TYMSHARE MANUALS
TYMCOM-X

STATPAK

FEBRUARY 1972

TYMSHARE, INC.
CUPERTINO, CALIFORNIA 95014

CONTENTS

SECTION 1

INTRODUCTION

Tymshare's statistical package for use on the TYMCOM-X computer is called STATPAK. The major uses of STATPAK are in business and financial applications such as management science, operations research, market analysis, financial analysis, and investment analysis. STATPAK is also a convenient tool for scientists and engineers involved in statistical work. Tasks that can be performed by STATPAK include:

- Data creation and editing.
- Data screening and plotting.
- Data transformations, including square root, natural and common logarithms, exponential, inverse, etc.
- Parametric and non-parametric hypothesis testing.
- Regression analysis.
- Time series analysis.
- Discriminant analysis.
- Analysis of variance and factor analysis.

A useful feature of STATPAK is the ability to store on a file results from a STATPAK analysis for use by a later STATPAK analysis or by a user-written program.

STATPAK contains eight modules and eighteen programs to perform statistical analyses. These are listed on the following page.

Section 2 of this manual describes the general use of STATPAK, including the creation of data files. Sections 3 through 10 describe each module and program with a sample execution of each. The Appendix contains a list of all STATPAK error messages and their explanations.

In all examples throughout this manual, everything typed by the user is underlined. User-typed Carriage Returns are represented by the symbol ↵.

Control characters are denoted in this manual by a superscript c. For example, A^c denotes Control A. The method of typing a control character depends upon the type of terminal being used. Consult the literature for your particular terminal.

Module Name	Program Name	Description
CREME	CRATE	Creates a data file in the proper format for all STATPAK programs except ANVAR.
	MODIF	Allows the user to modify a data file.
	MERGE	Allows the user to select data from two or more files and to merge this data on a new data file.
	TRANS	Performs 12 data transformations.
DATA	DSCRE	Generates basic statistics, performs conditional and unconditional analyses, and plots a histogram.
REGON	DREGR	Performs a multiple linear regression analysis.
	RGSTP	Performs a stepwise multiple linear regression analysis.
	RGPOL	Performs a polynomial regression analysis.
PARH	TSTAT	Computes the student's *t*-statistic under four different hypotheses.
	FSTAT	Computes the *F*-ratio and corresponding degrees of freedom.
NPARH	UTEST	Tests the hypothesis that two independent samples come from the same population.
	CRANK	Computes Kendall's rank correlation coefficient between two variables.
	CCORD	Computes the degree of association among several variables (the concordance coefficient) and the chi-square statistic.
TIMSA	TRIXP	Makes a forecast of a variable based on a set of past observations.
	XPOSE	Makes a forecast of a variable using exponentially weighted moving averages; includes linear trend and seasonal factors.
DISCA	DANSS	Performs a discriminant analysis and generates linear discriminant functions.
AVANC	ANVAR	Performs an analysis of variance.
	FCTOR	Performs a factor analysis.

SECTION 2

USING STATPAK

STATPAK is called from XEXEC[1] by typing:

```
-R STATPAK↵
```

STATPAK then requests the name of the module desired by the user. For example,

```
-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)?
```

The user may respond with the name of any one of the eight STATPAK modules, or he may type HELP and a Carriage Return for a list of the modules and the programs within each.

After the module is specified, STATPAK requests the name of the program to be called from that module. For example,

```
-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? CREME↵

PROGRAM NAME (TYPE "HELP" FOR AID)?
```

The user may respond to the last question with the name of a program in CREME, or he may type HELP for a list of all programs and modules.

STATPAK asks a series of questions during each analysis. Each response must be terminated by a Carriage Return. If the response consists of two or more numbers, those numbers must be separated by spaces.

When a module has run to completion, STATPAK requests the next module name. The user may type STOP to return to XEXEC.

1 - Refer to the *Tymshare TYMCOM-X XEXEC Reference Manual* for a description of XEXEC.

Error Correction

While executing STATPAK programs, the user can correct his entries by using Control A or Control Q. If he types an incorrect character in STATPAK, he may delete it with a Control A. The first time Control A is typed, a back slash (\) is printed, followed by a reprint of the deleted character. This key may be used repeatedly to delete successive characters to the left; each time Control A is typed, the next deleted character is reprinted. When all incorrect characters have been deleted and normal typing resumes, another back slash and the first new correct character are printed. For example,

ABXCPA^C\PA^CCA^CXC\D CORP↵

is interpreted as

ABCD CORP

Note that the user types the C before the second back slash is printed; however, the C appears on the right side of the back slash in the actual printout.

Control Q deletes an entire response. On some terminals, Control Q echoes as an up arrow. For example,

```
-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)



MODULE NAME (TYPE "HELP" FOR AID)? CMRE↑
CREME↵

PROGRAM NAME (TYPE "HELP" FOR AID)?
```

The user types CMRE and a Control Q. The carriage is returned by the system, and the user enters the correct response.

Data Files

All programs in STATPAK, except ANVAR[1] in module AVANC, accept data from a file written in the following format:

N M
variable name$_1$
x_1 x_2 $x_3 \ldots x_N$
variable name$_2$
x_1 x_2 $x_3 \ldots x_N$
⋮
variable name$_M$
x_1 x_2 $x_3 \ldots x_N$

1 - The ANVAR program is described on page 71.

where N is the number of observations for each variable.

M is the number of variables.

x is a data entry in free format. There may be as many as six data entries in each line.

NOTE: A variable name may have up to five alphanumeric characters. No variable can be named STOP.

The XEXEC TYPE command is used below to list a sample data file.

```
-TYPE DFILE↵

3 2                 There are two variables with
HGT                 three observations per variable.
66 70 64.5
WGT
120 135 115

-
```

The user may create a data file in one of the editing languages, EDITOR or TECO,[1] or he may use the CREME module, described on page 7. The example below demonstrates the creation of a data file in EDITOR.

```
-EDITOR↵            The user calls EDITOR from XEXEC.
*APPEND↵            The APPEND command indicates that he wishes to enter data.
10 3↵
AGE↵
46 24 32 41 50 63↵
29 28 52 36↵
HGT↵
64 72 71 68 65 75↵
70 64 77 67↵
WGT↵
173 170 154 129 192 203↵
122 136 147 153↵        The APPEND command is terminated by a Control D.
*WRITE SDATA↵       The contents of EDITOR are written on the file SDATA.
NEW FILE↵           The NEW FILE message indicates the user is creating a
                    new file. He confirms this by typing a Carriage Return.[2]
130 CHARACTERS
*QUIT↵

-       The QUIT command returns the user to XEXEC, indicated by the dash.
```

1 - See the *Tymshare TYMCOM-X TECO Reference Manual* or the *Tymshare EDITOR Reference Manual* for more information.

2 - The OLD FILE message here indicates that a file already exists by the specified name. A Carriage Return writes over the previous contents. An Alt Mode/Escape returns the user to EDITOR command level, indicated by the asterisk. He then reexecutes the WRITE command with another file name.

SECTION 3

CREATING AND MANIPULATING DATA: THE CREME MODULE

The CREME module consists of the programs CRATE, MODIF, MERGE, and TRANS, which perform the following functions:

- CRATE creates a data file.
- MODIF allows the user to edit a data file.
- MERGE allows the user to merge data from two or more data files.
- TRANS performs data transformations.

MODIF, MERGE, and TRANS accept data from a file created by CRATE or written in the proper format in one of the editing languages. The format is described in Section 2.

The CRATE Program

CRATE creates a data file in the proper format for all STATPAK programs except ANVAR.[1] The example below demonstrates the use of the CRATE program.

Example

The user calls CRATE to create the file PAYRL, containing three variables, LEVEL, SALRY, and HOURS.

```
-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)



MODULE NAME (TYPE "HELP" FOR AID)? CREME↵

PROGRAM NAME (TYPE "HELP" FOR AID)? CRATE↵


DATA FILE CREATION PROGRAM
FILE NAME (5 OR LESS CHAR.)? PAYRL↵

HOW MANY VARIABLES?3↵

HOW MANY OBSERVATIONS PER VARIABLE?10↵

NAME OF VARIABLE   1 (5 OR LESS CHAR.)? LEVEL↵
```

The file created by STATPAK is actually PAYRL.DAT. Refer to the Tymshare TYMCOM-X XEXEC Reference Manual *for a description of file name extensions.*

There are three variables with ten observations per variable.

The user names each variable and enters all observations for that variable.

1 - The ANVAR program is described on page 71.

```
ENTER OBSERVATIONS 6 TO A LINE
?1 2 5 5 4 3↵          The user separates the observations by spaces.

?3 3 1 1↵

NAME OF VARIABLE   2 (5 OR LESS CHAR.)? SALRY↵

ENTER OBSERVATIONS 6 TO A LINE
?5 4.25 2.6 2.75 3.15 3.70↵

?3.65 4 4.75 4.85↵

NAME OF VARIABLE   3 (5 OR LESS CHAR.)? HOURS↵

ENTER OBSERVATIONS 6 TO A LINE
?45 40 40 38 35 40↵

?40 45 35 40↵


THATS ALL FOR FILE PAYRL

CREATE ANOTHER FILE? NO↵      The user does not wish to create another file.


MODULE NAME (TYPE "HELP" FOR AID)? STOP↵      Control returns to XEXEC.
                                              The execution times are printed.

EXECUTION TIME:          38.64 SEC.
TOTAL ELAPSED TIME:      2 MIN. 19.10 SEC.
NO EXECUTION ERRORS DETECTED

EXIT

-TYPE PAYRL.DAT↵      The file created by CRATE is printed. Note
                      that values are stored as decimal numbers.

   10      3
LEVEL
   1.000000        2.000000        5.000000        5.000000        4.000000
      3.000000
   3.000000        3.000000        1.000000        1.000000
SALRY
   5.000000        4.250000        2.600000        2.750000        3.150000
      3.700000
   3.650000        4.000000        4.750000        4.850000
HOURS
   45.00000        40.00000        40.00000        38.00000        35.00000
      40.00000
   40.00000        45.00000        35.00000        40.00000

-
```

The MODIF Program

This program allows the user to edit a data file. The editing functions available in MODIF are:

Function	Description
ADDON	Adds a specified number of observations to each variable in the data file.
CHANG	Changes a variable name.
CORCT	Corrects individual data elements in the file.
DELET	Deletes a specified number of observations from the beginning or end for each variable.
EXPND	Allows the user to add a new variable with its observations.
LIST	Lists the data file on the terminal.
REMOV	Removes a variable and its observations.

Each of these functions is demonstrated in the example below.

Example

In this example, the user reads the file SDATA and edits the contents using MODIF. The user lists the file which contains variables AGE, HGT, and WGT, with ten observations each. The user adds two observations to each variable using the ADDON function. Then he changes the name of variable HGT to HT. Using the CORCT function, the user changes the eleventh observation for variable WGT, just entered, from 180 to 186. The user then deletes the first observation of each variable using DELET. A new variable, SEX, is added, and an old variable, AGE, is deleted. Finally, the user lists the file with all its modifications.

```
-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? CREME↵

PROGRAM NAME (TYPE "HELP" FOR AID)? MODIF↵        The user calls the
                                                  MODIF program.
  FILE NAME? SDATA↵

FUNCTION? LIST↵          The user lists file SDATA.
```

```
    10

            10                  3
AGE
    46.00000        24.00000        32.00000        41.00000        50.00000
       63.00000
    29.00000        28.00000        52.00000        36.00000
HGT
    64.00000        72.00000        71.00000        68.00000        65.00000
       75.00000
    70.00000        64.00000        77.00000        67.00000
WGT
    173.0000        170.0000        154.0000        129.0000        192.0000
       203.0000
    122.0000        136.0000        147.0000        153.0000
MORE MODIFICATIONS? YES↵

 FILE NAME? SDATA↵

FUNCTION? ADDON↵        The user requests the ADDON function.

HOW MANY OBSERVATIONS TO BE ADDED? 2↵

ENTER NEW OBSERVATIONS FOR VARIABLE AGE   6 TO A LINE
?42 47↵

ENTER NEW OBSERVATIONS FOR VARIABLE HGT   6 TO A LINE
?60 59↵

ENTER NEW OBSERVATIONS FOR VARIABLE WGT   6 TO A LINE
?180 175↵

MORE MODIFICATIONS? YES↵

 FILE NAME? SDATA↵

FUNCTION? CHANG↵        The user requests the CHANG function.

OLD VARIABLE NAME? HGT↵

NEW VARIABLE NAME(5 OR LESS CHAR.)? HT↵     The user changes the variable
                                            name HGT to HT.
CHANGE ANOTHER VARIABLE NAME? NO↵

MORE MODIFICATIONS? YES↵

 FILE NAME? SDATA↵

FUNCTION? CORCT↵        The user requests the CORCT function.

VARIABLE NAME? WGT↵

LINE NUMBER(TYPE 0 TO STOP)? 2↵
```

```
WHICH ENTRY IN LINE     2 ? 5↵
```
The fifth entry on line 2 is modified.

```
WHAT SHOULD IT BE? 186↵
```
The user changes the eleventh observation for variable WGT to 186.

```
ANOTHER ENTRY ON SAME LINE? NO↵

CORRECT ANOTHER VARIABLE? NO↵

MORE MODIFICATIONS? YES↵

 FILE NAME? SDATA↵

FUNCTION? DELET↵
```
The user requests the DELET function.

```
HOW MANY OBSERVATIONS TO BE DELETED?
   >0 => DROP OFF END;  <0 => DROP FROM BEGINNING): -1↵
```
The user deletes the first observation of each variable.

```
MORE MODIFICATIONS? YES↵

 FILE NAME? SDATA↵

FUNCTION? EXPND↵

HOW MANY VARIABLES TO BE ADDED? 1↵

REMINDER   11 OBSERVATIONS PER VARIABLE.
NEW VARIABLE NAME(5 OR LESS CHAR.)? SEX↵

?0 0 1 1 1 1↵

?1 0 0 1 0↵
```
There are 11 observations because the user has added 2 to the initial 10 observations, and deleted the first observation of each variable.

```
THATS ALL FOR FILE SDATA↵

MORE MODIFICATIONS? YES↵

 FILE NAME? SDATA↵

FUNCTION? REMOV↵
```
The user removes the variable AGE.

```
NAME OF VARIABLE TO BE REMOVED? AGE↵

REMOVE ANOTHER VARIABLE? NO↵

MORE MODIFICATIONS? YES↵

 FILE NAME? SDATA↵

FUNCTION? LIST↵
```
The user lists the modified file.

```
            11                  3
HT
    72.00000        71.00000        68.00000        65.00000        75.00000
       70.00000
    64.00000        77.00000        67.00000        60.00000        59.00000

WGT
    170.0000        154.0000        129.0000        192.0000        203.0000
       122.0000
    136.0000        147.0000        153.0000        186.0000        175.0000

SEX
  0.0000000       0.0000000        1.000000        1.000000        1.000000
       1.000000
    1.000000       0.0000000       0.0000000        1.000000       0.0000000

MORE MODIFICATIONS? NO↵

MODULE NAME (TYPE "HELP" FOR AID)?
```

The user may type another module name. He may type STOP to return to XEXEC.

