs. Sentences in the program end with a period, while paragraphs begin with a slash. The paragraphs that describe the data are common to all of the BMDP programs, but the paragraphs that describe the anlysis to be done are different for each program. A simple BMDP program using PD5 follows:

```
/INPUT    VARIABLES = 4.
          FORMAT = FREE.
/VARIABLES    NAMES = PULSE1, PULSE2, HEIGHT, WEIGHT.
/PLOT    TYPE = HIST.
          VARIABLE = PULSE1.
/END.
(The data is entered here.)
```

This program has four paragraphs. The INPUT paragraph contains a VARIABLES and FORMAT statement to identify how many variables will be read and in what format they will be read. The NAMES sentence assigns names to the variables. The PLOT paragraph describes the analysis we want performed and then identifies the variable to use in the analysis. The last paragraph signifies the end of the program which is then followed by the data lines. See Figure 6 on the next two pages for the results of the BMDP analysis. The BMDP output always includes the program statements prior to any analysis while SAS output does not provide this without additional commands.

Regression programs in BMDP all end with the letter R; the program to do standard multiple regression is BMDP1R. The following is an example of the BMDR1R program. (See Figure 7 for the output.)

```
/INPUT    VARIABLES = 3.
          FORMAT = FREE.
/VARIABLE NAMES = SEX, HEIGHT, WEIGHT.

/REGRESS    DEPENDENT = WEIGHT.
            INDEPENDENT = HEIGHT, SEX.

/END
 (Data lines)
```

```
BMDP5D - HISTOGRAM AND UNIVARIATE PLOTS
BMDP STATISTICAL SOFTWARE, INC.
1964 WESTWOOD BLVD. SUITE 202
(213) 475-5700
PROGRAM REVISED OCTOBER 1983
MANUAL REVISED -- 1983
COPYRIGHT (C) 1983 REGENTS OF UNIVERSITY OF CALIFORNIA

TO SEE REMARKS AND A SUMMARY OF NEW FEATURES FOR
THIS PROGRAM, STATE NEWS. IN THE PRINT PARAGRAPH.

     MAY 23, 1984  AT 10:29:07

PROGRAM CONTROL INFORMATION

/INPUT VARIABLES=4.
      FORMAT=FREE.
/VARIABLE NAMES=PULSE1,PULSE2,HEIGHT,WEIGHT.
/PLOT TYPE=HIST.
     VARIABLE=PULSE1.
/END
```

```
NUMBER OF VARIABLES TO READ IN. . . . . . . . . .         4
NUMBER OF VARIABLES ADDED BY TRANSFORMATIONS. .           0
TOTAL NUMBER OF VARIABLES . . . . . . . . . . .           4
NUMBER OF CASES TO READ IN. . . . . . . . . . . .    TO END
CASE LABELING VARIABLES . . . . . . . . . . . .
MISSING VALUES CHECKED BEFORE OR AFTER TRANS. .     NEITHER
BLANKS ARE. . . . . . . . . . . . . . . . . . . .   MISSING
INPUT UNIT NUMBER . . . . . . . . . . . . . . .           5
REWIND INPUT UNIT PRIOR TO READING. . DATA. . .          NO
NUMBER OF WORDS OF DYNAMIC STORAGE. . . . . . .       60414

VARIABLES TO BE USED
     1 PULSE1       2 PULSE2       3 HEIGHT       4 WEIGHT

INPUT FORMAT IS
FREE

MAXIMUM LENGTH DATA RECORD IS   80 CHARACTERS.

 GROUPING VARIABLE . . . . . . . . . . . . . . .   (NONE GIVEN)

NUMBER OF CASES READ. . . . . . . . . . . . . .    12
```

