STATPAK

TYMSHARE TYMCOM-IX MANUALS

Preliminary Manual

SEPTEMBER 1972

NOTE

This manual contains selected material (sections 1 through 10) of the forthcoming STATPAK Manual.

Please notify your Tymshare representative of errors it contains or of suggestions for its improvement.

TYMSHARE®

CONTENTS

	Page
SECTION 1 – INTRODUCTION	1
Symbol Conventions	3
Editing Characters	4
STATPAK Analyses	5
Data Manipulation Commands	8
Commands for Entering and Saving Data	8
Commands for Examining Data	8
Commands for Ordering Data	9
Commands for Modifying and Adding Data	9
Commands for Transforming Data	9
Utility Commands	10
Sample Analysis	11
SECTION 2 – STATPAK DATA FILES	13
Matrix Component Specification	14
Entering and Saving Data	16
The LINE Command	17
The INPUT Command	18
The SAVE Command	21
The LOAD Command	22
Examining the Data	24
The LIST Command	24
The FAST Command	25
The SIZE Command	27
The ROW and COLS Commands	27

Page

Ordering and Ranking the Data 28

The RANK Command 28

The ORDER Command 31

Modifying and Adding Data 33

The DELETE Command 33

The RENAME Command 34

The DUPLICATE Command 34

The NUMBER Command 37

The CHANGE Command 37

Transforming the Data 41

Expressions 41

The APPEND Command 44

The INSERT Command 48

The REPLACE Command 49

SECTION 3 – ELEMENTARY ANALYSES 52

Elementary Statistics 52

Descriptive Statistics 54

SECTION 4 – PLOTTING 57

Scatter Diagrams 57

Plots 60

Histograms 64

Cumulative Histograms 66

SECTION 5 – CORRELATION ANALYSES 68

Correlation 68

Spearman and Kendall Rank Correlations 70

Contingency Table 73

	Page
SECTION 6 – REGRESSION	75
Linear Regression	76
Multiple Regression	79
Stepwise Regression	82
Polynomial Regression	86
Curve Fitting	88
SECTION 7 – GOODNESS-OF-FIT	91
Data Input	91
Chi-Square Test	94
Kolmogorov-Smirnov Test	98
SECTION 8 – CONFIDENCE LIMITS	100
SECTION 9 – THE XPOS TIME SERIES ANALYSIS	103
SECTION 10 – DATA GENERATION	113
Step Function	114
Linear Function of a Single Variable	116
Linear Function of Several Variables	118
Polynomial Function	120
The XPOS Forecast Function	122
User-Defined Function	124

SECTION 1
INTRODUCTION

STATPAK, Tymshare's statistical package on the TYMCOM-IX, is a convenient tool for the businessman who needs to analyze sales, costs, or any numeric data, and for the scientist or engineer involved in statistical work. Tasks that can be performed with STATPAK include:

- Data creation and modification.
- Correlation analysis.
- Time series analyses for variable forecasting.
- Statistical analyses including mean, standard deviation, standard error, maximum, minimum, and range.
- Histograms and scatter diagrams.
- Transformations of the data, including square root, natural and common logarithms, exponential, sine, and cosine.
- Linear, multiple, stepwise, and polynomial regression.
- Chi-square and Kolmogorov-Smirnov goodness-of-fit.
- Curve fitting.
- Confidence limits on the mean.
- Contingency table.
- Data generation using a variety of functions to calculate variable values.

An outstanding feature of STATPAK is the ability to store on a file results from most STATPAK analyses. The results may be used later with a different STATPAK analysis, Tymshare's TYMTAB and FINPAK programs, or user-written programs.

The user calls STATPAK by typing

-STATPAK↵

in the EXECUTIVE. The system prints

1>

to indicate that the user may enter his first STATPAK command. When control is transferred to the STATPAK command level, the system prints the appropriate command number and a greater than sign (>); each time STATPAK executes a command, the command number is incremented by 1.

The user may type his input on successive lines by typing a Line Feed at the end of the line to be continued. For example, STATPAK prompts for the column names for the data being entered. If the user's entries require more than one line, the user types a Line Feed as the terminator for the line to be continued. For example,

COLUMN TITLES OR #: DAY,MONTH,CASH,OVHEAD,INVEN,SALES1,↓
SALES2,SALES3,DTOTAL,MTOTAL,PROFIT↵

The Line Feed permits the user to continue the input on the next line; the Line Feed does not insert any characters. The Carriage Return terminates the entry.

SYMBOL CONVENTIONS

In all examples in this manual, everything typed by the user is underlined. The symbols used to indicate user-typed characters are:

Carriage Return: ↵

Line Feed: ↓

Control characters are denoted by a superscript c. For example, A^c denotes Control A. The method for typing a control character depends on the type of terminal used. Consult the literature for your particular terminal or see your Tymshare representative.

Lowercase letters used in examples of command forms represent the input to be typed. In the following command,

p>SAVE file name↵

the characters file name indicate that the user should type a file name in that position.

Brackets indicate an option; they are not part of the statement or command. For example,

p>LIST [matrix component]↵

indicates that the user may optionally specify a matrix component.

Braces in a statement form indicate that the user must enter one of the items described within the braces. The braces are not part of the statement. For example,

$$\text{p>SAVE} \left\{ \begin{array}{l} \text{individual column} \\ \text{column list} \\ \text{column range} \end{array} \right\} \text{ON file name} \hookleftarrow$$

indicates that the user must specify one of the items described within the braces as part of the command.

STATPAK command level is indicated by sequential numbers and greater than signs. For example, successive commands appear as:

1 > command ↵
2 > command ↵
3 > command ↵

In all examples of command forms in this manual, the sequential number of the STATPAK command level prompt character is indicated with a p followed by a >.

The user may interrupt execution of any STATPAK command by typing an Alt Mode/Escape. If STATPAK has not completed execution of the interrupted command, the previous prompt number is repeated. If STATPAK has completed execution of the current command, the system prints the next prompt in sequence.

EDITING CHARACTERS

As the user enters information from the terminal, he may use the following editing characters: Control A, Control W, and Control Q

Control A deletes the previous character in the current line. On most terminals, a back arrow (←) prints when the user types a Control A. One character is deleted for each Control A typed. For example,

3# 3.3,6.8,1,7A^{C}←A^{C}←.7 ↵ *The first Control A deletes the 7; the second Control A deletes the comma.*

is accepted as:

3# 3.3,6.8,1.7

Control W deletes all preceding characters up to the first blank space encountered. On many terminals, a back slash (\) prints when the user types a Control W. For example,

```
7# 4.2,7.4, 1.3W^c\ 3.1 ↵
```

The Control W deletes 1.3 and stops the deletion at the blank position.

is accepted as:

```
7# 4.2,7.4, 3.1
```

Control Q deletes the entire current line and returns the carriage. On many terminals, an up arrow (↑) prints when a Control Q is used. The user may then retype the entire line. For example,

```
4# 5.3,6.4,7Q^c ↑
3.5,6.4,7.1 ↵
5#
```

Control Q deletes the entire line and returns the carriage. The user reenters the entire line.

NOTE: Control Q deletes only the current line. Successive Control Q's do not delete any preceding lines.

STATPAK ANALYSES

There are 20 analyses available in STATPAK; any of these analyses can be called by typing the name of the analysis after the STATPAK prompt. The analysis name can be abbreviated to the first three characters, except where additional characters are necessary to identify the name uniquely. For example, four letters are necessary to identify the CONTINGENCY and CONFIDENCE analyses. The LINEAR command may be shortened to LIN or anything other than LINE; LINE is interpreted as the LINE command. The 20 STATPAK analyses are listed below.

Analysis	Description
ELEMENTARY	Calculates six basic statistics for each variable.
DESCRIPTIVE	Calculates 18 statistics for a specified variable.
SCATTER	Plots two variables on the terminal.
PLOT	Produces a graph for one independent variable and as many as three dependent variables.
HISTOGRAM	Prints a histogram for any selected variable.
CUMULATIVE	Prints a cumulative frequency histogram for any selected variable.
CORRELATION	Computes correlation coefficients for each column of data against each other column.
SPEARMAN	Measures the degree of correlation between two columns ranked according to different criteria.
KENDALL	Calculates a coefficient of concordance for any number of ranked columns.
CONTINGENCY	Tests statistical independence of two variables.
LINEAR	Fits a set of data to a linear equation of the form $y=A+Bx$.
MULTIPLE	Fits a curve of the form $y=B_0+B_1x_1+B_2x_2+\cdots+B_kx_k$.

Analysis	Description
STEPWISE	Performs a multiple regression, using a stepping technique.
POLYNOMIAL	Fits a curve of the form $y=B_0+B_1x+B_2x^2+\cdots+B_kx^k$.
CURVE	Performs a least squares fit for six types of curves.
CHI-SQUARE	Performs a chi-square goodness-of-fit test.
KOLMOGOROV-SMIRNOV	Performs a Kolmogorov-Smirnov goodness-of-fit test.
CONFIDENCE	Computes a confidence level for an associated interval for the mean.
XPOS	Generates forecasting parameters, smoothing coefficients, and variable forecasts.
FORECAST	Computes a column of variable values using the results of any one of a variety of STATPAK analyses or user-defined functions.

DATA MANIPULATION COMMANDS

With the exception of the CHI-SQUARE, KOLMOGOROV-SMIRNOV, and Data Generation analyses, each analysis operates on data previously entered into STATPAK. For entering or modifying data, a complete set of data manipulation commands is available. Any of these commands may be typed at the STATPAK command level; with the exception of the LINE command, each may be abbreviated to three characters.

Commands for Entering and Saving Data

LINE	Changes the number of characters per line.
INPUT	Accepts a data matrix entered at the terminal.
SAVE	Saves all or part of a data matrix on a file.
LOAD	Reads a data matrix from an existing file.

Commands for Examining Data

LIST	Lists all or part of a data matrix with the title and the headings for columns and rows.
FAST	Lists all or part of a data matrix, but does not print the title or the headings.
SIZE	Prints the number of rows and columns in the data matrix.
ROW	Prints the names of the rows in the data matrix.
COLS	Prints the names of the columns in the data matrix.

Commands for Ordering Data

RANK	Creates a column of ranking numbers corresponding to ascending values in a specified column, and inserts the column of ranking numbers in a specified position in the data matrix.
ORDER	Orders a specified column or the entire matrix based on ascending values in the specified column.

Commands for Modifying and Adding Data

DELETE	Deletes all or part of a data matrix.
RENAME	Renames a column in a data matrix.
DUPLICATE	Creates a duplicate of an existing row or column and inserts the duplicate in a specified position in the matrix.
NUMBER	Sequentially numbers the rows in a data matrix and inserts the column of sequential numbers in a specified position in the matrix.
CHANGE	Changes any part of a data matrix.

Commands for Transforming Data

APPEND	Adds new rows or columns to the data matrix.
INSERT	Inserts one or more new rows or columns in a specified position in the matrix.
REPLACE	Replaces a column in a matrix with transformed data.

UTILITY COMMANDS

The following standard commands are available in STATPAK; each command may be abbreviated to the first three letters.

HELP (or ?)	Prints a list of commands with their descriptions; this command may be typed whenever assistance is required.
CAPABILITIES	Describes program capabilities.
INSTRUCTIONS	Prints operating instructions for STATPAK.
CREDITS	Prints credits for development of STATPAK.
CHARGES	Indicates additional charge, if any, for the program. There is no additional charge for STATPAK.
VERSION	Prints the number of the latest STATPAK update.
EXPLAIN	Explains in detail any STATPAK command the user types immediately after the word EXPLAIN.
QUIT (or Q)	Returns the user to the EXECUTIVE.
SAMPLE	Prints a sample STATPAK session.