The MERGE Program

This program allows the user to merge selected variables from different data bases into one data base. In order to merge variables, each data file must have the same number of observations per variable. If this is not the case, MERGE excludes the file(s) with an appropriate message.

Example

The user wishes to merge variables from four files. From file STAT1 variables VAR1 and VAR3 are desired, from STAT2 variable VAR5, from STAT3 variable VAR7, and from STAT4 variable VAR8 is desired. The merge information is to be saved on file STATS. Note that STAT4 has a different number of observations and therefore is excluded from the merge.

The contents of the four files STAT1, STAT2, STAT3, and STAT4 are listed with TYPE commands.

```
-TYPE STAT1↵


10 3
VAR1
1 2 3 4 5 6
7 8 9 10
VAR2
3 9 12 15 19 21
24 27 30 33
VAR3
5 4 6 3 2 9
8 4 5 2
```

```
-TYPE STAT2↵

10 2
VAR4
3 7 8 5 2 5
9 4 5 3
VAR5
3 4 5 8 9 17
2 34 12 5


-TYPE STAT3↵

10 2
VAR6
7 6 3 2 1 9
7 4 3 2
VAR7
23 21 28 32 23 24
21 26 27 23


-TYPE STAT4↵

15 1
VAR8
1 2 3 4 5 6
7 8 9 10 11 12
13 14 15


-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? CREME↵

PROGRAM NAME (TYPE "HELP" FOR AID)? MERGE↵      The user requests the
                                                 MERGE program.
NAME OF OUTPUT FILE.?STATS↵      The file STATS.DAT will contain the merged results.

NAME OF INPUT FILE?.TYPE STOP WHEN DONE.
STAT1↵

TYPE VARIABLE NAMES IN RESPONSE TO ?.TYPE STOP WHEN DONE.
```

```
?VAR1↵

?VAR3↵

?STOP↵

NAME OF INPUT FILE?.TYPE STOP WHEN DONE.
STAT2↵

TYPE VARIABLE NAMES IN RESPONSE TO ?.TYPE STOP WHEN DONE.

?VAR5↵

?STOP↵

NAME OF INPUT FILE?.TYPE STOP WHEN DONE.
STAT3↵

TYPE VARIABLE NAMES IN RESPONSE TO ?.TYPE STOP WHEN DONE.

?VAR7↵

?STOP↵

NAME OF INPUT FILE?.TYPE STOP WHEN DONE.
STAT4↵

TYPE VARIABLE NAMES IN RESPONSE TO ?.TYPE STOP WHEN DONE.

?VAR8↵

INCOMPATIBLE NO. OF OBSERVATIONS    15 IN FILE STAT4.
FILE EXCLUDED IN MERGE ROUTINE.

NAME OF INPUT FILE?.TYPE STOP WHEN DONE.
STOP↵

MODULE NAME (TYPE "HELP" FOR AID)? STOP↵


EXECUTION TIME:          57.31 SEC.
TOTAL ELAPSED TIME:      2 MIN. 6.17 SEC.
NO EXECUTION ERRORS DETECTED

EXIT
```

-TYPE STATS.DAT↵ *The file STATS.DAT contains the results of the MERGE routine.*

```
          10          4
VAR1
   1.000000    2.000000    3.000000    4.000000    5.000000
      6.000000
   7.000000    8.000000    9.000000    10.00000
VAR3
   5.000000    4.000000    6.000000    3.000000    2.000000
      9.000000
   8.000000    4.000000    5.000000    2.000000
VAR5
   3.000000    4.000000    5.000000    8.000000    9.000000
      17.00000
   2.000000    34.00000    12.00000    5.000000
VAR7
   23.00000    21.00000    28.00000    32.00000    23.00000
      24.00000
   21.00000    26.00000    27.00000    23.00000
```

-

The TRANS Program

TRANS allows the user to create a data file that is a simple function of his original data file. The input data file is limited to 60 variables and 500 observations per variable. The total number of observations on the file may not exceed 4000.

The transformations that TRANS can perform are listed in the table below.

Name	Function	Comments
ABS	$\|x\|$	Absolute value.
SQR	$\sqrt{x}$	Square root; x must be greater than or equal to zero.
PWR	Cx^D	
CPW	CD^x	
INV	C/x	x must not equal zero.
AMC	C+Dx	
EXP	Ce^x	Exponential; e is base of natural logarithm.
LOG	$\log_e x$	Natural logarithm; x must not equal zero.
LGT	$\log_{10} x$	Base 10 logarithm; x must not equal zero.
LIN	Cx+Dy	
MUL	Cxy	
DIV	Cx/Dy	y must not equal zero.

where x and y are variables, and

C and D are constants supplied by the user.

As an example, the LIN transformation, Cx+Dy, is performed on the first set of data below to produce the second set. The constants, C and D, are set to 2 and 3, respectively.

Original Data

variable x: 3 4 5 8
variable y: 4 2 6 7

Transformed Data

variable x: 18 14 28 37
variable y: 4 2 6 7

Thus, the observations of variable x are replaced by 2x+3y, and the observations of y are unchanged.

Example

In the example below, the user reads a data file that has four variables with ten observations each. The variables and values are listed below.

V1	V2	V3	V4
5	32	10	56
10	14	11	58
12	76	12	60
14	20	13	62
11	12	14	64
9	42	15	60
7	15	16	42
16	76	17	34
20	24	18	47
12	18	19	78

The user transforms V1 to V1+V3, V2 to V2/V3, and V3 to $3(V3)^2$. He writes the transformed data on file TRANF. The original file, VARFL, is unchanged.

```
-TYPE VARFL↵          The contents of the file VARFL are displayed.

10 4
V1
5 10 12 14 11 9
7 16 20 12
V2
32 14 76 20 12 42
15 76 24 18
V3
10 11 12 13 14 15
16 17 18 19
V4
56 58 60 62 64 60
42 34 47 78


-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? CREME↵

PROGRAM NAME (TYPE "HELP" FOR AID)? TRANS↵

ENTER FILE NAME? VARFL↵          The original data is found on file VARFL.

ENTER NAME OF VARIABLE X? V1↵

TRANSFORMATION TYPE? LIN↵

ENTER NAME OF VARIABLE Y? V3↵

ENTER VALUES OF CONSTANTS C AND D? 1 1↵          The user replaces V1 with V1+V3.

ENTER NAME OF VARIABLE X? V2↵

TRANSFORMATION TYPE? DIV↵

ENTER NAME OF VARIABLE Y? V3↵

ENTER VALUES OF CONSTANTS C AND D? 1 1↵          The user replaces V2 with V2/V3.

ENTER NAME OF VARIABLE X? V3↵

TRANSFORMATION TYPE? PWR↵

ENTER VALUES OF CONSTANTS C AND D? 3 2↵          The user replaces V3 with 3(V3)^2.

ENTER NAME OF VARIABLE X? STOP↵          The user replies with STOP when he is finished
                                         making transformations.
ENTER OUTPUT FILE NAME? TRANF↵           The transformed data is written on file TRANF.
```

```
MODULE NAME (TYPE "HELP" FOR AID)? STOP↵

EXECUTION TIME:          52.90 SEC.
TOTAL ELAPSED TIME:      1 MIN. 44.42 SEC.
NO EXECUTION ERRORS DETECTED

EXIT

-TYPE TRANF.DAT↵          The user lists his transformed data.

              10                  4
V1
    15.00000          21.00000          24.00000          27.00000          25.00000
       24.00000
    23.00000          33.00000          38.00000          31.00000
V2
    3.200000          1.272727          6.333333          1.538462          0.8571429
       2.800000
  0.9375000           4.470588          1.333333          0.9473684
V3
    300.0000          363.0000          432.0000          507.0000          588.0000
       675.0000
    768.0000          867.0000          972.0000          1083.000
V4
    56.00000          58.00000          60.00000          62.00000          64.00000
       60.00000
    42.00000          34.00000          47.00000          78.00000
```

SECTION 4

BASIC STATISTICAL ANALYSES: THE DATA MODULE

The DATA module has only one program, DSCRE, which allows data screening, optionally prints a histogram and frequency table, and generates basic statistics.

This program accepts data from a data file written in the standard format for STATPAK, described on page 4. DSCRE accepts a maximum of 40 variables and 75 observations per variable.

The user selects a variable for analysis. He may screen the observations to be used for this variable by stipulating the conditions by which observations are chosen.

For example, the user selects for analysis the variable CASH from the data file below.

4 3
DEBT
24.5 32.35 37.2 26.75
CRED
730.0 680.0 645.0 752.5
CASH
198.25 162.0 170.0 156.5

The user may stipulate that CRED observations be greater than 650 and DEBT observations be greater than 25 using the conditions GT 650 and GT 25. This causes the following CASH observations to be used for analysis: 162.0 and 156.5.

The conditions that may be used in DSCRE are:

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

The user may enter a maximum of 30 conditions for data screening. More than one condition corresponding to the same variable can be entered on the same line. For example,

LT 20 GT 15

After the user has selected the variable and observations to be analyzed, he may print or save the chosen subset of data. DSCRE then computes and prints the following:

- Histogram of up to 12 frequency classes (optional).
- Statistics summary including maximum, minimum, average, median, variance, and standard deviation.
- Frequency table of up to 12 frequency classes (optional).
- Chi-square measure of goodness-of-fit.

Example

The user chooses variable TOTAL from the file ACCT for analysis. He selects those observations of TOTAL that correspond to zero values of PAST and values between 0 and 3999, inclusive of variable PRES. The user saves this subset of data on file XACCT.DAT. The user chooses to print a histogram and frequency table as well as the basic statistics.

```
-TYPE ACCT↵

30 3
PRES
842.21 0.10 5146.67 355.31 .00 2824.94
3849.72 .00 1198.91 .00 .00 .00
271.00 .00 .00 .00 86.91 1060.10
425.33 18.78 .00 1582.87 734.30 160.93
11.77 .00 -203.65 2874.29 .00 80.00
PAST
.00 .00 156.45 .00 248.53 336.18
.00 .00 .00 .00 69.12 25.00
.00 207.30 .00 .00 .00 .00
.00 .00 .00 .00 .00 .00
.00 92.48 .00 .00 270.00 .00
TOTAL
842.21 0.10 6401.15 355.31 248.53 4000.30
3849.70 .00 1198.91 549.58 232.36 25.00
271.00 207.30 .00 .00 86.91 1060.10
425.33 18.78 .00 1582.82 734.30 160.93
20.74 92.48 -203.65 4392.70 613.67 210.00
```

```
-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? DATA↵

PROGRAM NAME (TYPE "HELP" FOR AID)? DSCRE↵

ENTER THE NAME OF DATA FILE?ACCT↵          The file ACCT contains the original data.

NAME OF VARIABLE CHOSEN FOR ANALYSIS?.TOTAL↵

ARE THERE ANY VARIABLES SUBJECT TO CONDITIONS?.YES↵

START ENTERING CONDITIONS NOW.

ENTER NAME OF VARIABLE?PRES↵

ENTER CODE OF CONDITION AND VALUE?GE 0 LT 4000↵

ENTER NAME OF VARIABLE?PAST↵

ENTER CODE OF CONDITION AND VALUE?EQ 0↵

ENTER NAME OF VARIABLE?STOP↵          The user types STOP to stop entering conditions.

DO YOU WANT A LISTING OF THE SUBSET VECTOR?.NO↵


STORE THE SUBMATRIX OF OBSERVATIONS.?
TYPE FILENAME OTHERWISE TYPE NO.
XACCT↵          The file XACCT.DAT contains the submatrix.

DO YOU WANT A HISTOGRAM?.YES↵

DO YOU WANT A FREQUENCY TABLE.?YES↵

ENTER NO OF FREQ CLASSES?12↵


DATA SCREENING PROBLEM  1


SUMMARY STATISTICS FOR VARIABLE TOTAL

TOT=    15759.42      AVER=    750.4486      STD DEVIATION =    1182.519
MINIMUM =  0.0000000      MAXIMUM =    4392.700      VARIANC=     1398351.


MEDIAN =     320.3010
```

HISTOGRAM 1

```
FREQUENCY   12    2    3    1    1    0    0    0    0    0    1    1
-------------------------------------------------------------------
   12       *
   11       *
   10       *
    9       *
    8       *
    7       *
    6       *
    5       *
    4       *
    3       *         *
    2       *    *    *
    1       *    *    *    *    *                             *    *
-------------------------------------------------------------------
INTERVAL     1    2    3    4    5    6    7    8    9   10   11   12
CLASS
```

FREQUENCY TABLE

CLASS	RANGE			FREQUENCY
1	0.0000000	--	366.0583	12
2	366.0583	--	732.1167	2
3	732.1167	--	1098.175	3
4	1098.175	--	1464.233	1
5	1464.233	--	1830.292	1
6	1830.292	--	2196.350	0
7	2196.350	--	2562.408	0
8	2562.408	--	2928.467	0
9	2928.467	--	3294.525	0
10	3294.525	--	3660.583	0
11	3660.583	--	4026.642	1
12	4026.642	--	4392.700	1

MEASURE OF GOODNESS OF FIT
CHI SQUARE = 71.00000 WITH DEGREES OF FREEDOM 11

ENTER THE NAME OF DATA FILE?STOP↵

MODULE NAME (TYPE "HELP" FOR AID)?