TABLE OF CONTENTS

VARIABLE NO. NAME	GROUP NAME	PLOT TYPE	PAGE NO.
1 PULSE1		HIST	4


```
HISTOGRAM OF VARIABLE     1 PULSE1
                                  SYMBOL  COUNT      MEAN     ST.DEV.
                                    X       12      72.750     10.101
                         EACH SYMBOL REPRESENTS        1 OBSERVATIONS
INTERVAL                                                                      FREQUENCY PERCENTAGE
NAME          5   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80  INT. CUM. INT.  CUM.
         *********************************************************************************
*60.0300 *                                                                                  0    0   0.0   0.0
*60.9000 *                                                                                  0    0   0.0   0.0
*61.7700 *X                                                                                 1    1   8.3   8.3
*62.6400 *XXX                                                                               3    4  25.0  33.3
*63.5100 *                                                                                  0    4   0.0  33.3
*64.3800 *                                                                                  0    4   0.0  33.3
*65.2500 *                                                                                  0    4   0.0  33.3
*66.1200 *                                                                                  0    4   0.0  33.3
*66.9900 *                                                                                  0    4   0.0  33.3
*67.8600 *                                                                                  0    4   0.0  33.3
*68.7300 *X                                                                                 1    5   8.3  41.7
*69.6000 *                                                                                  0    5   0.0  41.7
*70.4700 *                                                                                  0    5   0.0  41.7
*71.3400 *                                                                                  0    5   0.0  41.7
*72.2100 *X                                                                                 1    6   8.3  50.0
*73.0800 *                                                                                  0    6   0.0  50.0
*73.9500 *                                                                                  0    6   0.0  50.0
*74.8200 *X                                                                                 1    7   8.3  58.3
*75.6900 *                                                                                  0    7   0.0  58.3
*76.5600 *X                                                                                 1    8   8.3  66.7
*77.4300 *                                                                                  0    8   0.0  66.7
*78.3000 *                                                                                  0    8   0.0  66.7
*79.1700 *                                                                                  0    8   0.0  66.7
*80.0400 *XX                                                                                2   10  16.7  83.3
*80.9100 *                                                                                  0   10   0.0  83.3
*81.7800 *                                                                                  0   10   0.0  83.3
*82.6500 *                                                                                  0   10   0.0  83.3
*83.5200 *                                                                                  0   10   0.0  83.3
*84.3900 *X                                                                                 1   11   8.3  91.7
*85.2600 *                                                                                  0   11   0.0  91.7
*86.1300 *                                                                                  0   11   0.0  91.7
*87.0000 *                                                                                  0   11   0.0  91.7
*87.8700 *                                                                                  0   11   0.0  91.7
*88.7400 *                                                                                  0   11   0.0  91.7
*89.6100 *                                                                                  0   11   0.0  91.7
*90.4800 *                                                                                  0   11   0.0  91.7
*91.3500 *                                                                                  0   11   0.0  91.7
*92.2200 *X                                                                                 1   12   8.3 100.0
         *********************************************************************************
              5   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80

NUMBER OF INTEGER WORDS OF STORAGE USED IN PRECEDING   PROBLEM     606
CPU TIME USED       0.257 SECONDS
```

Figure IV-6 BMDP Histogram

In addition, this program will also print the predicted values and residuals, along with the variables in the data set. To do this, insert /PRINT DATA. between the REGRESS and END paragraphs. The output is shown below in Figure 8.

```
PAGE   2  BMDP1R

  REGRESSION TITLE IS
     JUNE  6, 1984       14:37:01

  DEPENDENT VARIABLE. . . . . . . . . . . . . . . .       3 WEIGHT
  TOLERANCE . . . . . . . . . . . . . . . . . . . . . 0.0100
ALL DATA CONSIDERED AS A SINGLE GROUP

MULTIPLE R              0.9571          STD. ERROR OF EST.        8.1313
MULTIPLE R-SQUARE       0.9160

ANALYSIS OF VARIANCE
                     SUM OF SQUARES     DF     MEAN SQUARE       F RATIO    P(TAIL)
        REGRESSION      6485.8437        2      3242.9219        49.048     0.0000
        RESIDUAL         595.0596        9        66.1177

                                               STD. REG
   VARIABLE        COEFFICIENT  STD. ERROR      COEFF         T    P(2 TAIL) TOLERANCE

INTERCEPT            -41.11378
SEX          1       -24.09193     7.88865     -0.489    -3.054    0.0137      0.36427
HEIGHT       2         3.18480     0.97951      0.521     3.251    0.0100      0.36427

NUMBER OF INTEGER WORDS OF STORAGE USED IN PRECEDING    PROBLEM     614
CPU TIME USED       0.140 SECONDS
```