SAMPLE ANALYSIS

To introduce STATPAK, a sample analysis of monthly operating costs is presented, using ELEMENTARY, one of 20 analyses available. The user logs in, then proceeds as follows:

```
-STATPAK

1>INPUT

TITLE: YEAR72

COLUMN TITLES OR #: COSTS

    COSTS
1#  4697          Each row is terminated by a Carriage Return.
2#  3684
3#  9628
4#  2749
5#  7492
6#  3958
7#  5727
8#  6948
9#  3757
10#  9386
11#  6243
12#  8572
13#               Data entry is terminated by an additional Carriage Return.

2>SAVE ABC        The user saves the data on file ABC.

NEW FILE- OK?   YES

3>ELEMENTARY      The user requests the ELEMENTARY statistics analysis.
```

VARIABLE	MEAN	STD DEV	STD ERR	MAXIMUM	MINIMUM	RANGE
COSTS	6070.083	2360.151	681.317	9628.000	2749.000	6879.000

```
4>QUIT            The user exits from STATPAK.

-
```

Thus, the mean operating cost for the year is computed as $6070.083. The mean, standard deviation, standard error, maximum, minimum, and range are all computed whenever an ELEMENTARY analysis is performed.

To illustrate that the user may shorten his input, a rerun of the previous example is shown below with all STATPAK commands abbreviated.

```
-STATPAK

1>INP

TITLE: YEAR72

COLUMN TITLES OR #: COSTS

     COSTS
1#   4697
2#   3684
3#   9628
4#   2749
5#   7492
6#   3958
7#   5727
8#   6948
9#   3757
10#   9386
11#   6243
12#   8572
13#

2>SAV ABC

NEW FILE- OK? Y

3>ELE


VARIABLE    MEAN       STD DEV    STD ERR    MAXIMUM     MINIMUM     RAN

COSTS     607 0.083   2360.151   681.317   9628.000   2749.000   6879.


4>Q


-
```

SECTION 2
STATPAK DATA FILES

Each STATPAK analysis operates on data provided by the user. The data is arranged in rows and columns, each column representing a variable and each row representing one observation per variable. The row and column arrangement of data is called a data matrix; it may contain as many as 2000 values.

The user creates a data file by entering the data at the terminal, then saves all or part of the data matrix on a file. Once the data is in STATPAK, whether entered at the terminal or from an existing file, the user may list, modify, transform, or order the data matrix with simple, flexible commands.

STATPAK contains all the line editing characters available in Tymshare's EDITOR.[1] During data entry, the user may type any of these control characters and STATPAK performs the appropriate function, using the previous line as an image. For example,

3# 23.5,42.7,36.0,92.4
4# Z^c6 23.5,42.7,36.5,87.2
5# D^c 23.5,42.7,36.5,87.2

In row 4, the user types a Control Z and a 6 to instruct STATPAK to copy the previous line up to and including the character 6. The user then types the rest of the values for row 4. Row 5 values are identical to row 4 values, so the user merely types a Control D to instruct STATPAK to copy the preceding line.

---

1 - See the Tymshare EDITOR Reference Manual.

MATRIX COMPONENT SPECIFICATION

Some commands require the specification of one or more matrix components; other commands optionally operate on one or more matrix components. The following list details the general forms for matrix component specification. The column names COL1, COL2, . . . ,COL12 are used for illustration only; any six-character column name is permitted.

Form Name	Specification	Refers To	Example
Individual column	column name	The column named	COL1
Individual row	row name	The row named	34#
Column list	column $name_1$, column $name_2$, . . .	Each column named	COL1,COL2, . . .
Row list	row $name_1$, row $name_2$, . . .	Each row named	1#,2#, . . .
Column range	column $name_1$: column $name_2$	column $name_1$ through column $name_2$	COL1:COL5
Row range	row $name_1$: row $name_2$	row $name_1$ through row $name_2$	2#:15#
Individual element	row name column name	The element in location corresponding to the row and column named.	8# COL3

Form Name	Specification	Refers To	Example
Submatrix	row $name_1$: row $name_2$ column $name_1$: column $name_2$	The submatrix with row $name_1$ through row $name_2$ and column $name_1$ through column $name_2$	1#:4# COL3:COL7
Element list	row $name_1$, row $name_2$, . . . , row $name_n$ column $name_1$, column $name_2$, . . . , column $name_m$	The locations defined by the rows and columns, where the order is: row $name_1$, column $name_1$ row $name_1$, column $name_2$ ⋮ row $name_1$, column $name_m$ row $name_2$, column $name_1$ row $name_2$, column $name_2$ ⋮ row $name_2$, column $name_m$ ⋮ row $name_n$, column $name_m$	2#,4#,8# COL3,COL9,COL10, COL12

STATPAK permits the user to specify the final row or column in the matrix by typing a dollar sign ($) as part of the component specification. For example, a row range may be specified as

row name$_1$:$

to indicate the range of rows from row name$_1$ through the final row in the matrix. A dollar sign appearing alone indicates the last column. To indicate only the last row, the user enters:

$#

In the commands which permit all the component specification forms, any order or combination is permitted. For example,

2>LIST 1#:5# C3 and 2>LIST C3 1#:5#

cause STATPAK to produce identical listings. The command

3>LIST 4#:9# C7 27#,28#

instructs STATPAK to list the values in rows 4 through 9 and rows 27 and 28 for column C7.

ENTERING AND SAVING DATA

The STATPAK user may enter his data from the terminal or from an existing file. STATPAK includes editing characters and error messages to assist the user during data entry. The user may save all or part of the newly created data matrix on a file.

STATPAK eliminates the possibility of entering two data matrices simultaneously. Each time the user types a LOAD command, the current data is cleared and the new data matrix is loaded. If the user types INPUT when a matrix is loaded, STATPAK asks the user to verify his intentions by responding appropriately to the question:

CLEAR EXISTING DATA?

After the user responds, STATPAK prints

CLEARED. or NOT CLEARED.

The LINE Command

The LINE command permits the user to change the number of characters the system prints on a line. STATPAK presets the automatic Carriage Return to position 72, but the user may change the number by simply typing

p> LINE n

where n is an integer from 13 to 256; the value of n represents the number of characters per line, including an automatic Carriage Return at position n. For example,

2> LINE 120

instructs STATPAK to print as many as 120 characters on a physical line.

NOTE: The LINE command cannot be abbreviated.

The INPUT Command

The user must use the INPUT command to enter data from the terminal. After the user calls STATPAK and the prompt appears, the user types INPUT followed by a Carriage Return. The system prompts for an identifying title; every data file has a title of six or fewer characters; the first character must be a letter from A to Z. For example,

```
-STATPAK
1>INPUT
TITLE: SALES1
```

Next, the system prompts for column titles or the number of columns by printing:

```
COLUMN TITLES OR #:
```

The user responds with a list of his column titles or simply a number indicating the number of columns. In the latter case, STATPAK assigns the column titles as C1, C2,...,Cn, where n is the number of columns specified. For example ,

```
COLUMN TITLES OR #: 7
```

names seven columns as C1, C2, C3, C4, C5, C6, and C7. In the following example, the user enters his own column titles:

```
COLUMN TITLES OR #: MONTH,SALES,COST,INVEN
```

STATPAK automatically assigns the row names as 1#, 2#, 3#, and so on, for all rows in the data, and prompts for the input for each row. The user terminates the data matrix by typing a Carriage Return in response to the row prompt. For example,

```
-STATPAK

1>INPUT

TITLE: YR1970

COLUMN TITLES OR #: MONTH,SALES,COST,INVEN

     MONTH,SALES,COST,INVEN
1#   1,598.54,375.43,895.
2#   2,643.44,340.19,905.80
3#   3,425.00,380.64,998.75
4#   4,765.74,450.00,712.
5#   5,896.32,433.51,780.43
6#   6,624.90,520.53,621.50
7#   7,913.45,378.64,746.31
8#   8,672.56,428.57,877.60
9#   9,745.22,643.26,689.21
10#   10,830.22,416.73,488.56
11#   11,456.34,417.50,683.
12#   12,669.80,522.65,704.35
13#

2>LIST
```

The user types a value for each column and terminates the row input with a Carriage Return.

A Carriage Return as the only response to the row prompt, terminates data entry.

```
TITLE- YR1970
     MONTH    SALES     COST     INVEN
1#       1   598.54   375.43    895.00
2#       2   643.44   340.19    905.80
3#       3   425.00   380.64    998.75
4#       4   765.74   450.00    712.00
5#       5   896.32   433.51    780.43
6#       6   624.90   520.53    621.50
7#       7   913.45   378.64    746.31
8#       8   672.56   428.57    877.60
9#       9   745.22   643.26    689.21
10#     10   830.22   416.73    488.56
11#     11   456.34   417.50    683.00
12#     12   669.80   522.65    704.35


3>
```

If the user types fewer values in a row than the number of columns specified for the matrix, the system prompts for each missing value. For example,

```
-STATPAK
1>INPUT
TITLE: EXAMP
COLUMN TITLES OR #: 3
     C1,C2,C3
1#   2
C2:  3
C3:  5
2#   4
C2:  6,7
3#   3,7
C3:  2
4#
2>LIST

TITLE- EXAMP
    C1  C2  C3
1#   2   3   5
2#   4   6   7
3#   3   7   2

3>
```

The SAVE Command

The SAVE command allows the user to save the entered data and titles for use with future analyses. Also, the command is used to save the results of an analysis. When the user types

p>SAVE file name ↵

STATPAK saves the entire matrix on a data file and creates a description file. The data file contains the data values, and has the specified file name plus the file name extension 'DAT.A'. The description file contains the title, the number of rows and columns, the row names, and the column names; it has the file name extension 'NAM.A'. For example,

p>SAVE SURVEY ↵

instructs STATPAK to save SURVEY'DAT.A' and SURVEY'NAM.A' in the user's directory.

The user may save one or more columns of a data matrix on a separate file; however, only the data is saved and no corresponding description file is created. The user may name an individual column, a column list, or a column range in the command form

```
       ( individual column )
p>SAVE { column list       } ON file name ↵
       ( column range      )
```

where the column specifications are as detailed on page 14. For example,

p>SAVE C1,C2 ON HIST ↵

creates a file named HIST which contains a two-column matrix, that is, the data from C1 and C2. Note that when saving part of the matrix, no description file is created and no file name extension assigned.

When the user enters the SAVE command, directing STATPAK to write data or calculated results on a file, STATPAK responds with the message:

NEW FILE- OK?   or   OLD FILE- OK?

The user types either YES (or Y) to confirm the command, or NO (or N) to abort the request. If the user types YES (or Y) in response to the message

OLD FILE- OK?

STATPAK writes the new data over the previous contents of the file.

The LOAD Command

The LOAD command permits the user to enter an existing data file for input to STATPAK. The file must contain the data written in the same order as entered with an INPUT command, that is, row by row. To enter an existing data file, the user types:

p> LOAD file name ↵

Note that the file name extension is not included. For example, the command

1> LOAD EXAM ↵

loads the data in file EXAM'DAT.A' and the description in file EXAM'NAM.A' for use with STATPAK.

The user may type simply LOAD followed by a Carriage Return, and the system prompts for the name of the file. STATPAK automatically loads the corresponding description file, such as EXAM'NAM.A', as well as the specified data file, and returns control to the STATPAK command level.