The DSCRE program has terminated. The user may call another module or type STOP to return to XEXEC.

SECTION 5

REGRESSION ANALYSES: THE REGON MODULE

The regression analyses available in REGON are the multiple linear regression analysis, the stepwise multiple linear regression analysis, and the polynomial regression analysis. These analyses are performed by the programs DREGR, RGSTP, and RGPOL, respectively. These programs accept data from a file written in the standard format for STATPAK, described on page 4.

The DREGR Program

DREGR performs a multiple linear regression analysis between a dependent variable, y, and a set of independent variables, $x_1, x_2, \ldots, x_m$, based on a set of observations. A linear relationship of the form

$$y = a + b_1 x_1 + b_2 x_2 + \cdots + b_m x_m$$

is established, where a is an intercept, and b_i is a regression coefficient.

DREGR accepts up to 40 variables and 75 observations per variable. The number of observations per variable should exceed the number of variables by at least two.

When executing the program, the user enters the name of the data file, the dependent variable, and each of the independent variables. The list of independent variables is terminated by typing STOP, followed by a Carriage Return. DREGR then computes and prints the regression analysis statistics, an analysis of variance for the regression, and a table of residuals. The program then requests the next dependent variable. The user types STOP to exit the program.

Example

A multiple linear regression analysis is performed on the data in file DATA2, below. FACT6 is the dependent variable. FACT1, FACT2, FACT3, FACT4, and FACT5 are the independent variables.

```
-TYPE DATA2↵

15 8
FACT1
29 30 30 30 35 35
43 43 44 44 44 44
44 44 45
FACT2
289 391 424 313 243 365
396 356 346 156 278 349
141 245 297
```

```
FACT3
216 244 246 239 275 219
267 274 255 258 249 252
236 236 256
FACT4
85 92 90 91 95 95
100 79 126 95 110 88
129 97 111
FACT5
14 16 18 10 30 21
39 19 56 28 42 21
56 24 45
FACT6
1 2 2 0 2 2
3 2 3 0 4 1
1 1 3
FACT7
1 2 3 4 5 6
7 8 9 10 11 12
13 14 15
FACT8
10 16 20 23 25 26
30 36 48 62 78 94
107 118 127


-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)



MODULE NAME (TYPE "HELP" FOR AID)? REGON↵

PROGRAM NAME (TYPE "HELP" FOR AID)? DREGR↵

ENTER NAME OF DATA FILE?DATA2↵


NAME OF DEPENDENT VARIABLE?FACT6↵

ENTER NAMES OF INDEPENDENT VARIABLES,ONE TO A ROW IN RESPONSE TO ?.
 TYPE STOP WHEN DONE.
```

? FACT1↵

? FACT2↵

? FACT3↵

? FACT4↵

? FACT5↵

? STOP↵

VARIABLE NAME	MEAN	STANDARD DEVIATION	CORRELATION X VS Y	REGRESSION COEFFICIENT	STD. ERROR OF REG.COEF	CAL T VAL
FACT1	38.933	6.508	0.266	-0.035	0.053	-0.655
FACT2	305.933	83.365	0.412	0.008	0.003	2.958
FACT3	248.133	17.423	0.309	-0.006	0.017	-0.368
FACT4	98.867	14.287	0.360	-0.073	0.060	-1.216
FACT5	29.267	14.911	0.505	0.133	0.065	2.032
DEPENDENT FACT6	1.800	1.146				

INTERCEPT 5.525384

MULTIPLE CORRELATION 0.8427965

STD. ERROR OF ESTIMATE 0.7695865

ANALYSIS OF VARIANCE FOR THE REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F VAL
DUE TO REGRESSION	5	13.06963	2.613926	4.413453
DEVIATION FROM REGRESSION	9	5.330370	0.5922633	
TOTAL	14	18.40000		

NEED TABLE OF RESIDUALS.?
TYPE TTY, DSK, BOTH OR NONE.
BOTH↵ *The table of residuals is printed on the terminal and on a file.*

```
MULTIPLE REGRESSION.....Y= FACT6

                TABLE OF RESIDUALS

CASE NO.    Y VALUE      Y ESTIMATE      RESIDUAL
     1    1.000000        1.196935     -0.1969345
     2    2.000000        1.589456      0.4105439
     3    2.000000        2.264102     -0.2641021
     4   0.0000000       0.2468169     -0.2468169
     5    2.000000        1.621938      0.3780617
     6    2.000000        1.800457      0.1995433
     7    3.000000        3.499589     -0.4995890
     8    2.000000        2.003885     -0.3884822E-02
     9    3.000000        3.476657     -0.4766569
    10   0.0000000       0.4241671     -0.4241671
    11    4.000000        2.259888       1.740112
    12    1.000000        1.656622     -0.6566221
    13    1.000000        1.666003     -0.6660032
    14    1.000000       0.6291739      0.3708261
    15    3.000000        2.664310      0.3356900
ENTER OUTPUT FILENAME.?RES2↵        The file RES2.DAT contains the residuals.

NAME OF RESIDUAL ARRAY?.ARR2↵       The residual array on the file is named ARR2.

NAME OF DEPENDENT VARIABLE?STOP↵

MODULE NAME (TYPE "HELP" FOR AID)?
```

The contents of RES2.DAT are shown below.

```
              15                1
ARR2                            RESIDUAL DATA   FACT6
 -0.1969345       0.4105439      -0.2641021      -0.2468169       0.3780617
    0.1995433
 -0.4995890      -0.3884822E-02  -0.4766569      -0.4241671        1.740112
   -0.6566221
 -0.6660032       0.3708261       0.3356900
```

The RGSTP Program

This program performs a stepwise multiple linear regression analysis between a dependent variable, y, and a set of independent variables, $x_1, x_2, \ldots, x_m$, based on a set of observations. A linear relationship of the form

$$y = a + b_1 x_1 + b_2 x_2 + \cdots + b_m x_m$$

is established, where a is an intercept, and b_i is a regression coefficient.

RGSTP accepts up to 40 variables and 75 observations per variable. The number of observations per variable should exceed the number of variables by at least three.

RGSTP computes and prints the following:

- The mean and standard deviation for each variable (optional).
- A correlation matrix (optional).
- The regression analysis as each variable is entered.
- A table of residuals (optional).

When executing RGSTP, the user enters the name of the data file, the dependent variable, and the independent variables with the appropriate codes. The code signifies how the independent variable is to be used in the regression. The codes are:

Code	Meaning
0	The variable is free to enter or leave the regression.
1	The variable is forced into the regression regardless of the entry criterion described below.
2	The variable is forced out of the regression regardless of the entry criterion.

To terminate the list of independent variables and codes, the user types STOP, followed by a Carriage Return.

RGSTP requests an entry criterion, PCT, which is the proportion of the sum of squares of the dependent variable. To be entered into the regression, a variable must account for a proportion of variability at least as large as PCT. The value of PCT should be between 0 and 1, inclusive. The user who has no special choice of this criterion may use the value 0.

The user may perform as many analyses as he chooses and may terminate the program by typing STOP in response to:

NAME OF DEPENDENT VARIABLE?

Example

The data below corresponds to 20 lots on a tract of land. The variables are AREA in square feet, ELEVN (elevation) in feet above sea level, SLOPE in degrees, VIEW on a scale of 1 to 9 ranging from poor to excellent, and PRICE in thousands of dollars. The data is stored on a file named FILE1.

```
-TYPE FILE1↵

20 5
AREA
14.7 14.2 12.7 13.8 14.4 17.4
21.8 14 17.5 23 18.3 19.4
15.2 18.3 21.7 16.7 13.6 14.5
12.1 17.4
ELEVN
155 155 158 158 155 157
172 170 175 185 185 205
215 195 178 160 205 190
203 125
SLOPE
1.5 1.8 2.9 1 .5 1
5.7 5.4 17.5 14.5 14.4 12.2
5 13.1 15.2 10.1 7.4 5.8
5.1 17.3
VIEW
2 2 1 1 2 2
4 6 9 9 9 9
8 6 8 8 7 7
7 1
PRICE
4.1 3.9 3.2 2.9 3.9 4.1
5.8 5.1 6.8 6.8 6.5 7
5.8 5.1 5.3 4.9 6 5.3
4.8 4.3
```

The user performs a stepwise multiple linear regression with PRICE as the dependent variable, and AREA, ELEVN, VIEW, and SLOPE as independent variables. The entry criterion is set to zero.

```
-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? REGON↵

PROGRAM NAME (TYPE "HELP" FOR AID)? RGSTP↵

ENTER NAME OF DATA FILE.?FILE1↵
```

```
NAME OF DEPENDENT VARIABLE.?PRICE↵

START ENTERING INDEPENDENT VARIABLES AND CODES NOW.
CODE 0 MEANS A VARIABLE IS FREE TO ENTER OR LEAVE REGRESSION
CODE 1 MEANS A VARIABLE IS FORCED INTO REGRESSION
CODE 2 MEANS A VARIABLE IS FORCED OUT OF REGRESSION.

ENTER NAME OF INDEPENDENT VARIABLE.?AREA↵

ENTER CODE.?0↵

ENTER NAME OF INDEPENDENT VARIABLE.?ELEVN↵

ENTER CODE.?0↵

ENTER NAME OF INDEPENDENT VARIABLE.?SLOPE↵

ENTER CODE.?0↵

ENTER NAME OF INDEPENDENT VARIABLE.?VIEW↵

ENTER CODE.?0↵

ENTER NAME OF INDEPENDENT VARIABLE.?STOP↵

DO YOU WANT A PRINTOUT OF MEANS AND STD.DEVS.?YES↵

DO YOU WANT A PRINTOUT OF CORRELATION MATRIX.?YES↵

NEED RESIDUAL TABLE?.
TYPE TTY,DSK,BOTH OR NONE.
BOTH↵

ENTER OUTPUT FILENAME?.RES1↵
```

The residuals are printed on the file RES1.DAT and the terminal.

```
ENTER THE VALUE OF PCT,THE PROPORTION OF SUM OF
SQUARES THAT SHOULD BE USED AS CRITERION IN ENTERING A
VARIABLE INTO REGRESSION.VALUE SHOULD BE BETWEEN 0 AND 1.
0↵

NUMBER OF OBSERVATIONS                    20
NUMBER OF VARIABLES                        5
CONSTANT TO LIMIT VARIABLES             0.00000


VARIABLE      MEAN       STANDARD
NAME                     DEVIATION


PRICE     5.080000        1.198947
AREA      16.53500        3.156326
ELEVN     175.0500        22.80230
SLOPE     7.870000        5.871976
VIEW      5.400000        3.135535
```

CORRELATION MATRIX

ROW	PRICE				
	1.000000	0.5775156	0.6445836	0.6644410	0.8786537
ROW	AREA				
	0.5775156	1.000000	0.7039703E-01	0.6297176	0.3963015
ROW	ELEVN				
	0.6445836	0.7039703E-01	1.000000	0.1515843	0.7490887
ROW	SLOPE				
	0.6644410	0.6297176	0.1515843	1.000000	0.6075628
ROW	VIEW				
	0.8786537	0.3963015	0.7490887	0.6075628	1.000000

DEPENDENT VARIABLE PRICE
NUMBER OF VARIABLES FORCED 0
NUMBER OF VARIABLES DELETED 0

STEP 1
VARIABLE ENTERED VIEW
SUM OF SQUARES REDUCED IN THIS STEP 21.08575
PROPORTION REDUCED IN THIS STEP 0.7720323
CUMULATIVE SUM OF SQUARES REDUCED 21.08575
CUMULATIVE PROPORTION REDUCED 0.7720323 OF 27.31200

FOR 1 VARIABLES ENTERED
MULTIPLE CORRELATION COEFFICIENT 0.8786537
(ADJUSTED FOR D.F.) 0.8786537
F VALUE FOR ANALYSIS OF VARIANCE 60.95856
STANDARD ERROR OF ESTIMATE 0.5881352
(ADJUSTED FOR D.F.) 0.5881352

VARIABLE NAME	REGRESSION COEFFICIENT	STD ERROR OF REG COEFF	COMPUTED T-VALUE
VIEW	0.3359743	0.4303172E-01	7.807596
INTERCEPT	3.265739		

STEP 2
VARIABLE ENTERED AREA
SUM OF SQUARES REDUCED IN THIS STEP 1.703536
PROPORTION REDUCED IN THIS STEP 0.6237634E-01
CUMULATIVE SUM OF SQUARES REDUCED 22.78938
CUMULATIVE PROPORTION REDUCED 0.8344092 OF 27.31200

FOR 2 VARIABLES ENTERED
MULTIPLE CORRELATION COEFFICIENT 0.9134600
(ADJUSTED FOR D.F.) 0.9084105
F VALUE FOR ANALYSIS OF VARIANCE 42.83134
STANDARD ERROR OF ESTIMATE 0.5157871
(ADJUSTED FOR D.F.) 0.5299208

VARIABLE NAME	REGRESSION COEFFICIENT	STD ERROR OF REG COEFF	COMPUTED T-VALUE
VIEW	0.2947526	0.4110384E-01	7.170926
AREA	0.1033309	0.4083308E-01	2.530568

INTERCEPT 1.779760

STEP 3
VARIABLE ENTERED ELEVN
SUM OF SQUARES REDUCED IN THIS STEP 0.1664590
PROPORTION REDUCED IN THIS STEP 0.6094720E-02
CUMULATIVE SUM OF SQUARES REDUCED 22.95584
CUMULATIVE PROPORTION REDUCED 0.8405039 OF 27.31200

FOR 3 VARIABLES ENTERED
MULTIPLE CORRELATION COEFFICIENT 0.9167900
(ADJUSTED FOR D.F.) 0.9064986
F VALUE FOR ANALYSIS OF VARIANCE 28.10531
STANDARD ERROR OF ESTIMATE 0.5217853
(ADJUSTED FOR D.F.) 0.5516253