Figure IV-7   BMDP Regression

```
LIST OF PREDICTED VALUES, RESIDUALS, AND VARIABLES
         NOTE - NEGATIVE CASE NUMBER DENOTES A CASE WITH MISSING VALUES.
                THE NUMBER OF STANDARD DEVIATIONS FROM THE MEAN IS DENOTED BY UP TO 3 ASTERISKS TO THE RIGHT
                     OF EACH RESIDUAL OR VARIABLE.
MISSING VALUES AND VALUES OUT OF RANGE ARE DENOTED BY VALUES
GREATER THAN OR EQUAL TO  0.2127E+38 IN ABSOLUTE VALUE.

  CASE          RESIDUAL        PREDICTED       VARIABLES
  LABEL   NO.                     VALUE           1 SEX          2 HEIGHT        3 WEIGHT


           1  -0.877    *      160.9          1.000          73.50    *     160.0
           2  -2.285           167.3          1.000          73.00    *     165.0
           3  -9.100    *      164.1          1.000          72.00          155.0
           4  -4.731           157.7          1.000          70.00          153.0
           5   1.451           110.5          2.000    *     62.75    *     112.0     *
           6  -2.269           127.3          2.000    *     68.00          125.0
           7   19.53    **     170.5          1.000          74.00    *     190.0     *
           8   1.824           148.2          1.000          67.00          150.0
           9   3.639           151.4          1.000          68.00          155.0
          10  0.6929           119.3          2.000    *     65.50          120.0
          11  0.2853           117.7          2.000    *     65.00          118.0
          12 -0.1602           108.2          2.000    *     62.00    *     108.0     *
```

Figure IV-8   BMDP Predicted Values and Residuals

To perform a one-way analysis of variance, program BMDP1V is used. This program requires an additional DESIGN paragraph with a DEPENDENT sentence. An additional GROUPING sentence identifies the classification variable, placed in the VARIABLE paragraph. The BMDP1V program and the output is shown below.

```
/INPUT       VARIABLES = 3.
             FORMAT = FREE.
/VARIABLE    NAMES = PULSE2, RAN, SMOKES.
             GROUPING = RAN.
/DESIGN      DEPENDENT = PULSE2.
/END
 (Data lines)
```

*** ANALYSIS OF VARIANCE ***

PULSE2
BY RAN

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF OF F
MAIN EFFECTS	1587.000	1	1587.000	9.263	0.017
RAN	1587.000	1	1587.000	8.263	0.017
EXPLAINED	1587.000	1	1587.000	8.263	0.017
RESIDUAL	1920.667	10	192.067		
TOTAL	3507.667	11	318.879		

Figure IV-9 BMDP ANOVA

The program BMDP2V can be used for three-way and higher interactions. You can include or exclude interactions and main effects in the DESIGN paragraph as shown below. The numbers in these statements indicate the order the factors were presented in the grouping statement. For example, RAN corresponds to 1, SMOKES corresponds to 2, and 23 represents the interaction between SMOKES and SEX.

```
/DESIGN    DEPENDENT = PULSE2.
           GROUPING = RAN, SMOKES, SEX.
           EXCLUDE = 23, 123.
           (Excludes the interaction between SMOKES and SEX and
           between all three independent variables.)
/DESIGN    DEPENDENT = PULSE2.
           GROUPING = RAN, SMOKES, SEX.
           INCLUDE = 1, 2, 3, 12, 13.
           (Includes all three independent variables and the
           interactions between RAN and SMOKES and between RAN and
           SEX.)
           (Data lines)
```

V. PROGRAM REFERENCE GUIDE:

Following is a partial list of capabilities and the procedure or program name used to access the package.

CAPABILITY	SAS	BMPD
Histograms	CHART	5D
PLOTS	PLOT	6D
DATA SCREENING		4D, AM
Descriptive Statistics	TTEST, MEANS, UNIVARIATE, SUMMARY, CORR	1D, 2D, 3D, 8D, 9D
Regression		
Multiple Linear	REG	1R
Stepwise	STEPWISE	2R
All possible subsets	RSQUARE	9R
Nonlinear	NLIN	3R, AR
Logistics	LOGISTIC, CATMOD	LR
Other	RSREG	4R, 5R, 6R
Analysis of Variance		
Balanced	ANOVA	1V, 2V, 8V
Unbalanced	GLM	1V, 2V, 3V
ANCOVA	GLM	1V
Repeated Measures	GLM	2V
Other	Nested VARCOMP	
Models for Categorical Data		
Contingency Tables	FUNCAT	
Log-linear Models	CATMOD	4F
Nonparametrics	NPAR1WAY	
Multivariate Techniques		
Canonical Correlation	CANCORR	6M
Discriminant Analysis	DISCRIM NEIGHBOR CANDISC STEPDISC	7M
Principal Components	PRINCOMP	
Factor Analysis	FACTOR	4M, 8M
Cluster Analysis	CLUSTER FASTCLUS VARCLUS, TREE	1M, 2M, 3M, KM
MANOVA	GLM	4V

VI. REFERENCES:

Lefkowitz, Jerry M., Introduction to Statistical Computer Packages, Duxbury Press, 1985.

Afifi, A.A. and Clark V., Computer Aided-Multivariate Analysis, Van Nostrand Reinhold, 1984.