When no corresponding description file exists, STATPAK prompts for descriptive information about the loaded data. For example, the file EXAMP contains columns of data only; no description file exists for EXAMP. When the user loads EXAMP, STATPAK prompts for the descriptive information:

```
-STATPAK

1>LOAD EXAMP

TITLE: TEST

COLUMN TITLES OR #: 3

2>LIST


TITLE- TEST
          C1        C2        C3
1#       5.5      7.80       9.2
2#      -6.3      8.99       4.0
3#      62.7     34.10     -10.0
4#      50.0     30.00      20.0
5#      12.0    -34.20      32.1
6#       9.0      4.00       1.0
7#      -3.0      4.00      51.0
8#      71.0     17.00     -14.0
9#       1.0      5.00      51.0
10#      2.0      4.00       6.0
11#     81.0     32.00     -41.0
12#      6.0      8.00      10.0
13#     32.0     55.00      61.0
14#     12.0     54.00      71.0
15#      3.0      5.00      91.0


3>
```

EXAMINING THE DATA

Using simple STATPAK commands, the user may list any part of the data matrix, determine the basic parameters of the matrix, and review row or column titles. The LIST and FAST commands print titled and untitled listings, respectively, for all or part of the matrix. The SIZE command prints the matrix dimensions, and the ROWS and COLS commands print the row names and column names, respectively.

The LIST Command

The user may request a listing of all or part of a data matrix with the LIST command. LIST prints the title, the row names, and the column names for the data. The command form is:

p> LIST matrix component ↵

where the matrix component may be specified in any of the forms described on pages 14 and 15. The user simply types

p> LIST ↵

to request a titled listing of the entire data matrix. For example,

```
2>LIST ↵
```

STATPAK prints the entire matrix.

```
TITLE- LINEAR
     TIME VAR1 VAR2
1#     15    3    1
2#     10   15    2
3#     16   21    3
4#     20   29    4
5#     23   33    5
6#     25   35    6
7#     26   37    7
8#     30   46    8
9#     36   60    9
10#    48   72   10
11#    62   90   11
12#    78  107   12
13#    94  114   13
14#   107  123   14
15#   118  135   15
```

3>LIST 7#:10# ↵ *The user asks STATPAK to list a range of rows.*

	TIME	VAR1	VAR2
7#	26	37	7
8#	30	46	8
9#	36	60	9
10#	48	72	10

4>

The FAST Command

The FAST command lists all or part of a data matrix without a title or headings. The FAST command functions in the same manner as the LIST command, except that the FAST command suppresses the printing of any headings. The general form of the FAST command is

p> FAST matrix component ↵

where the matrix component may be specified in any of the forms listed on pages 14 and 15. If the user wants an untitled listing of the entire matrix, he types:

p>FAST ↵

Example

```
2>FAST ↵
```

The user wishes to see all the values in the matrix without a title or headings.

15	3	1
10	15	2
16	21	3
20	29	4
23	33	5
25	35	6
26	37	7
30	46	8
36	60	9
48	72	10
62	90	11
78	107	12
94	114	13
107	123	14
118	135	15

```
3>FAST 7#:10# ↵
```

The user specifies a row range.

26	37	7
30	46	8
36	60	9
48	72	10

```
4>
```

The SIZE Command

The SIZE command prints the number of rows and columns in the current data matrix. For example,

```
4>SIZE
15 ROWS,    3 COLUMNS
5>
```

The ROW and COLS Commands

The ROW command prints the row names for the current data matrix, and the COLS command prints the column names for the current data matrix. For example,

```
5>ROW
1#,2#,3#,4#,5#,6#,7#,8#,9#,10#,11#,12#,13#,14#,15#
6>COLS
TIMF,VAR1,VAR2
7>
```

ORDERING AND RANKING THE DATA

STATPAK includes a command for ordering the data sequentially and a command for ranking the data. The ORDER command orders a column or the entire matrix; the RANK command creates a column of numbers corresponding to the rank of each value in a column.

The RANK Command

The RANK command creates a column of ranking numbers based on ascending values of a specified column. The general form is:

```
                                          [BEFORE column name]
p> RANK column name IN column name        [AFTER column name]
```

The ranking numbers correspond to the ascending values in the first column named. The second column name names the column of ranking numbers which appears before or after the column named last. For example,

```
p> RANK HEIGHT IN HTRANK BEFORE WEIGHT
```

creates a column of ranking numbers corresponding to ascending values of HEIGHT, names the column of ranking numbers HTRANK, and inserts the column in the data matrix immediately preceding the column WEIGHT.

If the user merely types

```
p> RANK
```

STATPAK prompts for each specification. Successive prompts are:

```
COLUMN TO BE RANKED:
NEW COLUMN NAME:
BEFORE OR AFTER:
COLUMN:
```

If the user types a Carriage Return in response to the prompt

BEFORE OR AFTER:

STATPAK does not prompt for

COLUMN:

but merely appends the new column of ranking numbers.

If several equal values are to be ranked, equal rank numbers are assigned to each, the rank numbers each being the average rank number. For example, if four values are equal and they occur in positions 3, 4, 5, and 6, the rank value assigned to each is 4.5.

Example

```
-STATPAK
1>LOAD EXAM
2>LIST

TITLE- EXAFIT
          C1       C2       C3
1#       5.5     7.80      9.2
2#      -6.3     8.99      4.0
3#      62.7    34.10    -10.0
4#      50.0    30.00     20.0
5#      12.0   -34.20     32.1
6#       9.0     4.00      1.0
7#      -3.0     4.00     51.0
8#      71.0    17.00    -14.0
9#       1.0     5.00     51.0
10#      2.0     4.00      6.0
11#     81.0    32.00    -41.0
12#      6.0     8.00     10.0
13#     32.0    55.00     61.0
14#     12.0    54.00     71.0
15#      3.0     5.00     91.0
```

```
3>RANK C1 IN C1RANK AFTER C1

4>RANK C3 IN C3RANK

BEFORE OR AFTER:

5>LIST


TITLE- EXAFIT
        C1  C1RANK      C2      C3  C3RANK
1#     5.5     6.0    7.80     9.2     7.0
2#    -6.3     1.0    8.99     4.0     5.0
3#    62.7    13.0   34.10   -10.0     3.0
4#    50.0    12.0   30.00    20.0     9.0
5#    12.0     9.5  -34.20    32.1    10.0
6#     9.0     8.0    4.00     1.0     4.0
7#    -3.0     2.0    4.00    51.0    11.5
8#    71.0    14.0   17.00   -14.0     2.0
9#     1.0     3.0    5.00    51.0    11.5
10#    2.0     4.0    4.00     6.0     6.0
11#   81.0    15.0   32.00   -41.0     1.0
12#    6.0     7.0    8.00    10.0     8.0
13#   32.0    11.0   55.00    61.0    13.0
14#   12.0     9.5   54.00    71.0    14.0
15#    3.0     5.0    5.00    91.0    15.0


6>SAVE EXRS

NEW FILE- OK? YES

7>QUIT

-
```

The ORDER Command

The ORDER command orders a column or entire matrix according to ascending values in the specified column. To order a column, the user types

p> ORDER column name ⤶

and STATPAK orders only the column named. The command form

p> ORDER BASED ON column name ⤶

instructs STATPAK to order the entire matrix based on ascending values in the column named.

Example

```
-STATPAK ⤶
1>LOAD DATA ⤶

2>LIST ⤶

TITLE- RANK
      A  B  C
1#    2  4  6
2#    1  3  1
3#    3  1  4
4#    4  2  5
5#    5  6  2
6#    6  5  8
7#    7  8  3
8#    8  7  7
```

```
3>ORDER B
```

The user wishes to order column B and leave the rest of the matrix unchanged.

```
4>LIST


TITLE- RANK
      A  B  C
1#    2  1  6
2#    1  2  1
3#    3  3  4
4#    4  4  5
5#    5  5  2
6#    6  6  8
7#    7  7  3
8#    8  8  7
```

```
5>ORDER BASED ON A
```

The user instructs STATPAK to reorder the entire matrix according to ascending values in column A.

```
6>LIST


TITLE- RANK
      A  B  C
1#    1  2  1
2#    2  1  6
3#    3  3  4
4#    4  4  5
5#    5  5  2
6#    6  6  8
7#    7  7  3
8#    8  8  7


7>
```

MODIFYING AND ADDING DATA

STATPAK offers several commands for changing the current data matrix. The DELETE command deletes all or part of the matrix; the RENAME command changes column names; the DUPLICATE command duplicates columns or rows; the NUMBER command creates a column containing sequential numbers for the rows; and the CHANGE command changes individual elements or any part of the matrix.

The APPEND, INSERT, and REPLACE commands also allow the user to modify or add data to the matrix; however, since these commands include an additional transformation capability, they are discussed separately on page 41.

The DELETE Command

The DELETE command allows the user to delete all or part of a data matrix. The form of the command is

p> DELETE matrix component ↲

where any of the first six matrix components listed on page 14 may be specified in the DELETE command. If the user merely types DELETE followed by a Carriage Return, the system prompts to determine if the user wishes to delete the entire matrix. For example,

```
p> DELETE ↲
 ALL?
```

The user responds YES (or Y) and a Carriage Return to delete the entire matrix. If the user types NO (or N) followed by a Carriage Return, the system prompts for the matrix component specification. For example,

```
p> DELETE
ALL? NO
 DATA TO BE DELETED: C5:C9
```

instructs STATPAK to delete a range of columns from C5 through C9.

The RENAME Command

The RENAME command permits the user to change column names. The command form is:

```
p> RENAME old column name AS new column name
```

For example, if the user wishes to change a column name from COST to OVHEAD, he types:

```
p> RENAME COST AS OVHEAD
```

If the user enters an incomplete command, STATPAK prompts for each item. For example,

```
p> RENAME
 COLUMN: COST
 NEW COLUMN NAME: OVHEAD
```

The DUPLICATE Command

The DUPLICATE command creates a new row or column of data which is a duplicate of an existing row or column; the command also allows the user to duplicate a range or list of rows or columns. The user must name the new column or columns and may designate the position of the new row or column. Rows are numbered automatically, so the user does not specify new row names. For example,

```
p> DUPLICATE COL2 AFTER COL5 AS DUPC2
```

duplicates the column named COL2, inserts the duplicate column after COL5, and names the new column DUPC2. The following example illustrates the duplication of rows.

```
1>LOAD SALES1

2>LIST


TITLE- ELEDES
     MONTH HIPER   PRODA   PRODB
1#       1     2  275.00   236.5
2#       2     2  292.50   225.0
3#       3     2  300.00   241.7
4#       4     3  250.00   475.0
5#       5     4  262.50   550.0
6#       6     4  301.50   565.0
7#       7     4  288.75   535.0
8#       8     4  306.25   555.0
9#       9     4  318.75   548.5
10#     10     4  323.75   550.0
11#     11     5  257.00   605.0
12#     12     5  279.50   615.0


3>DUPLICATE 3# BEORE 9#
```

The user instructs STATPAK to duplicate row 3 and insert the new row before row 9.

```
4>LIST


TITLE- ELEDES
     MONTH HIPER   PRODA   PRODB
1#       1     2  275.00   236.5
2#       2     2  292.50   225.0
3#       3     2  300.00   241.7
4#       4     3  250.00   475.0
5#       5     4  262.50   550.0
6#       6     4  301.50   565.0
7#       7     4  288.75   535.0
8#       8     4  306.25   555.0
9#       3     2  300.00   241.7
10#      9     4  318.75   548.5
11#     10     4  323.75   550.0
12#     11     5  257.00   605.0
13#     12     5  279.50   615.0
```

Rows 9 through 12 in the old matrix become rows 10 through 13, with the addition of a new row 9.