VARIABLE NAME	REGRESSION COEFFICIENT	STD ERROR OF REG COEFF	COMPUTED T-VALUE
VIEW	0.2532135	0.6746307E-01	3.753365
AREA	0.1162890	0.4450826E-01	2.612752
ELEVN	0.6676355E-02	0.8538421E-02	0.7819191

INTERCEPT 0.6211119

```
STEP  4
VARIABLE ENTERED                                    SLOPE
SUM OF SQUARES REDUCED IN THIS STEP          0.1975852
PROPORTION REDUCED IN THIS STEP              0.7234373E-02
CUMULATIVE SUM OF SQUARES REDUCED             23.15343
CUMULATIVE PROPORTION REDUCED                0.8477383        OF     27.31200

FOR  4 VARIABLES ENTERED
MULTIPLE CORRELATION COEFFICIENT        0.9207270
        (ADJUSTED FOR D.F.)             0.9050907
F VALUE FOR ANALYSIS OF VARIANCE         20.87864
STANDARD ERROR OF ESTIMATE              0.5265341
        (ADJUSTED FOR D.F.)             0.5737773

     VARIABLE      REGRESSION      STD ERROR OF     COMPUTED
     NAME          COEFFICIENT     REG COEFF        T-VALUE
     VIEW        0.2048664         0.8896190E-01    2.302856
     AREA        0.9873099E-01     0.4949519E-01    1.994759
     ELEVN       0.1067605E-01     0.9832821E-02    1.085756
     SLOPE       0.2949830E-01     0.3494189E-01   0.8442104

     INTERCEPT  0.2402108



                TABLE OF RESIDUALS

CASE NO.      Y VALUE      Y ESTIMATE       RESIDUAL
      1    4.100000          3.800324        0.2996761
      2    3.900000          3.759808        0.1401921
      3    3.200000          3.471321       -0.2713212
      4    2.900000          3.523879       -0.6238786
      5    3.900000          3.741206        0.1587937
      6    4.100000          4.073500        0.2649951E-01
      7    5.800000          5.216432        0.5835675
      8    5.100000          4.825862        0.2741379
      9    6.800000          6.196330        0.6036704
     10    6.800000          6.757616        0.4238433E-01
     11    6.500000          6.290630        0.2093699
     12    7.000000          6.547859        0.4521412
     13    5.800000          5.822695       -0.2269495E-01
     14    5.100000          5.744443       -0.6444435
     15    5.300000          6.370315        -1.070315
     16    4.900000          5.534050       -0.6340502
     17    6.000000          5.423894        0.5761057
     18    5.300000          5.305414       -0.5414307E-02
     19    4.800000          5.186600       -0.3865997
     20    4.300000          4.007823        0.2921771
NAME OF RESIDUAL ARRAY?.ARR1

NAME OF DEPENDENT VARIABLE.?STOP

MODULE NAME (TYPE "HELP" FOR AID)?
```

The RGPOL Program

RGPOL performs a polynomial regression analysis between a dependent variable, y, and an independent variable, x, based on a set of observations. A polynomial relationship of the form

$$y = a + b_1 x + b_2 x^2 + \cdots + b_k x^k$$

is established, where a is an intercept, and b_k is a regression coefficient.

RGPOL accepts up to 75 observations per variable. The number of observations per variable should exceed the degree of the polynomial by at least two.

The analysis begins with a first degree polynomial of the form

$$y = a + b_1 x$$

then proceeds to the next degree polynomial,

$$y = a + b_1 x + b_2 x^2$$

and continues to the polynomial

$$y = a + b_1 x + b_2 x^2 + \cdots + b_k x^k$$

where k is the degree specified by the user, between 1 and 10, inclusive. At each step, RGPOL checks on the improvement in fit caused by going to a higher degree polynomial. If there is no improvement, higher degree fits are ignored.

RGPOL allows a lead/lag analysis of data. This analysis is particularly useful with a first degree polynomial of the form:

$$y = ax + b$$

The user can analyze an independent variable whose effect on the dependent variable leads or lags the corresponding observations.

Example 1

This example demonstrates the lead/lag feature of RGPOL. The user has a 12-month table of advertising and sales. He believes advertising affects sales with a two-month lag. The data, in thousands of dollars, is stored on the file MKTG.

```
-TYPE MKTG↵

12 2
ADVTG
10 12 11 11 10 12
12 12 10 10 12 12
SALES
200 210 205 222 217 218
198 205 210 211 207 202
```

```
-R STATPAK↵
TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? REGON↵
PROGRAM NAME (TYPE "HELP" FOR AID)? RGPOL↵
ENTER THE NAME OF DATA FILE.?MKTG↵
NAME OF DEPENDENT VARIABLE.?SALES↵
NAME OF INDEPENDENT VARIABLE.?ADVTG↵
HIGHEST DEGREE OF POLYNOMIAL TO BE TRIED.?1↵
```

The user assumes a linear relationship.

```
IF THE DEPENDENT VARIABLE LEADS/LAGS THE INDEPENDENT
VARIABLE ENTER PERIODS AS A POSITIVE/NEGATIVE NUMBER.
OTHERWISE TYPE 0.
2↵

NUMBER OF OBSERVATIONS      10
```

The number of observations is 10 instead of 12 because there is a lead of 2 periods.

```
POLYNOMIAL REGRESSION OF DEGREE  1
----------------------------------

  INTERCEPT    160.0000

  REGRESSION COEFFICIENTS
  4.500000

                    ANALYSIS OF VARIANCE FOR   1  DEGREE POLYNOMIAL

SOURCE OF VARIATION  D.F.  SUM OF SQUQRES   MEAN SQUARE   F VAL   IMPROVEMENT IN SS

DUE TO REGRESSION      1    162.0000        162.0000       3.60     162.000
DEV FROM REGRESSN      8    360.5000        45.06250

TOTAL                  9    522.5000

STANDARD ERROR OF FORECAST        45.062

NEED RESIDUAL TABLE?.
TYPE TTY,DSK,BOTH OR NONE.
BOTH↵
```

The user requests that the residual table be printed on the terminal and written on a file.

```
OUTPUT FILENAME? .MRES↵
```

The residuals are written on file MRES.DAT.

```
POLYNOMIAL REGRESSION.....X= ADVTG Y= SALES


POLYNOMIAL REGRESSION OF DEGREE  1
----------------------------------
```

TABLE OF RESIDUALS

OBSERVATION	X VALUE	Y VALUE	Y ESTIMATE	RESIDUAL
1	10.00000	205.0000	205.0000	0.0000000
2	12.00000	222.0000	214.0000	8.000000
3	11.00000	217.0000	209.5000	7.500000
4	11.00000	218.0000	209.5000	8.500000
5	10.00000	198.0000	205.0000	-7.000000
6	12.00000	205.0000	214.0000	-9.000000
7	12.00000	210.0000	214.0000	-4.000000
8	12.00000	211.0000	214.0000	-3.000000
9	10.00000	207.0000	205.0000	2.000000
10	10.00000	202.0000	205.0000	-3.000000

```
RESIDUAL ARRAY NAME?.RESID↵


MODULE NAME (TYPE "HELP" FOR AID)? STOP↵


EXECUTION TIME:          2 MIN. 0.09 SEC.
TOTAL ELAPSED TIME:      4 MIN. 22.43 SEC.
NO EXECUTION ERRORS DETECTED

EXIT

-TYPE MRES.DAT↵          The user prints the file of residuals.


            10                  1
RESID                            RESIDUAL DATA  SALES
  0.0000000        8.000000        7.500000        8.500000       -7.000000
    -9.000000
  -4.000000       -3.000000        2.000000       -3.000000
```

Example 2

The user performs a polynomial regression analysis of the fourth degree on the variables FACT8 and FACT7 of data file DATA2.

```
-TYPE DATA2↵

15 8
FACT1
29 30 30 30 35 35
43 43 44 44 44 44
44 44 45
FACT2
289 391 424 313 243 365
396 356 346 156 278 349
141 245 297
FACT3
216 244 246 239 275 219
267 274 255 258 249 252
236 236 256
FACT4
85 92 90 91 95 95
100 79 126 95 110 88
129 97 111
FACT5
14 16 18 10 30 21
39 19 56 28 42 21
56 24 45
FACT6
1 2 2 0 2 2
3 2 3 0 4 1
1 1 3
FACT7
1 2 3 4 5 6
7 8 9 10 11 12
13 14 15
FACT8
10 16 20 23 25 26
30 36 48 62 78 94
107 118 127
```

```
-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? REGON↵

PROGRAM NAME (TYPE "HELP" FOR AID)? RGPOL↵

ENTER THE NAME OF DATA FILE.?DATA2↵

NAME OF DEPENDENT VARIABLE.?FACT8↵

NAME OF INDEPENDENT VARIABLE.?FACT7↵

HIGHEST DEGREE OF POLYNOMIAL TO BE TRIED.?4↵

IF THE DEPENDENT VARIABLE LEADS/LAGS THE INDEPENDENT
VARIABLE ENTER PERIODS AS A POSITIVE/NEGATIVE NUMBER.
OTHERWISE TYPE 0.
0↵        There is no lead or lag.

NUMBER OF OBSERVATIONS     15


POLYNOMIAL REGRESSION OF DEGREE  1
----------------------------------


  INTERCEPT  -13.87619

  REGRESSION COEFFICIENTS
  8.567857


                        ANALYSIS OF VARIANCE FOR   1  DEGREE POLYNOMIAL

SOURCE OF VARIATION  D.F.  SUM OF          MEAN          F     IMPROVEMENT
                           SQUQRES         SQUARE        VAL     IN SS

DUE TO REGRESSION     1    20554.29        20554.29      135.57 20554.289
DEV FROM REGRESSN    13    1971.045        151.6188

TOTAL                14    22525.33

STANDARD ERROR OF FORECAST       151.619
```

POLYNOMIAL REGRESSION OF DEGREE 2

INTERCEPT 15.07252

REGRESSION COEFFICIENTS
-1.649337 0.6385746

ANALYSIS OF VARIANCE FOR 2 DEGREE POLYNOMIAL

SOURCE OF VARIATION	D.F.	SUM OF SQUQRES	MEAN SQUARE	F VAL	IMPROVEMENT IN SS
DUE TO REGRESSION	2	22236.51	11118.25	461.94	1682.219
DEV FROM REGRESSN	12	288.8257	24.06881		
TOTAL	14	22525.33			

STANDARD ERROR OF FORECAST 24.069

POLYNOMIAL REGRESSION OF DEGREE 3

INTERCEPT 18.52399

REGRESSION COEFFICIENTS
-3.885354 0.9769363 -0.1409850E-01

ANALYSIS OF VARIANCE FOR 3 DEGREE POLYNOMIAL

SOURCE OF VARIATION	D.F.	SUM OF SQUQRES	MEAN SQUARE	F VAL	IMPROVEMENT IN SS
DUE TO REGRESSION	3	22247.81	7415.936	293.94	11.300
DEV FROM REGRESSN	11	277.5259	25.22963		
TOTAL	14	22525.33			

STANDARD ERROR OF FORECAST 25.230

POLYNOMIAL REGRESSION OF DEGREE 4

INTERCEPT -6.042034

REGRESSION COEFFICIENTS

19.98636	-5.247136	0.5774151	-0.1848480E-01

ANALYSIS OF VARIANCE FOR 4 DEGREE POLYNOMIAL

SOURCE OF VARIATION	D.F.	SUM OF SQUQRES	MEAN SQUARE	F VAL	IMPROVEMENT IN SS
DUE TO REGRESSION	4	22508.35	5627.087	3312.91	260.541
DEV FROM REGRESSN	10	16.98535	1.698535		
TOTAL	14	22525.33			

STANDARD ERROR OF FORECAST 1.699

NEED RESIDUAL TABLE?.
TYPE TTY, DSK, BOTH OR NONE.
TTY

POLYNOMIAL REGRESSION.....X= FACT7 Y= FACT8

POLYNOMIAL REGRESSION OF DEGREE 4

TABLE OF RESIDUALS

OBSERVATION	X VALUE	Y VALUE	Y ESTIMATE	RESIDUAL
1	1.000000	10.00000	9.256118	0.7438824
2	2.000000	16.00000	17.26570	-1.265702
3	3.000000	20.00000	20.78576	-0.7857559
4	4.000000	23.00000	22.17168	0.8283164
5	5.000000	25.00000	23.33525	1.664749
6	6.000000	26.00000	25.74459	0.2554083
7	7.000000	30.00000	30.42420	-0.4241996
8	8.000000	36.00000	37.95494	-1.954942
9	9.000000	48.00000	48.47404	-0.4740391
10	10.00000	62.00000	61.67508	0.3249168
11	11.00000	78.00000	76.80804	1.191956
12	12.00000	94.00000	92.67921	1.320786
13	13.00000	107.0000	107.6513	-0.6513062
14	14.00000	118.0000	119.6433	-1.643349
15	15.00000	127.0000	126.1308	0.8692398

MODULE NAME (TYPE "HELP" FOR AID)?

SECTION 6

PARAMETRIC HYPOTHESIS TESTING: THE PARH MODULE

The PARH module consists of the programs TSTAT and FSTAT. TSTAT computes the student's *t*-statistic under four different hypotheses. FSTAT computes the *F*-ratio and the corresponding degrees of freedom.

Both programs accept data from the terminal or from one or more data files written in the standard format for STATPAK, described on page 4. FSTAT and TSTAT accept up to 100 observations per variable.

The TSTAT Program

TSTAT calculates the student's *t*-statistic and corresponding degrees of freedom for a set of data. The *t*-test is valid for normally distributed parent populations. TSTAT can test any of the hypotheses below with the *t*-statistic.