The matrix components to be duplicated may be specified in any of the first six forms listed on page 14. The complete form of the DUPLICATE command is:

$$\underline{\text{p>DUPLICATE matrix component}\left\{\begin{matrix}\text{BEFORE}\\ \text{AFTER}\end{matrix}\right\}\left\{\begin{matrix}\text{column name}\\ \text{row name}\end{matrix}\right\}\text{AS new column name}}$$

Note that the final AS clause is included only for column duplication; row names are automatically assigned and correspond to the row's position in the matrix. The user may type merely DUPLICATE, or part of the command, and STATPAK prompts for the needed information. For example,

```
5>DUPLICATE

ROWS OR COLUMNS: COLUMNS
COLUMNS TO BE DUPLICATED: HIPER
BEFORE OR AFTER: AFTER
COLUMN: PRODA
NEW NAMES: HP1

6>DUPLICATE

ROWS OR COLUMNS: ROWS
ROWS TO BE DUPLICATED: 4#,8#
BEFORE OR AFTER: BEFORE 2#

7>LIST


TITLE- ELEDES
     MONTH HIPER   PRODA HP1   PRODB
1#       1     2  275.00   2   236.5
2#       4     3  250.00   3   475.0
3#       8     4  306.25   4   555.0
4#       2     2  292.50   2   225.0
5#       3     2  300.00   2   241.7
6#       4     3  250.00   3   475.0
7#       5     4  262.50   4   550.0
8#       6     4  301.50   4   565.0
9#       7     4  288.75   4   535.0
10#      8     4  306.25   4   555.0
11#      3     2  300.00   2   241.7
12#      9     4  318.75   4   548.5
13#     10     4  323.75   4   550.0
14#     11     5  257.00   5   605.0
15#     12     5  279.50   5   615.0


8>
```

The NUMBER Command

The NUMBER command creates a new column containing the sequential numbers for the rows of the data matrix. The user may name the new column and specify its position with the NUMBER command, or the system prompts for the information. The complete form of the NUMBER command is:

p> NUMBER new column name [BEFORE column name]
[AFTER column name] ↵

For example, to insert a column named SEQ containing sequential row numbers after a column named FREQ, the user types:

p> NUMBER SEQ AFTER FREQ ↵

The CHANGE Command

The user may change all or any part of the matrix; the new data may be entered from the terminal or from a file. Any of the matrix components listed on pages 14 and 15 may appear in a CHANGE command.

When the user wishes to enter the new data at the terminal, the form of the CHANGE command used is:

p> CHANGE matrix component ↵

For example, the user types

p> CHANGE 2# COL2 ↵

to change the element in row 2# and column COL2; he enters the new data at the terminal. Alternatively, the user may type CHANGE followed by a Carriage Return and STATPAK prompts to determine if the user wishes to change the entire matrix. For example,

p> CHANGE ↵
ALL?

The user responds YES or NO, as appropriate, followed by a Carriage Return. If the user responds NO, the system prompts

DATA TO BE CHANGED:

and the user enters any matrix component specification described on pages 14 and 15.

When entering the new data from the terminal, the STATPAK prompts are similar to the INPUT prompts. The example below illustrates the CHANGE command with various matrix components and the STATPAK prompts for data entry.

```
3>LIST
```

```
TITLE- STATUS
     C1   C2   C3    C4
1#   19   56   25    76
2#   48   46    6   270
3#   63   32    5   310
4#   64   31    5   260
5#   69   25    6   220
6#   66   24   10   200
```

```
4>CHANGE 2# C3
```
The user wishes to change the value in the second row of column C3.

```
     C3
2#   18
```
STATPAK prompts for the new value, and the user enters the data.

```
5>CHANGE C2
```
The user wishes to change the entire column C2.

```
     C2
1#   52
2#   45
3#   27
4#   25
5#   30
6#   21
```

```
6>CHANGE 3# C1,C2,C3          The user specifies the third value for columns C1, C2, and C3.

    C1,C2,C3
3#  65,30,7

7>LIST          The user asks STATPAK to list the changed matrix.


TITLE- STATUS
     C1  C2  C3   C4
1#   19  52  25   76
2#   48  45  18  270
3#   65  30   7  310
4#   64  25   5  260
5#   69  30   6  220
6#   66  21  10  200


8>
```

When the user decides to change the matrix and enter the new data from a file, he must include FROM and the file name at the end of the CHANGE command. The form of the command is

p> CHANGE matrix component FROM file name

or simply:

p> CHANGE FROM file name

STATPAK prompts to determine whether the user wants to change the entire matrix or specific matrix components.

STATPAK allows the user to enter the data from a free-format file. On the file, the order of the data corresponds to the usual form for input, that is, row by row. The example below illustrates the same changes as the previous CHANGE commands, but the user enters the data from a file rather than from the terminal. The user may separate the data items with a space or a comma. Carriage Returns are ignored, so more than one row may appear on the same line in the file.

```
-TYPE MOD1            The user lists the files containing the new data.

18

-TYPE MOD2

52,45 27 25,30
21

-TYPE MOD3

65 30
7

-STATPAK

1>LOAD SAMP

2>LIST


TITLE- STATUS
      C1   C2   C3    C4
1#    19   56   25    76
2#    48   46    6   270
3#    63   32    5   310
4#    64   31    5   260
5#    69   25    6   220
6#    66   24   10   200


3>CHANGE 2# C3 FROM MOD1

4>CHANGE C2 FROM MOD2

5>CHANGE 3# C1,C2,C3 FROM MOD3

6>LIST


TITLE- STATUS
      C1   C2   C3    C4
1#    19   52   25    76
2#    48   45   18   270
3#    65   30    7   310
4#    64   25    5   260
5#    69   30    6   220
6#    66   21   10   200


7>
```

TRANSFORMING THE DATA

STATPAK permits the user to modify or add data to the current matrix with the APPEND, INSERT, and REPLACE commands. If desired, the user may also specify transformations using functions or arithmetic operators in the same STATPAK command.

Expressions

STATPAK permits the user to specify transformations with any of the data manipulation commands described in this section. The transformation is requested in the form

column name=expression

where the column name may be a new column to be added or an existing column to be replaced. For example, the following commands request valid transformations:

```
p>REPLACE COL6=C4*C5+VAR1*VAR2 ↵
p>INSERT COL8=MEAN*1.2-STD ↵
p>APPEND COL1=FREQ1*PROB1+FREQ2*PROB2 ↵
```

An expression may contain column titles, numbers, arithmetic operators, and functions. The remaining discussion details valid expressions and the method by which STATPAK evaluates them.

The user may specify addition, subtraction, multiplication, division, and exponentiation, using the arithmetic operators in the table below.

Operator	Meaning
+	Addition
–	Subtraction
*	Multiplication
/	Division
** or ↑	Exponentiation

In addition, STATPAK contains the following functions which the user may incorporate in any expression.

Function	Meaning	Example
SQR	Square root	COL6=SQR(MEAN*C1)
LGT	Base 10 logarithm	TRANS=LGT(COL4+COL3)
LOG	Base e logarithm	COL5=LOG(15*FREQ)
EXP	Exponential (exp(2)= e^2)	COL5=EXP(COL2+4)
SIN	Sine of angle in radians	COL8=SIN(COL3)
COS	Cosine of angle in radians	COL2=COS(COL4)*COL2

STATPAK evaluates an expression from left to right, with no hierarchy of operations. For example, the expression

COL1+COL2/7*COL3

is evaluated from left to right as:

$$\frac{COL1+COL2}{7} \cdot COL3$$

The user may order the operations with parentheses. The portion of an expression enclosed in parentheses is evaluated first. If parentheses appear within parentheses, the part of the expression within the inner set is evaluated first. For example,

A+B*SQR(C)/D

is evaluated as

$$\frac{(A+B)*SQR(C)}{D}$$

but

A+(B*SQR(C)/D)

is evaluated as:

$$A + \frac{B*SQR(C)}{D}$$

The APPEND Command

The APPEND command permits the user to add one or more rows or columns to the end of the existing matrix. For example, the user may type

p> APPEND FREQ1 ↵

and STATPAK prompts for a value of FREQ1 for each row and adds the column of data to the end of the matrix. The user may enter the new data directly at the terminal or from a file.

In addition, the APPEND command allows the user to combine column transformations. For example, to add a column named TRANS, which is the sum of COL1 and COL2, the user types

p> APPEND TRANS=COL1+COL2 ↵

and STATPAK creates the data for TRANS as specified, appends the column to the end of the current matrix, and returns control to STATPAK command level.

There are three forms of the APPEND command. To add columns of data to the end of the matrix, the form is:

p> APPEND $\left\{\begin{matrix}\text{individual column} \\ \text{column list}\end{matrix}\right\}$ ↵

To add one or more rows to the end of the matrix, the form is:

p> APPEND ↵

Finally, to perform transformations and create a new column at the end of the matrix, the form is:

p> APPEND column name=expression ↵

where the expression may contain column names, numbers, and any of the functions and arithmetic operators listed on page 42.

Except for transformations, the user must enter the values for the new data. When entering data directly from the terminal, STATPAK prompts the user in the same form as the INPUT or CHANGE commands. In the following example, the user adds two columns to the end of the current matrix, entering the data directly at the terminal.

```
2>LIST


TITLE- RECPTS
       DEPT1     DEPT2     DEPT3
1#    234.45    132.87    456.33
2#    342.76    532.34    458.90
3#    265.40    365.48    550.81
4#    402.45    351.39    469.08


3>APPEND DEPT4,DEPT5

      DEPT4,DEPT5
1#    14.39,201.55
2#    45.88,195.46
3#    59.35,215.60
4#    75.36,180.50

4>LIST


TITLE- RECPTS
       DEPT1     DEPT2     DEPT3     DEPT4     DEPT5
1#    234.45    132.87    456.33    14.39    201.55
2#    342.76    532.34    458.90    45.88    195.46
3#    265.40    365.48    550.81    59.35    215.60
4#    402.45    351.39    469.08    75.36    180.50
```

Next, the user adds two rows to the existing matrix. Note that a Carriage Return terminates the data entry procedure.

```
5>APPEND

     DEPT1,DEPT2,DEPT3,DEPT4,DEPT5
5#   456.98,421.55,368.59,88.00,216.57
6#   385.49,362.43,419.36,95.05,251.78
7#
```

Finally, the user wants to add a column named TOTAL, which is the sum of the other columns.

```
6>APPEND TOTAL=DEPT1+DEPT2+DEPT3+DEPT4+DEPT5

7>LIST


TITLE- RECPTS
       DEPT1    DEPT2    DEPT3    DEPT4    DEPT5     TOTAL
1#    234.45   132.87   456.33    14.39   201.55   1039.59
2#    342.76   532.34   458.90    45.88   195.46   1575.34
3#    265.40   365.48   550.81    59.35   215.60   1456.64
4#    402.45   351.39   469.08    75.36   180.50   1478.78
5#    456.98   421.55   368.59    88.00   216.57   1551.69
6#    385.49   362.43   419.36    95.05   251.78   1514.11


8>
```

When the user wishes to append data from a file, he types the appropriate APPEND command, including a FROM clause which specifies the file containing the data to be appended. In the example below, the user performs the same additions as above, but enters the data from a file. The file contains the elements in the same order as in all STATPAK data entry procedures; the data is written row by row with a comma or a space between values.

```
-TYPE NEWDPT
```

With the EXECUTIVE TYPE command, the user lists the files containing the new data.

```
14.39,201.55,45.88,195.46
59.35
215.60 75.36 180.50

-TYPE ADDRTS

456.98 421.55 368.59 88.00 216.57
385.49 362.43,419.36,95.05,251.78

-TYPE SUM

1039.59 1575.34
1456.64
1478.78
1551.69
1514.11

-STATPAK

1>LOAD TRANS

2>LIST


TITLE- RECPTS
      DEPT1    DEPT2    DEPT3
1#   234.45   132.87   456.33
2#   342.76   532.34   458.90
3#   265.40   365.48   550.81
4#   402.45   351.39   469.08
```

```
3>APPEND DEPT4,DEPT5 FROM NEWDPT ↵
```
The user wants to add two new columns.

```
4>APPEND FROM ADDRTS ↵
```
APPEND implies APPEND rows unless a column name is specified.

```
5>APPEND TOTAL FROM SUM ↵

6>LIST ↵
```

TITLE- RECPTS

	DEPT1	DEPT2	DEPT3	DEPT4	DEPT5	TOTAL
1#	234.45	132.87	456.33	14.39	201.55	1039.59
2#	342.76	532.34	458.90	45.88	195.46	1575.34
3#	265.40	365.48	550.81	59.35	215.60	1456.64
4#	402.45	351.39	469.08	75.36	180.50	1478.78
5#	456.98	421.55	368.59	88.00	216.57	1551.69
6#	385.49	362.43	419.36	95.05	251.78	1514.11

```
7>
```

The INSERT Command

The INSERT command allows the user to insert rows or columns in an existing matrix. The new data may be entered at the terminal or from a file, or the user may specify column transformations. The general forms of the INSERT command when entering the new data at the terminal are:

p> INSERT {individual column / column list} {BEFORE / AFTER} column name ↵

to insert one or several new columns in a specified position in the matrix;

p> INSERT {BEFORE / AFTER} row name ↵

to insert one or more rows; and, to specify transformations, p> INSERT individual column=expression $\left\{\begin{matrix}\text{BEFORE}\\ \text{AFTER}\end{matrix}\right\}$ column name ↵

where the expression may contain column names, numbers, and any of the functions and arithmetic operators listed on page 42.

When entering data from a file, the user includes the FROM clause at the end of the command forms above. For example, the command

p> INSERT BEFORE 3# ↵

implies data entry at the terminal, whereas

p> INSERT BEFORE 3# FROM FILEX ↵

specifies data entry from a file named FILEX.

Note that the user does not specify a name for any inserted row; the new row name is automatically determined by STATPAK, and subsequent rows are renumbered appropriately.

The REPLACE Command

The user may replace one or several existing columns with transformed data. For example,

p> REPLACE C5=C5*2-C4 ↵

replaces each element in the column named C5 with twice its value, minus the value for that row in column C4.

STATPAK accepts a REPLACE command with several transformations, performs any calculations and substitutions, and returns control to STATPAK command level.

The form of the REPLACE command to replace one or more columns is

p> REPLACE column name=expression,column name=expression,...

where an expression may contain numbers, column names, arithmetic operators, and functions. See page 41 for an explanation of valid STATPAK expressions and their evaluation.

The example below illustrates the replacement of one column of data, C1, with transformed values. The user accomplishes these calculations with one simple REPLACE command. The new column C1 is listed to demonstrate the results.

```
2>LIST


TITLE- ESTATS
      MEAN   STD    X     C1
1#   23.45   5.6   .34   24.0
2#   34.25   3.2   .56   22.3
3#   28.40   7.2   .55   35.1


3>REPLACE C1=SQR(MEAN)+STD*X

4>LIST C1


               C1
1#   3.5504568017
2#   5.0693159750
3#   6.8910407708


5>
```

An example of a REPLACE command specifying more than one transformation is:

p>REPLACE C1=MEAN-C1/2-STD,MEAN=C1*4

Note that the values for C1 in the second expression are the current values of C1 after the first transformation is performed.

SECTION 3
ELEMENTARY ANALYSES

There are two elementary analyses in STATPAK. The ELEMENTARY analysis calculates six statistics for each variable in the data matrix; the DESCRIPTIVE analysis calculates 18 statistics for a single variable in the data matrix.

ELEMENTARY STATISTICS

STATPAK's ELEMENTARY analysis of a data matrix always produces six items of statistical information for each column of the matrix. These six items are the mean, standard deviation, standard error, maximum value, minimum value, and range of values.

To access the ELEMENTARY analysis, the user types:

p> ELEMENTARY

STATPAK automatically calculates the six statistics for each variable in the data matrix and prints them on the terminal. If the user wants to calculate the statistics and save them on a file, he types

p> ELEMENTARY TO file name

and STATPAK calculates the statistics and stores them on the named file. The results are not printed on the terminal but are saved on the file for future use.

Example

```
-STATPAK

1>LOAD SALES1

2>LIST

TITLE- ELEDES
          MONTH   HIPER    PRODA    PRODB
1#        1.000   2.000  275.000  236.500
2#        2.000   2.000  292.500  225.000
3#        3.000   2.000  300.000  241.700
4#        4.000   3.000  250.000  475.000
5#        5.000   4.000  262.500  550.000
6#        6.000   4.000  301.500  565.000
7#        7.000   4.000  288.750  535.000
8#        8.000   4.000  306.250  555.000
9#        9.000   4.000  318.750  548.500
10#      10.000   4.000  323.750  550.000
11#      11.000   5.000  257.000  605.000
12#      12.000   5.000  279.500  615.000

3>ELEMENTARY

VARIABLE     MEAN     STD DEV    STD ERR    MAXIMUM    MINIMUM     RANGE
MONTH       6.500       3.606      1.041     12.000      1.000    11.000
HIPER       3.583       1.084       .313      5.000      2.000     3.000
PRODA     287.958      23.741      6.854    323.750    250.000    73.750
PRODB     475.142     149.260     43.088    615.000    225.000   390.000

4>ELEMENTARY TO STATS
  NEW FILE- OK? YES

5>QUIT

-
```

DESCRIPTIVE STATISTICS

STATPAK's DESCRIPTIVE analysis calculates for a single specified variable 18 items of information, including the mean, variance, standard deviation, standard error of the mean, coefficient of variation, range, percentile and quartile data, moment coefficient of skewness, and Pearson's second coefficient of skewness.

To execute the DESCRIPTIVE analysis, the user types

p> DESCRIPTIVE variable ↲

or simply

p> DESCRIPTIVE ↲

and STATPAK requests the name of the variable to be analyzed. The 18 statistics are then calculated and printed.

The user is then asked whether he wishes to see the ordered array, the deviations from the mean ($x-\bar{x}$ for each x, where $\bar{x}$ is the mean), or the standardized values [$(x-\bar{x})/s$ for each x, where s is the standard deviation]. If he indicates that he wants either or both of the last two options but not the ordered array, the data values are printed in the order of their original entry. If an ordered array is requested, the items are listed in ascending order.

Example

```
-STATPAK

1>LOAD COSALES

2>DESCRIPTIVE C1


                 MEAN =   262.059
             VARIANCE =    63.471
   STANDARD DEVIATION =     7.967
       STANDARD ERROR =     1.699
COEFF. OF VARIATION = .304E-01

        MINIMUM =  249.600
10TH PERCENTILE =  252.210
   1ST QUARTILE =  256.750
         MEDIAN =  261.400
   3RD QUARTILE =  267.825
90TH PERCENTILE =  270.750
        MAXIMUM =  280.900

                    RANGE =   31.300
10-90 PERCENTILE RANGE =   18.540
       QUARTILE DEVIATION =    5.537
        AVERAGE DEVIATION =    6.355

 MOMENT COEFF. OF SKEWNESS =      .360
PEARSON COEFF. OF SKEWNESS =      .248
```

```
PRINT ORDERED ARRAY?YES
  DEVIATIONS FROM MEAN?YES
  STANDARDIZED VALUES?YES


      ARRAY     DEVIATIONS    STD.VALUES
      249.600      -12.459        -1.564
      250.300      -11.759        -1.476
      252.100       -9.959        -1.250
      253.200       -8.859        -1.112
      255.500       -6.559         -.823
      256.300       -5.759         -.723
      258.100       -3.959         -.497
      258.300       -3.759         -.472
      259.300       -2.759         -.346
      259.300       -2.759         -.346
      261.400        -.659    -.827E-01
      261.400        -.659    -.827E-01
      262.800         .741     .930E-01
      263.200        1.141          .143
      265.400        3.341          .419
      266.400        4.341          .545
      268.300        6.241          .783
      270.100        8.041         1.009
      270.300        8.241         1.034
      270.800        8.741         1.097
      272.300       10.241         1.285
      280.900       18.841         2.365
3>QUIT

-
```

SECTION 4
PLOTTING

STATPAK contains four commands which provide visual terminal displays of the information contained in the data matrix. The SCATTER command plots two variables on the terminal. The PLOT command produces a graph for one independent variable and as many as three dependent variables.

Histograms may be created within STATPAK. The HISTOGRAM command prints a histogram for any selected variable. The CUMULATIVE analysis produces a cumulative frequency histogram for any selected variable.

Each STATPAK plotting command allows the user to choose the plotting symbol to be used. This symbol can be any keyboard character available on the terminal being used.

SCATTER DIAGRAMS

The SCATTER analysis plots two variables on the terminal, one variable represented by the horizontal axis, and the other represented by the vertical axis. The user may specify the two variables to be plotted and the symbol to be used for the plot by typing:

p>SCATTER variable$_1$,variable$_2$ WITH character

For example,

p>SCATTER X,Y WITH +

instructs STATPAK to plot x versus y with the plot symbol +.

The user may simply type

p> SCATTER

and STATPAK prompts for the necessary specifications.

In either case, the first variable named is represented on the horizontal axis; the second variable named is represented on the vertical axis.

Example

```
-STATPAK
1>LOAD SALES2
2>LIST

TITLE- SALES2
            MONTH    REPR      INV    OHCOST
1#          1.000   2.000  275.000   236.500
2#          2.000   2.000  292.500   225.000
3#          3.000   2.000  300.000   241.700
4#          4.000   3.000  250.000   475.000
5#          5.000   4.000  262.500   550.000
6#          6.000   4.000  302.500   565.000
7#          7.000   4.000  288.750   535.000
8#          8.000   4.000  306.250   555.000
9#          9.000   4.000  318.750   548.500
10#        10.000   4.000  323.750   550.000
11#        11.000   5.000  257.000   605.000
12#        12.000   5.000  279.500   615.000
13#        13.000   5.000  281.000   612.750
14#        14.000   5.000  287.500   621.500
15#        15.000   5.000  295.000   610.000
```

3>SCATTER MONTH,OHCOST WITH *

```
     1.000       4.000       7.000      10.000      13.000      16.000
         + . . . . . + . . . . . + . . . . . + . . . . . + . . . . . +
625.000 +
        .                                            *       *
        .                                        *       *       *
        .
        .
        .
        .                    *
545.000 +                *           *   *   *
        .                        *
        .
        .
        .
        .
        .
465.000 +            *
        .
        .
        .
        .
        .
        .
385.000 +
        .
        .
        .
        .
        .
        .
305.000 +
        .
        .
        .
        .
        .
        .*       *
225.000 +   *
         + . . . . . + . . . . . + . . . . . + . . . . . + . . . . . +
     1.000       4.000       7.000      10.000      13.000      16.000
```

4>QUIT

-

PLOTS

STATPAK's PLOT analysis produces a graph on the terminal for one independent variable and as many as three dependent variables. The independent variable is indicated on the vertical axis; the dependent variable or variables are indicated on the horizontal axis.