Code	Hypothesis
1	The sample observations corresponding to the variable come from a population with a specified mean.
2	The sample observations corresponding to two variables come from populations with the same mean, assuming the populations have equal variances. *NOTE: The validity of this assumption can be checked by computing the* F*-ratio in FSTAT.*
3	The same as hypothesis 2, but assuming the populations have unequal variances.
4	The same as hypothesis 2, but assuming there is no information on the population variances. *NOTE: The two variables must have the same number of observations to test this hypothesis.*

Example 1

The heights, in inches, of ten individuals chosen at random from a normal population are:
63, 63, 66, 67, 68, 69, 70, 70, 71, 71
The user tests the hypothesis that the mean height in this population is 66 inches.

```
-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)



MODULE NAME (TYPE "HELP" FOR AID)? PARH↵

PROGRAM NAME (TYPE "HELP" FOR AID)? TSTAT↵

DO YOU WANT INSTRUCTIONS.?YES↵

CODE 1  TEST MEAN=SPECIFIED VALUE
CODE 2  TEST MEAN A=MEAN B POPULATION VARIANCES EQUAL
CODE 3  TEST MEAN A=MEAN B POPULATION VARIANCES UNEQUAL
CODE 4  TEST MEAN A=MEAN B NO ASSUMPTION ON VARIANCES.

ENTER CODE OF HYPOTHESIS TO BE TESTED.?1↵

DO YOU WANT TO ENTER THE VALUES ON-LINE.?YES↵

VARIABLE NAME.?HGT↵

ENTER NO OF OBSERVATIONS.?10↵

ENTER THE  10 VALUES NOT EXCEEDING 6 IN A ROW
63 63 66 67 68 69↵
70 70 71 71↵

ENTER VALUE OF POPULATION MEAN.?66↵
MEAN OF VARIABLE HGT          67.80000


THE COMPUTED T VALUE IS     1.890378
WITH DEGREES OF FREEDOM    9



MODULE NAME (TYPE "HELP" FOR AID)?
```

The user tests the hypothesis that the population mean is 66 inches.

From *t*-test tables, the user can see that his hypothesis is true at a 5% level of significance.

Example 2

A test population was fed on diet A during a certain period. A random sample of ten members of this population shows the following increase in weight:

10, 6, 16, 17, 13, 12, 8, 14, 5, and 9

Another sample fed on diet B shows the following increase in weight for the same period:

7, 13, 22, 15, 12, 14, 18, 8, 21, and 23

The user tests whether diets A and B significantly affect the increase in weight of the populations, assuming the population variances are equal.

The data for diet A and diet B is stored on file TDATA.

```
-TYPE TDATA↵

10 2
DIETA
10 6 16 17 13 12
8 14 5 9
DIETB
7 13 22 15 12 14
18 8 21 23


-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? PARH↵

PROGRAM NAME (TYPE "HELP" FOR AID)? TSTAT↵

DO YOU WANT INSTRUCTIONS.?NO↵

ENTER CODE OF HYPOTHESIS TO BE TESTED.?2↵

DO YOU WANT TO ENTER THE VALUES ON-LINE.?NO↵

ENTER NAME OF DATA FILE.?TDATA↵

VARIABLE NAME.?DIETA↵

ENTER NAME OF DATA FILE.?TDATA↵

VARIABLE NAME.?DIETB↵
```

The user wishes to test that the population means are equal, assuming equal variances.

```
MEAN OF VARIABLE DIETA        11.00000
MEAN OF VARIABLE DIETB        15.30000


THE COMPUTED T VALUE IS     1.957919
WITH DEGREES OF FREEDOM  18



MODULE NAME (TYPE "HELP" FOR AID)?
```

From *t*-test tables, the user can see that the two samples come from populations with the same mean at a 5% level of significance.

The FSTAT Program

FSTAT uses random samples from two normal populations to test the equality of the variances of those populations. The program computes the *F*-ratio and the corresponding degrees of freedom.

Example

Two random samples taken from two normal populations are shown below:

Sample 1	Sample 2
20	27
16	33
26	42
27	35
23	32
22	34
18	38
24	28
25	41
19	43
	30
	37

The user calls FSTAT to obtain estimates of the variances of the populations and test the hypothesis that the two populations have the same variance.

```
-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? PARH↵

PROGRAM NAME (TYPE "HELP" FOR AID)? FSTAT↵

DO YOU WANT TO ENTER THE VALUES ON-LINE.?YES↵

VARIABLE NAME.?SAMP1↵

ENTER NO OF OBSERVATIONS.?10↵

ENTER THE  10 VALUES NOT EXCEEDING 6 IN A ROW
20 16 26 27 23 22↵
18 24 25 19↵

VARIABLE NAME.?SAMP2↵

ENTER NO OF OBSERVATIONS.?12↵

ENTER THE  12 VALUES NOT EXCEEDING 6 IN A ROW
27 33 42 35 32 34↵
38 28 41 43 30 37↵


VARIANCE OF VARIABLE SAMP1       13.33333


VARIANCE OF VARIABLE SAMP2       28.54545


THE COMPUTED F VALUE IS     2.140909
WITH DEGREES OF FREEDOM  11 AND   9


MODULE NAME (TYPE "HELP" FOR AID)?
```

The calculated *F*-ratio is 2.1409, and the 5% value of *F* with 11 and 9 degrees of freedom is 3.1, according to the *F*-test tables. Thus, the user may conclude that the population variances are equal at a 5% level of significance.

SECTION 7

NON-PARAMETRIC HYPOTHESIS TESTING: THE NPARH MODULE

This module consists of programs UTEST, CRANK, and CCORD. These programs accept data from the terminal or from one or more data files written in the standard format for STATPAK. All three programs accept a maximum of 100 observations per variable.

The UTEST Program

UTEST is used to determine whether two independent samples come from the same population. UTEST prints the Mann-Whitney *U*-statistic and a significance measure of this statistic.

NOTE: When two populations cannot be assumed normal and homogeneous, UTEST can be used as a non-parametric test equivalent to the parametric T-*test.*

Example

The figures below represent half-yearly sales in thousands of dollars for ABC Company and XYZ Company.

ABC Company	XYZ Company
21.2	19.3
23.4	22.5
25.6	26.2
24.8	24.7
27.5	26.2
26.2	27.5
29.3	28.7
29.4	27.8
28.7	29.2
26.9	27.8
27.8	26.4
28.6	25.7
29.2	26.8
29.3	28.7
28.7	29.3
28.8	28.6
	29.1
	29.2
	28.9
	29.2

The user tests the hypothesis that the sales of the two companies are not significantly different.

```
-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)



MODULE NAME (TYPE "HELP" FOR AID)? NPARH↵

PROGRAM NAME (TYPE "HELP" FOR AID)? UTEST↵

DO YOU WANT TO ENTER DATA ON-LINE.?YES↵

ENTER SIZE OF SAMPLE 1 AND SAMPLE 2.?16 20↵

ENTER NAME OF VARIABLE.?ABCCO↵

ENTER THE  16 VALUES OF ABCCO NOT EXCEEDING 6 IN A ROW.
21.2 23.4 25.6 24.8 27.5 26.2↵
29.3 29.4 28.7 26.9 27.8 28.6↵
29.2 29.3 28.7 28.8↵

ENTER NAME OF VARIABLE.?XYZCO↵

ENTER THE  20 VALUES OF XYZCO NOT EXCEEDING 6 IN A ROW.
19.3 22.5 26.2 24.7 26.2 27.5↵
28.7 27.8 29.2 27.8 26.4 25.7↵
26.8 28.7 29.3 28.6 29.1 29.2↵
28.9 29.2↵

MANN-WHITNEY U-STATISTIC IS     151.5000
MEASURE OF SIGNIFICANCE ON U IS  -0.2711977



MODULE NAME (TYPE "HELP" FOR AID)?
```

The measure of significance of the *U*-statistic is -0.2712. This measure is a standard normal deviate. The 5% value of a standard normal deviate table is 1.96. Thus, there is no evidence that the two samples come from different populations.

The CRANK Program

CRANK computes Kendall's rank correlation coefficient to determine the degree of association between two variables. The program also prints a significance measure of the coefficient if there are at least ten observations per variable.

NOTE: CRANK is useful in analyzing correlations when the populations are not normally distributed.

The data for the variables may be ranked or not ranked. The user indicates this by entering a code of 0 for unranked data and 1 for ranked data.

Example

The data below corresponds to the half-yearly sales of Acme Company, which manufactures razors, and Gilt Company, which manufactures blades. The data is stored on file CDATA.

```
-TYPE CDATA↵

16 2
ACME
261 264 260 265 279 285
282 293 302 296 291 302
314 311 324 322
GILT
31 42 45 54 54 64
61 54 61 72 84 85
101 114 107 127
```

The user calls CRANK to determine whether there is a significant relation between the sales of the two companies.

```
-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? NPARH↵

PROGRAM NAME (TYPE "HELP" FOR AID)? CRANK↵

DO YOU WANT TO ENTER DATA ON-LINE.?NO↵

ENTER NAME OF DATA FILE.?CDATA↵

ENTER NAME OF VARIABLE.?ACME↵
```

```
ENTER NAME OF DATA FILE.?CDATA↵

ENTER NAME OF VARIABLE.?GILT↵

ENTER VALUE OF CODE.
CODE = 1 IF DATA IS RANKED.
CODE = 0 IF DATA IS NOT RANKED.
?0↵

KENDALLS RANK CORRELATION COEFF.    0.7830425

MEASURE OF SIGNIFICANCE IS     4.230545



MODULE NAME (TYPE "HELP" FOR AID)?
```

The measure of significance computed is 4.2305. This statistic is a standard normal deviate. The user compares the value with the 5% level of a standard normal deviate, which is 1.96. Thus, there is a significant correlation between the variables.

The CCORD Program

CCORD computes the degree of association among several variables based on a set of observations. The observations may represent rankings of a particular set of data by the variables, or the data may be unranked.

CCORD computes and prints the concordance coefficient, the chi-square statistic to measure the significance of the concordance coefficient, and the degrees of freedom.

NOTE: The chi-square statistic is not computed if there are fewer than eight observations per variable.

Example

Four financial analysts, Adams, Evans, Jones, and Lane, ranked nine different investments as follows:

Investment:	1	2	3	4	5	6	7	8	9
Adams:	2	3	9	4	6	1	7	5	8
Evans:	9	4	1	3	5	7	8	6	2
Jones:	5	4	3	2	9	7	8	6	1
Lane:	6	2	1	8	4	5	7	9	3

This data is stored on file INVST, displayed below.

```
-TYPE INVST⊃

9  4
ADAMS
2  3  9  4  6  1
7  5  8
EVANS
9  4  1  3  5  7
8  6  2
JONES
5  4  3  2  9  7
8  6  1
LANE
6  2  1  8  4  5
7  9  3
```

The user calls CCORD to learn whether there is a significant degree of association among the analysts.

```
-R STATPAK⊃
TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? NPARH⊃
PROGRAM NAME (TYPE "HELP" FOR AID)? CCORD⊃
ENTER NO. OF OBSERVATIONS AND NO. OF VARIABLES.?9 4⊃
DO YOU WANT TO ENTER DATA ON-LINE.?NO⊃
IS THE ENTIRE DATA IN ONE FILE.?YES⊃
ENTER NAME OF DATA FILE.?INVST⊃
ENTER NAME OF VARIABLE.?ADAMS⊃
ENTER NAME OF VARIABLE.?EVANS⊃
ENTER NAME OF VARIABLE.?JONES⊃
ENTER NAME OF VARIABLE.?LANE⊃
```

There are nine observations and four variables.

```
ENTER VALUE OF CODE.
CODE = 1 FOR INPUT DATA RANKED
CODE = 0 FOR INPUT DATA NOT RANKED.
1↵          The user enters 1 to indicate that the data is ranked.

CONCORDANCE COEFFICIENT IS   0.2979167

COMPUTED CHI-SQUARE IS     9.533333
WITH DEGREES OF FREEDOM    8



MODULE NAME (TYPE "HELP" FOR AID)?
```

The computed value of chi-square is 9.5333. The 5% level with eight degrees of freedom from the chi-square tables is 15.51. This is greater than the computed value, so the user concludes that there is no significant association between the analysts.

SECTION 8

TIME SERIES ANALYSIS: THE TIMSA MODULE

The TIMSA module consists of the forecasting programs TRIXP and XPOSE. TRIXP is recommended for short-term forecasting, while XPOSE is recommended for long-term forecasting with seasonal variations. Both programs accept data from a file written in the standard STATPAK format.

The TRIXP Program

This program forecasts future values of a variable based on a set of past observations. The forecast is made as a function of time only, and seasonal factors are not used in the calculations.

TRIXP accepts a maximum of 500 observations from a data file and can forecast for up to 50 periods into the future.

When TRIXP is called, it requests a smoothing coefficient, Q, and smoothing constants, A, B, and C. These numbers are used to smooth the past data. Q is a number between 0 and 1, inclusive. The role of Q in the formulas below demonstrates its effect. The user may enter initial values for the smoothing constants A, B, and C, or he may set them to 0 and let TRIXP generate the initial values.

TRIXP generates the initial values of the smoothing constants as follows:

$$C = x_1 - 2x_2 + x_3$$

$$B = x_2 - x_1 - 1.5C$$

$$A = x_1 - B - 0.5C$$

where x_i $(i = 1,2,3)$ are the first three observations of variable x.

The initial values of the smoothing constants are used to smooth the data for one period ahead. Then the constants A, B, and C are updated, smoothing is done for the next period, and so on.

TRIXP updates the smoothing constants as follows:

$$A_{updated} = x_k + (1 - Q)^3 (S_k - x_k)$$

$$B_{updated} = B_{previous} + C_{previous} - 1.5Q^2 (2 - Q)(S_k - x_k)$$

$$C_{updated} = C_{previous} - Q^3 (S_k - x_k)$$

where x_k is the k^{th} observation of variable x.

S_k is the smoothed value for the k^{th} observation $(S_k = A+B+0.5C)$.

Q is the smoothing constant entered by the user.

The process of updating the smoothing constants is repeated until the entire series of past observations is exhausted. Then the final values for A, B, and C are used in forecasting.

The forecast for t periods into the future is computed as follows:

$$\text{forecast} = A + Bt + Ct^2/2$$

Example

File SALES, below, contains two sets of observations representing the last 18 months of sales of product A and product B.

```
-TYPE SALES↵

18 2
PRODA
419 414 413 412 419 417
422 430 438 441 447 450
454 463 470 472 470 472
PRODB
220 222 220 230 234 248
256 260 252 263 263 264
266 252 254 248 249 260
```

The user calls TRIXP to obtain sales forecasts for product B for the next three months.

```
-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? TIMSA↵

PROGRAM NAME (TYPE "HELP" FOR AID)? TRIXP↵

ENTER NAME OF DATA FILE.?SALES↵

ENTER NAME OF VARIABLE.?PRODB↵

ENTER THE VALUES OF THE SMOOTHING CONSTANTS A,B,C
IF YOU HAVE NO SPECIAL CHOICE ENTER 0.
0 0 0↵          The user allows the program to calculate A, B, and C.

ENTER THE VALUE OF THE SMOOTHING COEFFICIENT.?.5↵

DO YOU WANT THE SMOOTHED SERIES PRINTED OUT.?NO↵
```

```
FORECAST FOR VARIABLE       PRODB
NO. OF PERIODS IN HISTORY   18
SMOOTHING COEFFICIENT          0.5000000

SMOOTHING CONSTANTS(UPDATED)
 A =     258.1592
 B =     5.391922
 C =     1.657501

FOR HOW MANY PERIODS AHEAD YOU NEED FORECASTS.?3

              FORECAST FOR PRODB
              ------------------
          PERIODS AHEAD    FORECAST
          -------------    --------

                1          264.3799
                2          272.2581
                3          281.7937


MODULE NAME (TYPE "HELP" FOR AID)?
```

The XPOSE Program

This program forecasts future values of a variable based either on a set of as many as 100 past observations or on a set of forecasting parameters and smoothing coefficients supplied by the user. XPOSE uses the technique of exponentially weighted moving averages and incorporates linear trends and seasonal factors.[1]

When forecasting parameters are not supplied by the user, XPOSE generates them from past observations. XPOSE uses a user-specified number, H, of the earliest observations to generate starting values of the forecasting parameters. Using these parameters and a set of smoothing coefficients A, B, and C, with values between 0 and 1, the program computes forecasts to be compared with the remaining past observations.

The forecasts are made for one period ahead and compared with the observations for that period. Then XPOSE updates the parameters and makes a forecast for the next period. This process is repeated until the entire series of past observations is exhausted. The final values of the parameters are used to forecast future values of the variable.

XPOSE uses a least squares method to compute the optimal smoothing coefficients. The forecasting technique is improved by minimizing the sum of squares of deviations, that is,

$$\Sigma(\text{observed value} - \text{forecasted value})^2$$

This sum is computed for all possible combinations of (A,B,C) in intervals of .1. The combination (A,B,C) which minimizes the sum of squared deviations is termed the optimal set of smoothing coefficients.

1 - This technique is documented in "Forecasting Sales by Exponentially Weighted Moving Averages" by Peter R. Winters, in *Management Science*, April 1960.

The length of a time period is defined by the user. He enters 4 to indicate quarterly periods, 12 to indicate monthly periods, etc.

The value of H, defined above, is limited as follows: H cannot exceed 60; it must be greater than the number of periods in a year, less than the number of periods in history, and must be a multiple of the number of periods in a year. Thus, if the user specifies that there are 12 periods in a year, H must be 24, 36, 48, or 60.

XPOSE computes and prints the following:

- Past observations with trend and seasonal factors, deseasonalized data, and forecasts (optional).
- Forecasting parameters, including:
  - SO the most recent smoothed and seasonally adjusted average.
  - R the most recent estimate of the trend factor, that is, the rate of increase or decrease.
  - A,B,C the smoothing constants, with values between 0 and 1.
  - F(t) seasonal factors for season t.
- Forecasts of future values.
- The standard error of the forecasts.

Example 1

File HIST, below, contains the sales history of two products for the last 36 months.

```
-TYPE HIST↵

36 2
PRODA
343.9 366.3 355.8 395.4 398.7 531.4
271.3 256.7 301.1 377.8 354.8 376.1
349.6 380.1 364.7 398.3 399.7 546.6
276.2 253.3 334.5 337.6 374.6 390.2
362.0 395.3 392.9 401 429.8 561.4
303.5 270.2 349.2 380.9 419 409.9
PRODB
101 102 103.5 105.4 104 104.5
106 106 107.3 106.2 103 108.2
109 109.2 109.3 109.7 107.6 108.7
109.2 105.6 107 107.2 107.1 108
102.9 104.5 106.2 107 106.5 108
108 108.2 106.5 106.3 107.8 109
```

The user calls XPOSE to forecast the next year's sales for product A, variable PRODA.

```
-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)



MODULE NAME (TYPE "HELP" FOR AID)? TIMSA↵

PROGRAM NAME (TYPE "HELP" FOR AID)? XPOSE↵

 DO YOU HAVE THE PARAMETERS A,B,C,SO,R,F(J)
 READY TO BE USED IN FORECASTING OR DO YOU WANT TO COMPUTE
 THEM ? TYPE YES IF YOU HAVE THE PARAMETERS READY.
NO↵

ENTER THE NAME OF THE FILE.?HIST↵

ENTER NAME OF VARIABLE.?PRODA↵

ENTER THE NAME OF YOUR COMPANY OR DIVISION
 NAME SHOULD NOT EXCEED 15 CHARACTERS.
ACE PRODUCTS↵

NO. OF PERIODS IN HISTORY IS  36
ENTER THE VALUES OF H AND L.
 H = NO OF PERIODS TO BE USED FOR GENERATING INITIAL
     VALUES OF TREND,AVERAGE AND SEASONALS
 L = SEASONS WITHIN A YEAR
     EXAMPLE- 12 FOR MONTHLY DATA,4 FOR QUARTERLY DATA
24 12↵        The first 24 months of data generate initial values for the forecasting parameters.

 DO YOU WANT A PRINT OUT OF THE  PAST SERIES
 WITH SEASONALS,TREND,DESEASONALISED DATA AND
 FORECASTS?. TYPE 1 FOR YES 0 FOR NO
1↵
```

PERIOD T	SEASON J	AVERAGE SO	TREND R	SEASONAL F(J)	ACTUAL S(T)	FOREC BASED ON PREV PER S(T-1,1)
1	1	360.6679	0.4014	0.9549	343.9000	0.0000000
2	2	359.8491	0.1573	1.0207	366.3000	0.0000000
3	3	359.5432	0.0647	0.9906	355.8000	0.0000000
4	4	359.9543	0.1340	1.0976	395.4000	0.0000000
5	5	360.6169	0.2397	1.1043	398.7000	0.0000000
6	6	360.4105	0.1505	1.4759	531.4000	0.0000000
7	7	360.6503	0.1683	0.7521	271.3000	0.0000000
8	8	362.0906	0.4227	0.7070	256.7000	0.0000000
9	9	359.2623	-0.2275	0.8444	301.1000	0.0000000
10	10	364.4666	0.8589	1.0249	377.8000	0.0000000
11	11	363.5589	0.5056	0.9798	354.8000	0.0000000
12	12	363.2893	0.3505	1.0370	376.1000	0.0000000
13	1	364.1371	0.4500	0.9590	349.6000	0.0000000
14	2	366.1461	0.7618	1.0346	380.1000	0.0000000
15	3	367.1574	0.8117	0.9928	364.7000	0.0000000
16	4	366.9498	0.6078	1.0879	398.3000	0.0000000
17	5	366.4347	0.3833	1.0935	399.7000	0.0000000
18	6	367.5246	0.5246	1.4850	546.6000	0.0000000
19	7	367.8867	0.4921	0.7510	276.2000	0.0000000
20	8	366.3605	0.0884	0.6945	253.3000	0.0000000
21	9	372.3869	1.2760	0.8875	334.5000	0.0000000
22	10	364.8087	-0.4948	0.9453	337.6000	0.0000000
23	11	367.9175	0.2259	1.0105	374.6000	0.0000000
24	12	369.7669	0.5506	1.0516	390.2000	0.0000000

25	1	371.7466	0.8364	0.9708	362.0000	355.1474
26	2	374.4800	1.2158	1.0514	395.3000	385.4868
27	3	379.7090	2.0185	1.0263	392.9000	372.9789
28	4	379.1038	1.4937	1.0638	401.0000	415.2713
29	5	383.0888	1.9920	1.1162	429.8000	416.1788
30	6	383.6752	1.7109	1.4676	561.4000	571.8362
31	7	389.1301	2.4597	0.7742	303.5000	289.4405
32	8	391.0819	2.3581	0.6916	270.2000	271.9635
33	9	393.4461	2.3593	0.8875	349.2000	349.1732
34	10	397.2310	2.6444	0.9562	380.9000	374.1617
35	11	402.8308	3.2355	1.0342	419.0000	404.0685
36	12	402.8092	2.5841	1.0244	409.9000	427.0260

DO YOU WANT A PRINT OUT OF THE PARAMETERS
CALCULATED FOR YOUR DATA? TYPE YES OR NO
YES

```
SO =     402.8092
 R =       2.584084
 A =         0.2000000
 B =         0.8000000
 C =         0.2000000

 SEASONAL FACTORS
-------------------

 F( 1) =   0.9708323
 F( 2) =    1.051404
 F( 3) =    1.026345
 F( 4) =    1.063781
 F( 5) =    1.116244
 F( 6) =    1.467569
 F( 7) =   0.7741639
 F( 8) =   0.6916253
 F( 9) =   0.8875313
 F(10) =   0.9561737
 F(11) =    1.034208
 F(12) =    1.024406
```

```
FOR HOW MANY PERIODS AHEAD YOU NEED FORECASTS?
12↵


NEED FORECASTS ON TTY, DSK OR BOTH?BOTH↵

OUTPUT FILENAME?FOREC↵

               SALES FORECASTS ACE PRODUCTS
-------------------------------------------------------------------------

PERIODS AHEAD      FORECAST
---------------------------

        1            393.5689
        2            428.9492
        3            421.3779
        4            439.4964
        5            464.0557
        6            613.9041
        7            325.8439
        8            292.8908
        9            378.1469
       10            409.8639
       11            445.9859
       12            444.4060



STANDARD ERROR OF FORECAST =     12.79439

ENTER NAME OF VARIABLE.?STOP↵


MODULE NAME (TYPE "HELP" FOR AID)?
```

Example 2

In the previous example, the user obtained an actual sales figure for product A for the 37th month. He now wishes to update the previous forecast by incorporating this figure (383.1). He enters the parameters computed in Example 1.

```
-R STATPAK⏎

TYMSHARE PDP-10 STATPAK VERSION 3.02 31-JAN-72



MODULE NAME (TYPE "HELP" FOR AID)? TIMSA⏎

PROGRAM NAME (TYPE "HELP" FOR AID)? XPOSE⏎

 DO YOU HAVE THE PARAMETERS A,B,C,SO,R,F(J)
 READY TO BE USED IN FORECASTING OR DO YOU WANT TO COMPUTE
 THEM ? TYPE YES IF YOU HAVE THE PARAMETERS READY.
YES⏎

ENTER THE NAME OF YOUR COMPANY OR DIVISION
 NAME SHOULD NOT EXCEED 15 CHARACTERS.
ACE PRODUCTS⏎

ENTER NAME OF VARIABLE.?PRODA⏎

HOW MANY SEASONS YOU HAVE IN A YEAR?
12⏎

ENTER THE VALUES OF A,B,C,SO AND R
.2 .8 .2 402.8092 2.584084⏎

ENTER THE SEASONAL FACTORS , 5 IN A ROW
.9708323 1.051404 1.026345 1.063781 1.116244⏎
1.467569 .7741639 .6916253 .8875313 .9561737⏎
1.034208 1.024406⏎

WHICH SEASON ARE YOU IN RIGHT NOW?.
1⏎

ENTER THE ACTUAL SALES FOR THIS SEASON
383.1⏎

DO YOU WANT A PRINT OUT OF THE PARAMETERS
CALCULATED FOR YOUR DATA? TYPE YES OR NO
YES⏎

SO =     403.2366
 R =         2.152748
 A =           0.2000000
 B =           0.8000000
 C =           0.2000000
```

The user enters the values computed in Example 1.

This is the first season of the forecast in Example 1.

```
 SEASONAL FACTORS
-------------------

 F( 1) =   0.9542165
 F( 2) =    1.051404
 F( 3) =    1.026345
 F( 4) =    1.063781
 F( 5) =    1.116244
 F( 6) =    1.467569
 F( 7) =   0.7741639
 F( 8) =   0.6916253
 F( 9) =   0.8875313
 F(10) =   0.9561737
 F(11) =    1.034208
 F(12) =    1.024406

FOR HOW MANY PERIODS AHEAD YOU NEED FORECASTS?
12↵


NEED FORECASTS ON TTY,DSK OR BOTH?BOTH↵

OUTPUT FILENAME?NEW4↵

               SALES FORECASTS ACE PRODUCTS
-----------------------------------------------------------------------

PERIODS AHEAD      FORECAST
----------------------------

        1             426.2280
        2             418.2788
        3             435.8256
        4             459.7224
        5             607.5741
        6             322.1707
        7             289.3109
        8             373.1702
        9             404.0898
       10             439.2944
       11             437.3362
       12             409.4253



 DO YOU HAVE THE PARAMETERS A,B,C,SO,R,F(J)
 READY TO BE USED IN FORECASTING OR DO YOU WANT TO COMPUTE
 THEM ? TYPE YES IF YOU HAVE THE PARAMETERS READY.
STOP↵


MODULE NAME (TYPE "HELP" FOR AID)?
```

Notice that the additional data has caused a slight change in the forecasts from Example 1.

SECTION 9

DISCRIMINANT ANALYSIS: THE DISCA MODULE

The DISCA module consists of one program, DANSS, which performs a discriminant analysis and generates linear discriminant functions. The program accepts data from a file written in the standard format for STATPAK.

DANSS tests a set of observations to find whether the populations represented in the data vary significantly from each other. The data used by DANSS is grouped by samples from different populations.

The program performs a discriminant analysis of as many as ten groups with a maximum of ten variables each. The total number of observations in the analysis must not exceed 3000. The number of variables must be greater than the number of groups in the analysis.