The length of the vertical axis varies according to the number of rows in the data matrix but does not exceed nine inches; the horizontal axis is six inches in length.

A different plot symbol must be specified for each dependent variable. If points for two or more dependent variables coincide, the symbol for the variable last specified is printed. For example, if the user specifies that the dependent variables are A, B, and C, the symbol for variable C is printed if a point for all three variables coincides.

To execute the PLOT analysis, the user types

p> PLOT independent variable,dependent variable list ↲

and STATPAK prompts for each of the plot symbols. When the user types

p> PLOT ↲

STATPAK prompts for the independent variable, the dependent variable or variables, and the plot symbols.

NOTE: The plot symbols may be specified only in response to STATPAK prompts and are not part of the general form of the PLOT command.

Example

```
-STATPAK ⮐

1>LOAD ACO ⮐

2>LIST ⮐

TITLE- ACO
              TIME        VAR1        VAR2
1#           1.000      10.000      15.000
2#           2.000      16.000      21.000
3#           3.000      20.000      29.000
4#           4.000      23.000      33.000
5#           5.000      25.000      35.000
6#           6.000      26.000      36.000
7#           7.000      30.000      46.000
8#           8.000      36.000      60.000
9#           9.000      48.000      72.000
10#         10.000      62.000      90.000
11#         11.000      78.000     107.000
12#         12.000      94.000     114.000
13#         13.000     107.000     123.000
14#         14.000     118.000     135.000
15#         15.000     127.000     142.000
```

```
3>PLOT TIME,VAR1,VAR2

PLOT SYMBOL
 FOR VAR1: A
 FOR VAR2: B


   10.000      36.400      62.800      89.200     115.600     142.000
        + . . . . . + . . . . . + . . . . . + . . . . . + . . . . . +
 1.000 .A B
 2.000 .   A B
 3.000 .     A   B
 4.000 .      A   B
 5.000 .       A   B
 6.000 .       A    B
 7.000 .         A      B
 8.000 .            A          B
 9.000 .                 A          B
10.000 .                        A           B
11.000 .                               A            B
12.000 .                                      A        B
13.000 .                                            A      B
14.000 .                                                 A       B
15.000 .                                                     A      B
        + . . . . . + . . . . . + . . . . . + . . . . . + . . . . . +
   10.000      36.400      62.800      89.200     115.600     142.000


4>QUIT

_
```

In some cases, due to the limitations of plotting on the terminal, the STATPAK plot may not represent the data precisely. For example,

```
-STATPAK

1>INPUT

TITLE: DEMO

COLUMN TITLES OR #: INDEP,DEP

    INDEP,DEP
1#  1,15
2#  1.4,17
3#  2.7,21
4#  4.4,34
5#  5.2,45
6#  6,60
7#

2>PLOT IDEP,DEP

PLOT SYMBOL
 FOR DEP: *


      15.000      24.000      33.000      42.000      51.000      60.000
           + . . . . . + . . . . . + . . . . . + . . . . . + . . . . . +
   1.000  .* *
   2.000  .
   3.000  .     *
   4.000  .              *
   5.000  .                          *
   6.000  .                                              *
           + . . . . . + . . . . . + . . . . . + . . . . . + . . . . . +
      15.000      24.000      33.000      42.000      51.000      60.000


3>QUIT

-
```

The second data point (1.4,17) actually lies below the first point, that is, between 1.000 and 2.000 on the independent variable scale. But since the point could not be printed there, it was printed on the line corresponding to the nearest base variable value, 1.000.

HISTOGRAMS

The HISTOGRAM analysis prints a histogram (bar chart) for any selected variable. The histogram illustrates the frequency distribution for the variable requested. Thus, the user can see at a glance what range of values occurs most often in a list of numbers.

The user types

p> HISTOGRAM variable WITH character ↵

or simply

p> HISTOGRAM ↵

and STATPAK requests the variable to be plotted and the plot symbol. After the user enters these specifications, STATPAK prompts for the number of intervals into which the data values are to be divided.

The user may specify any symbol on the keyboard for the histogram; each bar of the histogram is two symbols in width.

A maximum of 12 intervals may be specified. The intervals marked on the histogram include all values from the lower bound up to but not including the upper bound. For example, the interval

```
----------+----------+----------
        10.00      15.00
```

includes 10 and all values between 10 and 15, but not 15.

The example below contains a table of employee numbers and employee pay rates on the file EMPR. The histogram charts employee rates with six intervals.

```
-STATPAK ↵

1>LOAD EMPR ↵

2>LIST ↵

TITLE- HISTOG
             EMPNO       PAYRT
1#          289.000     750.500
2#          391.000     715.000
3#          424.000     780.000
4#          313.000     900.000
5#          243.000     715.500
6#          365.000     600.000
7#          396.000     615.000
8#          356.000     675.500
9#          346.000     500.000
10#         156.000     450.000
11#         278.000     555.000
12#         349.000     575.500
13#         141.000     790.000
14#         245.000     785.000
15#         297.000     915.500
16#         310.000    1000.000
17#         151.000    1500.000
18#         255.000    1200.500
19#         262.000    1100.000
20#         236.000     980.000
21#         357.000     875.500
22#         198.000     860.000
23#         220.000     750.000
24#         300.000     775.000
25#         195.000     715.500
26#         388.000     685.000
27#         381.000     700.000
28#         201.000     900.000
29#         333.000     790.500
30#         220.000     725.000


3>HISTOGRAM PAYRT WITH X ↵
# OF INTERVALS: 6 ↵
```

```
FREQUENCY      6     14     5     3     1     1
--------+----+----+----+----+----+----+----
   14              XX
                   XX
   12              XX
                   XX
   10              XX
                   XX
    8              XX
                   XX
    6         XX   XX
              XX   XX   XX
    4         XX   XX   XX
              XX   XX   XX   XX
    2         XX   XX   XX   XX
              XX   XX   XX   XX   XX   XX
--------+----+----+----+----+----+----+----
        450.000   802.000  1154.000  1506.000
             626.000   978.000  1330.000

4>QUIT
-
```

CUMULATIVE HISTOGRAMS

STATPAK can prepare a cumulative frequency histogram for any selected variable. In this type of histogram, each interval includes the total frequency of all values less than its upper bound. The resulting display is similar to that of the HISTOGRAM analysis.

To access this analysis, the user types

p> CUMULATIVE variable WITH character

and STATPAK prompts for the number of intervals. The user may omit the variable and character specification and simply type:

p> CUMULATIVE

STATPAK prompts for all required information. For a plot symbol, the user may select any symbol on the keyboard; each bar of the histogram is two symbols in width. The maximum number of intervals into which cumulative histogram data values may be divided is 12.

The CUMULATIVE analysis is illustrated below, using the same data file as in the previous example.

```
-STATPAK

1>LOAD EMPR

2>CUMULATIVE PAYRT WITH 1
# OF INTERVALS: 6


CUMULATIVE
FREQUENCY   6    20    25    28    29    30
  --------+----+----+----+----+----+----+----
   30                                     ↑↑
                                    ↑↑    ↑↑
   28                         ↑↑    ↑↑    ↑↑
                              ↑↑    ↑↑    ↑↑
   26                         ↑↑    ↑↑    ↑↑
                        ↑↑    ↑↑    ↑↑    ↑↑
   24                   ↑↑    ↑↑    ↑↑    ↑↑
                        ↑↑    ↑↑    ↑↑    ↑↑
   22                   ↑↑    ↑↑    ↑↑    ↑↑
                        ↑↑    ↑↑    ↑↑    ↑↑
   20             ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
                  ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
   18             ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
                  ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
   16             ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
                  ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
   14             ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
                  ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
   12             ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
                  ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
   10             ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
                  ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
    8             ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
                  ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
    6       ↑↑    ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
            ↑↑    ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
    4       ↑↑    ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
            ↑↑    ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
    2       ↑↑    ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
            ↑↑    ↑↑    ↑↑    ↑↑    ↑↑    ↑↑
  --------+----+----+----+----+----+----+----
       450.000    802.000   1154.000   1506.000
            626.000    978.000   1330.000

3>QUIT

-
```

SECTION 5
CORRELATION ANALYSES

STATPAK provides four correlation analyses. The CORRELATION analysis computes correlation coefficients for each column of data against each other column. The SPEARMAN and the KENDALL analyses measure the degree of correlation among columns ranked according to different criteria.

The correlation coefficient is a number from -1 to 1, inclusive. A correlation coefficient equal to or approximately equal to 1 indicates a high degree of correlation; a correlation coefficient near zero indicates very little correlation. A correlation coefficient equal to or approximately equal to -1 indicates that the data has a highly negative correlation.

The CONTINGENCY analysis enables the STATPAK user to determine whether two variables are statistically independent.

CORRELATION

The CORRELATION analysis computes correlation coefficients for each variable against each other variable. Thus, the user can see at a glance the relationship of one column to any other column. For example, in the data matrix below, the user can see the degree of relationship between sales and the month, or between the sales of PRODA and the sales of PRODB.

To obtain the CORRELATION analysis, the user types

p> CORRELATION

and STATPAK computes the correlation coefficients and prints them on the terminal. If the user wants to save the correlation matrix, he enters:

p> CORRELATION TO file name

The correlation coefficients are computed and automatically saved on the file the user names. In this case, no data is printed on the terminal.

Example

```
-STATPAK

1>LOAD SALFS1

2>LIST


TITLE- ELEDES
      MONTH HIPER    PRODA   PRODB
1#       1      2   275.00   236.5
2#       2      2   292.50   225.0
3#       3      2   300.00   241.7
4#       4      3   250.00   475.0
5#       5      4   262.50   550.0
6#       6      4   301.50   565.0
7#       7      4   288.75   535.0
8#       8      4   306.25   555.0
9#       9      4   318.75   548.5
10#     10      4   323.75   550.0
11#     11      5   257.00   605.0
12#     12      5   279.50   615.0
```

```
3>CORRELATION

CORRELATION MATRIX

 MONTH   HIPER   PRODA   PRODB

MONTH
 1.0000

HIPER
  .9191  1.0000

PRODA
  .1904  -.0308  1.0000

PRODB
  .8526   .9636  -.0076  1.0000


4>QUIT

-
```

SPEARMAN AND KENDALL RANK CORRELATIONS

A column of the data matrix is said to be ranked if its n rows are numbered from 1 to n according to some criterion. If several columns are ranked according to different criteria, the user may wish to know the degree of correlation among these rankings. There are two measures of this correlation: the Spearman rank correlation coefficient, used to compare two columns; and Kendall's coefficient of concordance, used for any number of columns. The ranked matrix must be read into STATPAK or created using the RANK command. The particular correlation calculation is performed on the specified columns of the matrix.

The user requests the Spearman rank correlation coefficient by typing

p>SPEARMAN variable$_1$ variable$_2$

or simply

p>SPEARMAN

and STATPAK prompts for the variables to be correlated, then prints the Spearman rank correlation coefficient, the t-statistic for that coefficient, and the degrees of freedom used to calculate the t-statistic. The t-statistic has n-2 degrees of freedom, n being the number of rows in the data matrix. The user may determine the significance level of a Spearman correlation coefficient by using a table for Student's t-distribution.

Example

```
-STATPAK

1>LOAD DATA

2>LIST

TITLE- RANK
              A        B        C
1#         2.000    4.000    6.000
2#         1.000    3.000    1.000
3#         3.000    1.000    4.000
4#         4.000    2.000    5.000
5#         5.000    6.000    2.000
6#         6.000    5.000    8.000
7#         7.000    8.000    3.000
8#         8.000    7.000    7.000

3>SPEARMAN B,C

SPEARMAN RANK CORRELATION: .095
STUDENTS T: .234   (6  DEGREES OF FREEDOM)

4>QUIT

-
```

To request calculation of the Kendall coefficient of concordance, the user types

p> KENDALL variable list ↲

or merely

p> KENDALL ↲

and STATPAK prompts for the variables the user wants to correlate.