DANSS computes and prints the following statistics:

- The means of the variables in each group.
- A pooled dispersion matrix (optional).
- The common means of each variable. These are the means of the variables considering all observations from all groups.
- The Mahalanobis *D*-square statistic. This statistic is used to determine whether the means of the variables differ significantly in each of the groups. Assuming normal populations, the user can compare the *D*-square statistic to the chi-square value with m(k–1) degrees of freedom, where m is the number of variables and k is the number of groups.
- The coefficients in each discriminant function.
- Evaluation of each observation on the basis of the discriminant functions developed (optional).

Example

The data below corresponds to four groups of office buildings. The first group consists of eight buildings, the second and third consist of seven each, and the fourth group has eight buildings. Measurements on five different characteristics have been made on each of these buildings.

	Building	BATHS	HGT	LOC	LSCP	OFSPC
Group 1	1	8	100	10	3	24
	2	8	80	12	4	22
	3	8	75	3	9	9
	4	2	17	2	16	7
	5	8	90	10	5	23
	6	8	95	3	17	6
	7	8	85	10	2	29
	8	8	105	10	7	28
Group 2	9	8	130	10	9	28
	10	9	115	7	11	8
	11	8	120	10	8	27
	12	14	275	6	1	14
	13	6	50	8	7	18
	14	2	20	9	7	19
	15	8	90	10	7	27
Group 3	16	15	250	11	3	20
	17	7	90	4	9	9
	18	7	75	13	4	21
	19	16	300	5	8	16
	20	5	100	9	6	23
	21	8	115	10	8	27
	22	7	100	3	17	6
Group 4	23	8	115	10	3	23
	24	8	100	12	4	23
	25	8	100	3	9	21
	26	2	35	2	15	7
	27	8	90	10	9	27
	28	8	95	9	8	26
	29	9	130	8	7	18
	30	8	140	10	7	26

These observations are stored on file BLDG, below.

```
-TYPE BLDG↵

30 5
BATHS
8 8 8 2 8 8
8 8 8 9 8 14
6 2 8 15 7 7
16 5 8 7 8 8
8 2 8 8 9 8
HGT
100 80 75 17 90 95
85 105 130 115 120 275
50 20 90 250 90 75
300 100 115 100 115 100
100 35 90 95 130 140
LOC
10 12 3 2 10 3
10 10 10 7 10 6
8 9 10 11 4 13
5 9 10 3 10 12
3 2 10 9 8 10
LSCP
3 4 9 16 5 17
2 7 9 11 8 1
7 7 7 3 9 4
8 6 8 17 3 4
9 15 9 8 7 7
OFSPC
24 22 9 7 23 6
29 28 28 8 27 14
18 19 27 20 9 21
16 23 27 6 23 23
21 7 27 26 18 26
```

The DANSS program tests whether the groups vary significantly from each other.

```
-R STATPAK↵
TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? DISCA↵
PROGRAM NAME (TYPE "HELP" FOR AID)? DANSS↵
ENTER NAME OF DATA FILE.?BLDG↵
ENTER NO OF VARIABLES IN ANALYSIS.?5↵
ENTER NO OF GROUPS.?4↵
NUMBER IN GROUP 1?8↵
NUMBER IN GROUP 2?7↵
NUMBER IN GROUP 3?7↵
NUMBER IN GROUP 4?8↵
ENTER NAME OF VARIABLE  1.?BATHS↵
ENTER NAME OF VARIABLE  2.?HGT↵
ENTER NAME OF VARIABLE  3.?LOC↵
ENTER NAME OF VARIABLE  4.?LSCP↵
ENTER NAME OF VARIABLE  5.?OFSPC↵
PRINT POOLED DISPERSION MATRIX.?YES↵
EVALUATE OBSERVATIONS ON THE BASIS OF
DISCRIMINANT FUNCTIONS DEVELOPED.?YES↵
```

GROUP 1 MEANS
BATHS 7.250000
HGT 80.87500
LOC 7.500000
LSCP 7.875000
OFSPC 18.50000

GROUP 2 MEANS
BATHS 7.857143
HGT 114.2857
LOC 8.571429
LSCP 7.142857
OFSPC 20.14286

GROUP 3 MEANS
BATHS 9.285714
HGT 147.1429
LOC 7.857143
LSCP 7.857143
OFSPC 17.42857

GROUP 4 MEANS
BATHS 7.375000
HGT 100.6250
LOC 8.000000
LSCP 7.750000
OFSPC 21.37500

POOLED DISPERSION MATRIX

ROW BATHS				
9.833104	181.7837	1.917582	-6.098901	4.621566
ROW HGT				
181.7837	3848.040	13.32692	-104.0007	56.80426
ROW LOC				
1.917582	13.32692	11.94505	-11.16209	22.60989
ROW LSCP				
-6.098901	-104.0007	-11.16209	19.61882	-22.74863
ROW OFSPC				
4.621566	56.80426	22.60989	-22.74863	62.78640

```
COMMON MEANS
BATHS      7.900000
HGT        109.4000
LOC        7.966667
LSCP       7.666667
OFSPC      19.40000

GENERALIZED MAHALANOBIS D-SQUARE    15.67007


NEED DISCRIMINANT FUNCTION COEFFICIENTS?.
TYPE TTY,DSK,BOTH OR NONE.
TTY↵          The coefficients are printed only on the terminal.


DISCRIMINANT FUNCTION  1
CONSTANT              -27.26358
COEFF. OF BATHS         2.588296
COEFF. OF HGT       -0.4746080E-01
COEFF. OF LOC           1.851615
COEFF. OF LSCP          2.423395
COEFF. OF OFSPC       0.3583278

DISCRIMINANT FUNCTION  2
CONSTANT              -28.85336
COEFF. OF BATHS         1.596163
COEFF. OF HGT       0.1119795E-01
COEFF. OF LOC           2.262848
COEFF. OF LSCP          2.562995
COEFF. OF OFSPC       0.3069434

DISCRIMINANT FUNCTION  3
CONSTANT              -31.16577
COEFF. OF BATHS         1.531905
COEFF. OF HGT       0.2630825E-01
COEFF. OF LOC           2.451687
COEFF. OF LSCP          2.674445
COEFF. OF OFSPC       0.2271499

DISCRIMINANT FUNCTION  4
CONSTANT              -28.79324
COEFF. OF BATHS         1.851321
COEFF. OF HGT       -0.5632087E-02
COEFF. OF LOC           1.933716
COEFF. OF LSCP          2.545502
COEFF. OF OFSPC       0.4351975
```

EVALUATION OF CLASSIFICATION FUNCTIONS FOR EACH OBSERVATION

GROUP 1

OBSERVATION	PROBABILITY ASSOCIATED WITH LARGEST DISCRIMINANT FUNCTION	LARGEST FUNCTION NO.
1	0.3623551	1
2	0.4260625	1
3	0.8548857	1
4	0.3428500	4
5	0.4170793	1
6	0.3367720	1
7	0.5585431	1
8	0.4311859	4

GROUP 2

OBSERVATION	PROBABILITY ASSOCIATED WITH LARGEST DISCRIMINANT FUNCTION	LARGEST FUNCTION NO.
1	0.3520473	4
2	0.4983430	3
3	0.3664007	4
4	0.6785006	3
5	0.5690117	1
6	0.4460260	2
7	0.3777263	4

GROUP 3

OBSERVATION	PROBABILITY ASSOCIATED WITH LARGEST DISCRIMINANT FUNCTION	LARGEST FUNCTION NO.
1	0.7327295	3
2	0.4551894	1
3	0.3981236	2
4	0.7815912	3
5	0.4440284	2
6	0.3814765	4
7	0.4816400	3

GROUP 4

OBSERVATION	PROBABILITY ASSOCIATED WITH LARGEST DISCRIMINANT FUNCTION	LARGEST FUNCTION NO.
1	0.3498248	2
2	0.3720276	2
3	0.5306962	1
4	0.3366389	2
5	0.3955042	4
6	0.4084113	4
7	0.3006484	2
8	0.3713639	2

MODULE NAME (TYPE "HELP" FOR AID)?

SECTION 10

VARIANCE AND FACTOR ANALYSIS: THE AVANC MODULE

The AVANC module performs an analysis of variance in the program ANVAR, and a factor analysis in the program FCTOR. FCTOR accepts data from a file written in the standard format for STATPAK. ANVAR accepts data from a file written in the format described below.

The ANVAR Program

ANVAR analyzes the variance of a set of observations and separates the total variance into components due to different factors. These components are compared by the *F*-test for any significant effect due to individual factors or interactions of factors.

ANVAR accepts data from a file written in the following form:

Line Number	Contents	Format
1	Number of factors, between 2 and 6, inclusive	Integer
2	Factor name, Factor level (greater than or equal to 2)	Five alphanumeric characters, Integer
3	Factor name, Factor level	Alphanumeric, Integer
⋮		
k	Factor name, Factor level	Alphanumeric, Integer
k+1	Up to six observations	Free form
k+2	Up to six observations	Free form
⋮		
k+N	Up to six observations	Free form

NOTE: The total number of observations may not exceed 4000.

The order of observations on the data file is such that the first factor varies most rapidly, then the second, and so on.

For example, the data file, DATFL, below, contains the data for the two factors A and B. A is a four-level factor and B is a three-level factor.

	A_1	A_2	A_3	A_4
B_1	23	40	21	17
B_2	38	19	25	28
B_3	20	26	33	29

The form of DATFL is:

```
2
A 4
B 3
23 40 21 17 38 19
25 28 20 26 33 29
```

ANVAR computes and prints the following statistics:

- The grand mean of all observations.
- The sum of squares, degrees of freedom, and mean squares for each factor and combination of factors.
- An analysis of variance table for each source of variation. The user may specify that one or more sources be pooled under the single source, ERROR. The *F*-ratio is computed between each explicit source and ERROR by the formula:

$$F\text{-ratio} = \frac{\text{mean square due to the source}}{\text{mean square due to ERROR}}$$

Example 1

Ten varieties of wheat were grown on three plots of land each. The following yields of bushels were obtained:

Variety:	1	2	3	4	5	6	7	8	9	10
Plot 1:	7	7	14	11	9	6	9	8	12	9
Plot 2:	8	9	13	10	9	7	13	13	11	11
Plot 3:	7	6	16	11	12	5	12	11	11	11

This data is stored on file PLOTS, below.

```
-TYPE PLOTS↵

2
VARIE 10
PLOTS 3
7 7 14 11 9 6
7 8 12 9 8 9
13 10 9 7 13 13
11 11 7 6 16 11
12 5 12 11 11 11
```

The user tests the significance of the difference between varieties using ANVAR.

```
-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? AVANC↵

PROGRAM NAME (TYPE "HELP" FOR AID)? ANVAR↵

ENTER NAME OF DATA FILE.?PLOTS↵


GRAND MEAN    9.866667


SOURCE OF                SUMS OF        DEGREES OF        MEAN
VARIATION                SQUARES        FREEDOM           SQUARES

 1 VARIE                 156.1333            9           17.34815
 2 PLOTS                 11.46667            2           5.733333
 3 VARIE PLOTS           43.86667           18           2.437037
   TOTAL                 211.4667           29


HOW MANY SOURCES YOU WANT TO POOL UNDER ERROR.?2↵

ENTER THE INDICES OF THE  2 SOURCES ,NOT EXCEEDING 8 IN A ROW.
2 3↵

SOURCE OF                SUMS OF        DEGREES OF            F
VARIATION                SQUARES        FREEDOM           RATIO

VARIE                    156.1333            9         6.270415
ERROR                    55.33333           20
TOTAL                    211.4667           29


MODULE NAME (TYPE "HELP" FOR AID)?
```

In this problem, VARIE is the only factor of interest. PLOTS is treated as a dummy second factor. In the analysis of variance table, the factor PLOTS and interaction factor VARIE PLOTS are pooled for the purpose of significance tests.

The *F*-ratio corresponding to VARIE is 6.2704. The user finds from *F*-test tables that this value of *F* for 9 and 20 degrees of freedom is highly significant at the 5% level.

Example 2

The data below represents yields of wheat in bushels. Four different varieties of wheat were tried in combination with three different fertilizers and three different pesticides. The experiment was repeated twice. Thus, the problem calls for a four-factor analysis, where the factors are fertilizer, variety, pesticide, and trial.

		F_1				F_2				F_3			
		V_1	V_2	V_3	V_4	V_1	V_2	V_3	V_4	V_1	V_2	V_3	V_4
Trial 1	P_1	3	10	9	8	24	8	9	3	2	8	9	8
	P_2	4	12	3	9	22	7	16	2	2	2	7	2
	P_3	5	10	5	8	23	9	17	3	2	8	6	3
Trial 2	P_1	2	14	9	13	29	16	11	3	2	7	5	3
	P_2	7	11	5	8	28	18	10	6	6	6	5	9
	P_3	9	10	27	8	28	16	11	7	8	9	8	15

where F represents fertilizer.
V represents variety.
P represents pesticide.

The data is stored on file WHEAT.

```
-TYPE WHEAT↵

4
VARIE 4  (variety)
FERTI 3  (fertilizer)
PESTI 3  (pesticide)
TRIAL 2  (trial)
3 10 9 8 24 8
9 3 2 8 9 8
4 12 3 9 22 7
16 2 2 2 7 2
5 10 5 8 23 9
17 3 2 8 6 3
2 14 9 13 29 16
11 3 2 7 5 3
7 11 5 8 28 18
10 6 6 6 5 9
9 10 27 8 28 16
11 7 8 9 8 15
```

The user tests for the significance of individual factors and the interaction of factors using ANVAR.

```
-R STATPAK↵

TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? AVANC↵

PROGRAM NAME (TYPE "HELP" FOR AID)? ANVAR↵

ENTER NAME OF DATA FILE.?WHEAT↵


GRAND MEAN    9.402778
```

SOURCE OF VARIATION	SUMS OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES
1 VARIE	229.0417	3	76.34722
2 FERTI	722.6944	2	361.3472
3 VARIE FERTI	1382.083	6	230.3472
4 PESTI	55.11111	2	27.55556
5 VARIE PESTI	42.00000	6	7.000000
6 FERTI PESTI	13.13889	4	3.284722
7 VARIE FERTI PESTI	140.7500	12	11.72917
8 TRIAL	141.6806	1	141.6806
9 VARIE TRIAL	18.81944	3	6.273148
10 FERTI TRIAL	6.027778	2	3.013889
11 VARIE FERTI TRIAL	176.9722	6	29.49537
12 PESTI TRIAL	40.77778	2	20.38889
13 VARIE PESTI TRIAL	50.55555	6	8.425926
14 FERTI PESTI TRIAL	62.63889	4	15.65972
15 VARIE FERTI PESTI TRIAL	151.0278	12	12.58565
TOTAL	3233.319	71	

```
HOW MANY SOURCES YOU WANT TO POOL UNDER ERROR.?7↵

ENTER THE INDICES OF THE  7 SOURCES ,NOT EXCEEDING 8 IN A ROW.
9 10 11 12 13 14 15↵
```

SOURCE OF VARIATION	SUMS OF SQUARES	DEGREES OF FREEDOM	F RATIO
VARIE	229.0417	3	5.272396
FERTI	722.6944	2	24.95396
VARIE FERTI	1382.083	6	15.90735
PESTI	55.11111	2	1.902935
VARIE PESTI	42.00000	6	0.4834069
FERTI PESTI	13.13889	4	0.2268368
VARIE FERTI PESTI	140.7500	12	0.8099942
TRIAL	141.6806	1	9.784193
ERROR	506.8194	35	
TOTAL	3233.319	71	

```
MODULE NAME (TYPE "HELP" FOR AID)?
```

The FCTOR Program

This program performs a factor analysis of a group of variables. FCTOR determines the minimum number of factors required to explain most of the variation in the variables, and computes their values.

FCTOR accepts a maximum of 40 variables and 3000 observations per variable.

FCTOR computes and prints the following statistics:

- Means and standard deviations for each variable.
- Correlation matrix (optional).
- Eigenvalues and eigenvectors.
- Initial factor matrix (optional).
- Rotated factor matrix.
- Check on communalities.

Example

The user performs a factor analysis on the data below with six variables and 23 observations per variable. The data is stored on file FACTR.

```
-TYPE FACTR↵

23 6
VAR1
7 13 9 7 6 10
7 16 9 8 8 9
11 9 10 11 16 9
7 8 6 10 8
VAR2
7 18 18 13 8 12
6 19 22 15 10 12
17 16 15 11 9 8
18 11 6 9 10
VAR3
9 25 24 25 20 30
11 25 26 26 20 28
21 26 24 30 16 19
22 23 28 26 26
VAR4
7 15 23 36 7 11
7 16 24 30 8 11
30 27 18 19 20 14
9 18 23 26 15
VAR5
15 13 12 11 15 10
15 13 13 13 17 8
10 14 12 19 18 16
15 9 7 10 11
VAR6
36 35 43 12 46 42
35 30 40 10 40 45
45 31 29 26 31 33
37 36 40 37 42
```

```
-R STATPAK
TYMSHARE PDP-10 STATPAK VERSION 3.01 (11/9/71)


MODULE NAME (TYPE "HELP" FOR AID)? AVANC
PROGRAM NAME (TYPE "HELP" FOR AID)? FCTOR
NAME OF DATA FILE?FACTR
ENTER NAME OF VARIABLE IN ANALYSIS?VAR1
ENTER NAME OF VARIABLE IN ANALYSIS?VAR2
ENTER NAME OF VARIABLE IN ANALYSIS?VAR3
ENTER NAME OF VARIABLE IN ANALYSIS?VAR4
ENTER NAME OF VARIABLE IN ANALYSIS?VAR5
ENTER NAME OF VARIABLE IN ANALYSIS?VAR6
ENTER NAME OF VARIABLE IN ANALYSIS?STOP
```

The user types STOP to stop entering values.

```
ENTER VALUE OF CONSTANT TO LIMIT EIGENVALUES.?1
```

The user allows no eigenvalues less than 1.

```
MEANS
VAR1        9.304348
VAR2        12.60870
VAR3        23.04348
VAR4        18.00000
VAR5        12.86957
VAR6        34.82609

STANDARD DEVIATIONS
VAR1        2.704118
VAR2        4.599794
VAR3        5.372305
VAR4        8.333939
VAR5        3.137810
VAR6        9.291502
```

DO YOU WANT THE CORRELATION MATRIX PRINTED.?YES

CORRELATION COEFFICIENTS

ROW VAR1
1.000000 0.3498669 0.1085591 0.1210187 0.2191729

-0.9548972E-01

ROW VAR2
0.3498669 1.000000 0.3980321 0.3557213 -0.8242919E
-01
-0.9100186E-01

ROW VAR3
0.1085591 0.3980321 1.000000 0.4172613 -0.4445602

-0.7815379E-01

ROW VAR4
0.1210187 0.3557213 0.4172613 1.000000 -0.3128765

-0.5036494

ROW VAR5
0.2191729 -0.8242919E-01 -0.4445602 -0.3128765 1.000000

-0.2299963

ROW VAR6
-0.9548972E-01 -0.9100186E-01 -0.7815379E-01 -0.5036494 -0.2299963

1.000000

EIGENVALUES
2.143067 1.459717 1.124474

CUMULATIVE PERCENTAGE OF EIGENVALUES
0.3571779 0.6004640 0.7878763

EIGENVECTORS

VECTOR 1
0.2174841 0.4629068 0.5167775 0.5572654 -0.2893265

-0.2776769

VECTOR 2
0.4772406 0.1444153 -0.2745525 0.5696335E-01 0.6671112

-0.4772047

VECTOR 3
0.5471254 0.4276067 0.1244227 -0.3650158 0.3236728E
-01
0.6066640

```
DO YOU WANT THE INITIAL FACTOR MATRIX PRINTED.?YES↵


FACTOR MATRIX (  3 FACTORS)

VARIABLE VAR1
  0.3183797       0.5765961       0.5801783

VARIABLE VAR2
  0.6776595       0.1744808       0.4534393

VARIABLE VAR3
  0.7565219      -0.3317109       0.1319393

VARIABLE VAR4
  0.8157930       0.6882240E-01  -0.3870672

VARIABLE VAR5
 -0.4235514       0.8059953       0.3432265E-01

VARIABLE VAR6
 -0.4064972      -0.5765527       0.6433138


ITERATION        VARIANCES
  CYCLE
    0           0.1928336
    1           0.3248272
    2           0.4338407
    3           0.4342646
    4           0.4342651
    5           0.4342651
    6           0.4342651
    7           0.4342651
    8           0.4342651


NEED ROTATED FACTOR MATRIX?
TYPE TTY, DSK, BOTH OR NONE.
BOTH↵          The rotated matrix is printed on the terminal and saved on a file.

OUTPUT FILENAME.?ROTAT↵        The file ROTAT.DAT contains the rotated matrix.
```

```
ROTATED FACTOR MATRIX (  3 FACTORS)
```

Three distinct factors explain most of the variability in this group of six variables.

```
VARIABLE VAR1
  0.2992361E-01   0.1922221       0.8559149

VARIABLE VAR2
  0.1305635      -0.3419365       0.7492030

VARIABLE VAR3
  0.1675858      -0.7579546       0.3117448

VARIABLE VAR4
  0.7478828      -0.4816886       0.1694775

VARIABLE VAR5
  0.1142076       0.8799314       0.2070729

VARIABLE VAR6
 -0.9373980      -0.1784606      -0.3068892E-01


CHECK ON COMMUNALITIES

VARIABLE        ORIGINAL            FINAL           DIFFERENCE
VAR1          0.7704356         0.7704350          0.5885959E-06
VAR2          0.6952731         0.6952725          0.5662441E-06
VAR3          0.6997654         0.6997649          0.4693866E-06
VAR4          0.8200758         0.8200751          0.6556511E-06
VAR5          0.8302023         0.8302018          0.4842877E-06
VAR6          0.9115057         0.9115050          0.7078052E-06


MODULE NAME (TYPE "HELP" FOR AID)? STOP↵


EXECUTION TIME:          3 MIN. 52.85 SEC.
TOTAL ELAPSED TIME:      7 MIN. 31.58 SEC.
NO EXECUTION ERRORS DETECTED

EXIT

-TYPE ROTAT.DAT↵
```

The retained data is printed.

```
               3                3     ROTATED FACTOR MATRIX
VAR1
  0.2992361E-01   0.1922221       0.8559149
VAR2
  0.1305635      -0.3419365       0.7492030
VAR3
  0.1675858      -0.7579546       0.3117448
VAR4
  0.7478828      -0.4816886       0.1694775
VAR5
  0.1142076       0.8799314       0.2070729
VAR6
 -0.9373980      -0.1784606      -0.3068892E-01
```

APPENDIX

ERROR MESSAGES

Program	Message and Description
General	**NO SUCH VARIABLE. TRY AGAIN.** The user tries to retrieve data for a nonexistent variable.
	VARIABLE ALREADY SPECIFIED. TRY AGAIN. The user tries to enter a variable name twice in the same analysis.
DSCRE	**ILLEGAL CONDITION CODE. TRY AGAIN.** The user attempts to use a condition code other than GT, GE, LT, LE, NE, and EQ.
	NO. OF FREQ. CLASSES EXCEEDS 12. TRY AGAIN. The user requests more than 12 frequency classes.
DREGR	**VARIABLE ALREADY IN REGRESSION. TRY AGAIN.** The user tries to enter a variable into the same regression analysis twice.
	THE MATRIX IS SINGULAR. THIS SELECTION IS SKIPPED. The matrix of cross products becomes singular.
	TOO FEW OBSERVATIONS. REGRESSION TERMINATED. The number of observations per variable is less than the number of variables plus 2.
RGSTP	**CODE IS OTHER THAN 0 OR 1 OR 2. TRY AGAIN.** The user specifies an illegal code for an independent variable.
	TOO FEW OBSERVATIONS. ANALYSIS TERMINATED. The number of observations per variable is less than the number of variables plus 3.
	EITHER THE MATRIX IS SINGULAR OR SUM OF SQUARES NEGATIVE. PROBLEM IGNORED. The problem contains illegal conditions.
RGPOL	**DEGREE EXCEEDS 10. TRY AGAIN.** The user enters a degree higher than ten.
	NO. OF OBS. IS NOT GREATER THAN DEGREE+1. The number of observations is less than the polynomial's degree plus 2.

Program	Message and Description
TSTAT	**ILLEGAL CODE. TRY AGAIN.** The user enters a hypothesis code other than 1, 2, 3, or 4.
	FOR HYPOTHESIS WITH CODE 4, THE NUMBER OF OBSERVATIONS FOR THE TWO VARIABLES MUST BE EQUAL. EXECUTION ABORTED. The user must start over.
TSTAT FSTAT	**MORE THAN 100 OBSERVATIONS PER VARIABLE. EXECUTION ABORTED.** The user tries to enter more than 100 observations for a variable.
UTEST CRANK	**YOU HAVE OVERLOOKED ONE OF THE FOLLOWING: SAMPLE SIZE 1 EXCEEDS SAMPLE SIZE 2, SAMPLE SIZE EXCEEDS 100. EXECUTION ABORTED.** The user makes an error in the sample size.
CRANK CCORD	**ILLEGAL CODE. TRY AGAIN.** The user enters a rank code other than 0 or 1.
CCORD	**THE NUMBER OF OBSERVATIONS IS DIFFERENT FROM THE NUMBER IN DATA FILE. EXECUTION ABORTED.** The user specifies a number of observations different from the number on the data file.
	YOU HAVE OVERLOOKED ONE OF THE FOLLOWING: NUMBER OF VARIABLES OR SOURCES EXCEEDS 20, NUMBER OF OBSERVATIONS PER VARIABLE EXCEEDS 100. EXECUTION ABORTED. The user enters too many variables or observations.
TRIXP	**SMOOTHING COEFFICIENT MUST HAVE A VALUE BETWEEN 0 AND 1. TRY AGAIN.** The user enters an illegal value for the smoothing coefficient.
XPOSE	**MORE THAN 100 PERIODS IN HISTORY. EXECUTION ABORTED.** The user enters more than 100 observations.
	H CANNOT EXCEED 60 AND L CANNOT EXCEED 12. RETYPE VALUES OF H AND L. The user specifies too many periods for generating the forecasting parameters, or designates more than 12 seasons in a year.

Program	Message and Description
	YOU HAVE OVERLOOKED ONE OF THE FOLLOWING: NUMBER OF PERIODS IN HISTORY MUST BE GREATER THAN THE NUMBER OF PERIODS USED FOR GENERATING INITIAL VALUES, NUMBER OF PERIODS USED FOR GENERATING INITIAL VALUES MUST BE GREATER THAN AND A MULTIPLE OF NUMBER OF SEASONS. RETYPE VALUES OF H AND L. The restrictions on H and L are described on page 56.
DANSS	**YOU HAVE OVERLOOKED ONE OF THE FOLLOWING:** **NO. OF GROUPS EXCEEDS 10** **NO. OF VARIABLES EXCEEDS 10** **TOTAL NO. OF OBSERVATIONS EXCEEDS 3000** **EXECUTION ABORTED** The user enters too many variables or groups.
ANVAR	**YOU HAVE OVERLOOKED ONE OF THE FOLLOWING:** **NUMBER OF FACTORS EXCEEDS 6** **NUMBER OF LEVELS FOR A FACTOR IS LESS THAN 2** **TOTAL NUMBER OF OBSERVATIONS EXCEEDS 4000** **EXECUTION ABORTED.** The user overlooks one of the restrictions on input data. *NOTE: The total number of observations is the product of the levels of all factors.*
FCTOR	**MORE THAN 40 VARIABLES OR MORE THAN 3000 OBSERVATIONS. EXECUTION TERMINATED.** The restrictions on variables and/or observations are exceeded.
	THERE IS NO EIGENVALUE GREATER THAN THE CONSTANT SPECIFIED. EXECUTION ABORTED. The user's constant to limit eigenvalues eliminates all the eigenvalues.