STATPAK prints the Kendall correlation coefficient and corresponding chi-square statistic for that coefficient. The chi-square statistic has n-1 degrees of freedom, where n is the number of rows in the data matrix. If n is greater than 7, the chi-square value can be used to test the correlation hypothesis.

Example

```
-STATPAK ↲

1>LOAD RDATA ↲

2>KENDALL A,B,C ↲

KENDALL COEFFICIENT OF CONCORDANCE: .526
CHI-SQUARE: 12.622  (8  DEGREES OF FREEDOM)

3>QUIT ↲

-
```

CONTINGENCY TABLE

A contingency table consists of a division of objects into the cells of a matrix using one criterion for the rows and another for the columns. There are two conclusions to be tested with a contingency table:

C_1: The two criteria being tested are statistically independent.

C_2: The two criteria being tested are not statistically independent.

It is desirable to control as much as possible the occurrence of a Type I error, the risk of concluding C_2 when, in fact, C_1 is correct.

If there are n classes into which one of the variables is divided and m classes into which the other variable is divided, and the risk of a Type I error is to be controlled, the statistical decision rule using the chi-square statistic is:

If chi-square $\leqslant$ A, conclude C_1.

If chi-square > A, conclude C_2.

where A is the action limit obtained from the chi-square distribution with (m-1)(n-1) degrees of freedom according to the specified risk of Type I error.

The contingency table itself must be entered into STATPAK as the data matrix. To perform the chi-square test, the STATPAK user types CONTINGENCY and a Carriage Return. The program computes and prints the chi-square statistic and returns the user to command level.

For example, a contingency table of hair color and grades for a school exam might appear as follows:

GRADE \ HAIR	Red	Brown	Black	Blonde
A	0	5	1	2
B	1	3	2	1
C	3	17	5	2
D	0	3	3	0
E	1	2	0	0

Total = 51 students

Seventeen students with brown hair got a C on the exam. If a user wanted to test the independence of hair color and grades on the exam, he could use the CONTINGENCY analysis to perform a chi-square test on the data.

Example

```
-STATPAK
1>LOAD STUDENTS
2>LIST

TITLE- STUDEN
    RED BROWN BLACK BLONDE
1#   0     5     1      2
2#   1     3     2      1
3#   3    17     5      2
4#   0     3     3      0
5#   1     2     0      0

3>CONTINGENCY

CHI-SQUARE: 10.31294  WITH 12 DEGREES OF FREEDOM.

4>QUIT

-
```

In the above example, m equals 4 and n equals 5. The action limit for the chi-square distribution with (4-1)(5-1)=12 degrees of freedom is 21.026 for a 5% risk of Type I error. The computed chi-square of 10.313 is less than the chi-square value from the table, indicating that hair color and grades are statistically independent.

SECTION 6
REGRESSION

Regression is a technique used to obtain a functional relationship among variables, where the values of one variable can be measured in terms of the associated variables. Five regression analyses can be performed in STATPAK. For each regression analysis, a complete set of statistics, including a table of residuals, is available. The user may save the coefficients and table of residuals calculated in any of these analyses. He has the additional option of printing the table of residuals at the terminal.

The LINEAR regression analysis fits a set of data to a linear equation of the form $y=A+Bx$, where y is the dependent variable, and x the independent variable. The least squares method is used.

The MULTIPLE regression analysis uses the linear least squares method to fit a curve of the form $y=B_0+B_1x_1+B_2x_2+\cdots+B_kx_k$ to a set of observations of y, the dependent variable, and $x_1, x_2, \ldots, x_k$, the independent variables.

The STEPWISE regression analysis performs a multiple regression using a stepping technique to add independent variables one at a time.

The POLYNOMIAL regression analysis fits a set of data to a polynomial of the form $y=B_0+B_1x+B_2x^2+\cdots+B_kx^k$, where the degree k may be specified by the user. The orthogonal polynomial method is used.

Curve fitting is done with the CURVE analysis. A least squares fit is performed for six types of curves on a specified independent variable and a specified dependent variable.

LINEAR REGRESSION

The functional relationship obtained by linear regression is of the form y=A+Bx, where y is the dependent variable, and x the independent variable. The least squares method is used to estimate the intercept, A, and the regression coefficient, B. Other statistics provided by the LINEAR regression analysis include the correlation coefficient, sum of the squares attributable to the regression, sum of the squares of deviations from the regression, F-value for analysis of variance, standard error of the estimate, standard error of regression coefficient, computed t-value, and table of residuals.

The user requests a LINEAR regression analysis by typing

p> LINEAR independent variable,dependent variable ↩

at the STATPAK command level. STATPAK prompts

COLUMN FOR RESIDUALS:

and the user types a column name to add a column of residuals to his current data matrix; if he does not wish a column of residuals, he types a Carriage Return. STATPAK then requests

COLUMN FOR COEFFICIENTS:

and the user responds with a column name or a Carriage Return.

When STATPAK finishes the LINEAR regression analysis, the user may type the SAVE command to save the column of residuals, the column of coefficients, or any part of the current data matrix.

The user may request a listing of the table of residuals by typing YES and a Carriage Return in response to the STATPAK question

PRINT TABLE OF RESIDUALS?

or he may omit the listing by typing NO and a Carriage Return, or simply a Carriage Return.

STATPAK then asks if the user wishes a plot of the observed values and the regression line. If the user responds by typing YES, STATPAK prints a plot of the observed values with the symbol O and the regression line with the symbol R. The output is similar to that of the PLOT analysis, with the independent variable indicated on the vertical axis. If points for the observed value and the regression line coincide, the O symbol is printed.

The limitations of plotting on the terminal, discussed on page 64, apply in this analysis when the observed y values are plotted. Points may appear to have the same x or y values when, in fact, they are different but have been rounded to the nearest scale value. The regression line, however, can be plotted with little difficulty since the equation for the line is available as the result of the regression analysis. The line is plotted as follows: Using the regression equation, $y=A+Bx$, an estimated y is calculated and plotted for each x scale value. Thus, regardless of the scatter of the original input data, the complete regression line, not just the estimated y values corresponding to the original data, is plotted.

Example

```
-STATPAK

1>LOAD PLOTDATA

2>LINEAR TIME,VAR1
COLUMN FOR RESIDUALS: LINR
COLUMN FOR COEFFICIENTS: LINC


                INTERCEPT =     6.32810    A      y=A + BX
  REGRESSION COEFFICIENT =     1.16537    B

STD. ERROR OF REG. COEF. = .680E-01
        COMPUTED T-VALUE =   17.131

 CORRELATION COEFFICIENT =     .979
  STD. ERROR OF ESTIMATE =    9.134

                ANALYSIS OF VARIANCE FOR THE REGRESSION

   SOURCE OF VARIATION          D.F.   SUM OF SQ.    MEAN SQ.     F VALUE
ATTRIBUTABLE TO REGRESSION       1     .245E+05     .245E+05     293.475
DEVIATION FROM REGRESSION       13     1084.681       83.437
    TOTAL                       14     .256E+05    1826.524
```

```
PRINT TABLE OF RESIDUALS?YES
```

Y OBSERVED	Y ESTIMATED	RESIDUAL	STD.RESIDUAL
3.000	23.809	-20.809	-2.278
15.000	17.982	-2.982	-.326
21.000	24.974	-3.974	-.435
29.000	29.635	-.635	-.696E-01
33.000	33.131	-.131	-.144E-01
35.000	35.462	-.462	-.506E-01
37.000	36.628	.372	.408E-01
46.000	41.289	4.711	.516
60.000	48.281	11.719	1.283
72.000	62.266	9.734	1.066
90.000	78.581	11.419	1.250
107.000	97.227	9.773	1.070
114.000	115.872	-1.872	-.205
123.000	131.022	-8.022	-.878
135.000	143.841	-8.841	-.968

```
PLOT Y OBSERVED AND REGRESSION LINE?YES

        3.000      31.168      59.336      87.505     115.673     143.841
            + . . . . . + . . . . . + . . . . . + . . . . . + . . . . . +
  10.000 .      OR
  17.714 .O        O RO
  25.429 .              OO
  33.143 .                   O     O
  40.857 .                      R
  48.571 .                          R  O
  56.286 .                             R
  64.000 .                                 R   O
  71.714 .                                     R
  79.429 .                                         R  O
  87.143 .                                            R
  94.857 .                                              O R
 102.571 .                                                   R
 110.286 .                                                 O    R
 118.000 .                                                      O   R
            + . . . . . + . . . . . + . . . . . + . . . . . + . . . . . +
        3.000      31.168      59.336      87.505     115.673     143.841
```

```
3>SAVE LINREG
  NEW FILE- OK? YES
```

The user saves the original matrix plus the column of residuals and the column of coefficients on the file LINREG.

```
4>QUIT

-
```

MULTIPLE REGRESSION

STATPAK's MULTIPLE regression analysis uses the linear least squares method to fit a curve of the form $y=B_0+B_1x_1+B_2x_2+\cdots+B_kx_k$ to a number of sets of observations of y, the dependent variable, and $x_1, x_2, \ldots, x_k$, the independent variables. While linear regression considers only two variables, one independent and the other dependent, multiple regression allows consideration of two or more independent variables.

To initiate the MULTIPLE regression analysis, the user types

p> MULTIPLE dependent variable,independent variable list ⤶

and STATPAK prompts for a column name for the residuals and a column name for the coefficients. If the user enters a column name, STATPAK creates a column containing that data. When the analysis is complete, the user may save all or part of the current data with the SAVE command. If the user does not wish to save the residuals or the coefficients, he types NO or a Carriage Return in answer to the appropriate question.

After the computations are performed and the statistics printed, STATPAK asks if the variance of the regression coefficient and the beta coefficients are to be printed, and if the Durbin-Watson statistic is to be computed. The user is then asked if he wants to see the table of residuals.

Example

```
-STATPAK

1>LOAD INFILE

2>COLS
    INDEP1,INDEP2,INDEP3,INDEP4,INDEP5,DEPVAR

3>MULTIPLE DEPVAR,INDEP1,INDEP2,INDEP3,INDEP4,INDEP5
COLUMN FOR RESIDUALS: MULTR
COLUMN FOR COEFFICIENTS: MULTC


             R-SQUARED =      .631
       F-VALUE( 5,  24) =    8.19959
STD. ERROR OF ESTIMATE =    11.9273

             INTERCEPT =     105.223
```

VARIABLE	COEFFICIENT	T-VALUE
INDEP1	-.649958	-3.77477
INDEP2	-.471735	-.283774
INDEP3	1.94162	4.95395
INDEP4	-7.176782E-03	-.346953
INDEP5	-.238574	-3.70846

ANALYSIS OF VARIANCE FOR THE REGRESSION

SOURCE OF VARIATION	D.F.	SUM OF SQ.	MEAN SQ.
ATTRIBUTABLE TO REGRESSION	5	5832.40	1166.48
DEVIATION FROM REGRESSION	24	3414.26	142.261
TOTAL	29	9246.67	318.851

PRINT VARIANCE OF REG. AND BETA COEFFICIENTS?YES

VARIABLE	VARIANCE	BETA COEFFICIENT
INDEP1	2.964760E-02	-.581502
INDEP2	2.76345	-3.969319E-02
INDEP3	.153613	.709147
INDEP4	4.278759F-04	-4.599132E-02
INDEP5	4.138648E-03	-.486741

COMPUTE DURBIN-WATSON?YES

DURBIN-WATSON = 1.496

PRINT TABLE OF RESIDUALS?YES

Y OBSERVED	Y ESTIMATED	RESIDUAL
85.0000	85.5949	-.594912
92.0000	92.8824	-.882429
90.0000	90.3968	-.396796
91.0000	91.5635	-.563522
95.0000	99.3285	-4.32848
95.0000	97.8704	-2.87044
100.000	107.579	-7.57902
79.0000	94.0251	-15.0251
126.000	114.042	11.9577
95.0000	90.1700	4.83001
110.000	111.056	-1.05561
88.0000	98.8440	-10.8440
129.000	119.218	9.78174
97.0000	95.7234	1.27664
111.000	113.319	-2.31895
94.0000	97.2016	-3.20163
96.0000	96.6932	-.693202
88.0000	80.6061	7.39395
147.000	131.002	15.9978
105.000	96.6428	8.35720
132.000	110.314	21.6863
108.000	96.0231	11.9769
101.000	111.241	-10.2408
136.000	128.858	7.14182
113.000	111.349	1.65086
88.0000	120.457	-32.4569
118.000	134.624	-16.6242
116.000	112.774	3.22627
140.000	127.767	12.2332
105.000	112.834	-7.83437

4>SAVE MULREG
NEW FILE- OK? YES

5>QUIT

-

STEPWISE REGRESSION

The STEPWISE regression analysis performs a multiple regression using a stepping technique to add independent variables one at a time. The criterion for choosing the independent variable for the next successive step is the F-statistic. The new F-statistic has an associated alpha between 0 and 1 that tests the significance level of the regression. The smaller the alpha the better the fit. The terminating criteria for the regression include manual, alpha less than a specified value, alpha increases, alpha changes by less than a specified amount, and an upper limit on alpha.

An option is provided for another type of F-test, that is, an F-statistic for the newly chosen variable. The associated alpha increases continuously as less significant variables are added. In this case, the stopping criteria are manual and alpha greater than a specified value.

The user selects the STEPWISE regression analysis and specifies the dependent variable and the independent variables to be included in every step by typing

p>STEPWISE dependent variable,independent variable list

where the independent variable list contains the independent variables to be included in each step. STATPAK prompts for the independent variables available for successive steps. When all the variables have been specified, STATPAK prompts for a column name for the residuals, then a column name for the regression coefficients. The user specifies a column name for each, and STATPAK creates the new columns of data which the user may explicitly save with the SAVE command at the end of the analysis. If the user does not wish a column of data for either or both of these parameters, he types a Carriage Return in response to the appropriate question.

STATPAK then asks if a list of stopping criteria should be printed, and requests the number of the stopping code. The regression is performed, and the program stops each time the criterion for stopping is met.

After the stepping procedure is completed, the user is asked if the last variable should be included, and if he wants to see the variance of the regression coefficients and the beta coefficients, the Durbin-Watson statistic, and the table of residuals.

Example

```
-STATPAK
1>LOAD INFILE
2>STEPWISE DEPVAR,INDEP1
OTHER IND. VARIABLES: INDEP3,INDEP4,INDEP5
COLUMN FOR RESIDUALS:
COLUMN FOR COEFFICIENTS:
DO YOU WANT LIST OF STOPPING CRITERIA?YES

     1- MANUAL

F-TEST FOR ENTIRE REGRESSION
     2- LOWER LIMIT ON ALPHA
     3- INCREASE IN ALPHA
     4- LIMIT ON CHANGE IN ALPHA

F-TEST FOR NEWLY CHOSEN VARIABLE
     5- UPPER LIMIT ON ALPHA

CODE # FOR STOPPING:1

STEP    1
  VARIABLE   INDEP3 SELECTED
```

Each calculation includes the independent variable INDEP1.

For each successive step, STATPAK selects one of these independent variables.

The user does not wish to save the residuals or coefficients.

Successive independent variables are selected on the basis of a computed F-statistic.

FOR REGRESSION:F(2, 27) = 9.7208
ALPHA= .001

FOR NEW VARIABLE: F(1, 27) = 18.73051
ALPHA = .000

R-SQUARED = .419

INTERCEPT = 38.1448

VARIABLE	COEFFICIENT	T-VALUE
INDEP1	-.527091	-2.81861
INDEP3	1.98253	4.32788

ANALYSIS OF VARIANCE FOR THE REGRESSION

SOURCE OF VARIATION	D.F.	SUM OF SQ.	MEAN SQ..
ATTRIBUTABLE TO REGRESSION	2	3870.88	1935.44
DEVIATION FROM REGRESSION	27	5375.78	199.103
TOTAL	29	9246.67	318.851

STOP?NO

STEP 2

VARIABLE INDEP5 SELECTED

FOR REGRESSION:F(3, 26) = 14.532
ALPHA= .000

FOR NEW VARIABLE: F(1, 26) = 14.46205
ALPHA = .001

R-SQUARED = .626

INTERCEPT = 99.5954

VARIABLE	COEFFICIENT	T-VALUE
INDEP1	-.669975	-4.25888
INDEP3	1.97305	5.27261
INDEP5	-.232277	-3.80290

ANALYSIS OF VARIANCE FOR THE REGRESSION

SOURCE OF VARIATION	D.F.	SUM OF SQ.	MEAN SQ.
ATTRIBUTABLE TO REGRESSION	3	5792.31	1930.77
DEVIATION FROM REGRESSION	26	3454.36	132.860
TOTAL	29	9246.67	318.851

```
STOP?YES

STEPPING COMPLETED

INCLUDE LAST VARIABLE?YES


            R-SQUARED =     .626
       F-VALUE( 3,  26) =   14.53237
STD. ERROR OF ESTIMATE =  11.5265

             INTERCEPT =    99.5954


VARIABLE        COEFFICIENT        T-VALUE

INDEP1           -.669975          -4.25888
INDEP3            1.97305           5.27261
INDEP5           -.232277          -3.80290


          ANALYSIS OF VARIANCE FOR THE REGRESSION

   SOURCE OF VARIATION         D.F.     SUM OF SQ.       MEAN SQ.
ATTRIBUTABLE TO REGRESSION      3       5792.31          1930.77
DEVIATION FROM REGRESSION      26       3454.36          132.860
     TOTAL                     29       9246.67          318.851


PRINT VARIANCE OF REG. AND BETA COEFFICIENTS?YES

VARIABLE        VARIANCE         BETA COEFFICIENT

INDEP1          2.474717E-02       -.599411
INDEP3           .140032            .720626
INDEP5          3.730643E-03       -.473894


COMPUTE DURBIN-WATSON?YES

DURBIN-WATSON =  1.598


PRINT TABLE OF RESIDUALS?NO

3>QUIT

-
```

POLYNOMIAL REGRESSION

The POLYNOMIAL regression analysis fits a set of data to a polynomial of the form $y=C_0+C_1x+C_2x^2+\cdots+C_kx^k$, where the degree of the polynomial, k, may be specified by the user. As the order of the polynomial is increased, the fit becomes better until the optimum fit is obtained. The orthonormalization method is used.

To perform a POLYNOMIAL regression analysis, the STATPAK user types

p> POLYNOMIAL independent variable,dependent variable ↩

and STATPAK prompts for a column of weights to be interpreted as frequencies of observations for the data in each respective row. The user types the corresponding variable name or, if he has no column of weights, a Carriage Return.

STATPAK next prompts for a column name for the residuals, and after the user responds, prompts for a column name for the regression coefficients. If the user provides a column name, STATPAK creates a new column of data for each column named. If the user does not wish to create new columns of data, he types a Carriage Return. When the analysis is completed, new columns may be saved with the SAVE command.

STATPAK then requests the degree of the polynomial to be fit. The degree entered must be less than the number of rows in the data matrix. The regression coefficients, index of determination, and standard error of the estimate for y are printed.

The user is asked if he wishes to see a table of residuals. The user may fit another polynomial to the data by entering its degree, and the analysis procedure is repeated. If he does not wish to fit another polynomial and has specified a name for a column of residuals and/or coefficients, STATPAK asks the degree of the polynomial for which the user wants to save the residuals and/or coefficients; the user types the degree of the polynomial and a Carriage Return.

Example

```
-STATPAK

1>LOAD POLYDATA

2>POLYNOMIAL IND,DEP
  COLUMN OF WEIGHTS: WEIGHT
COLUMN FOR RESIDUALS: POLYR
COLUMN FOR COEFFICIENTS: PCOEFS

DEGREE OF POLYNOMIAL:2

POWER OF X   COEFFICIENT

     0         -.117419

     1          9.73049

     2         -1.49064

INDEX OF DETERMINATION:    .920306
STANDARD ERROR OF ESTIMATE FOR Y:    5.79444


PRINT A TABLE OF RESIDUALS?YES

Y OBSERVED      Y ESTIMATED        RESIDUAL

 -52.0000         -42.7247         -9.27533
 -20.0000         -25.5410          5.54097
 -4.00000         -11.3386          7.33855
  2.00000         -.117419          2.11742
  4.00000          8.12243         -4.12243
  8.00000          13.3810         -5.38100
  20.0000          15.6583          4.34171

ANOTHER FIT FOR THIS DATA?YES

DEGREE OF POLYNOMIAL:3

POWER OF X   COEFFICIENT

     0          2.00000

     1          3.00000

     2         -2.00000

     3          1.00000

INDEX OF DETERMINATION:    1.00000
STANDARD ERROR OF ESTIMATE FOR Y:    .000000
```

```
PRINT A TABLE OF RESIDUALS?NO

ANOTHER FIT FOR THIS DATA?NO

SAVE RESIDUALS/COEFFS. FOR DEGREE #3

3>SAVE POLREG
  NEW FILE- OK? Y

4>QUIT

-
```

CURVE FITTING

The CURVE analysis performs a least squares fit for six types of curves on a specified independent and a specified dependent variable. The six curve types a listed below, where y is the dependent variable, and x the independent variable.

$y = A + (Bx)$	$y = A + (B/x)$
$y = A(x^B)$	$y = 1/(A+(Bx))$
$y = A(e^{Bx})$	$y = x/(A+(Bx))$

To perform curve fitting, the user types:

p> CURVE independent variable,dependent variable

STATPAK prompts for a column of weights, and the user types a column name or, if no column of weights exists in the matrix, a Carriage Return. The values in a column of weights are frequencies of observations.

Next, STATPAK prompts for a column name for residuals and a column name for coefficients. If the user wishes to save either the residuals or the coefficients, he must supply a new column name in response to the corresponding prompt. STATPAK then prints a table containing the general curve equations, the calculated values of A and B, and the index of determination for each curve. The closer the value of the index of determination is to 1, the better the fit to the respective curve equation.

The calculations are performed by appropriate transformations of the variables. A simple linear regression is then used to calculate A and B. If the data cannot be fit to a particular equation, a message is printed in that row of the table.

The next system prompt asks if the user wishes STATPAK to print a table of residuals. The user responds YES or NO, as appropriate. If the user responds by typing YES and a Carriage Return, the system asks for which curve he wishes the residuals. The user identifies the curve by number, and the table of residuals is printed. Finally, if the user named a column for residuals and/or coefficients prior to the printout of the analysis results, STATPAK asks for which curve he wishes to save the residuals and/or coefficients. If the user did not specify a column for either the residuals or the coefficients, control is transferred to the STATPAK command level and no results may be saved.

Example

```
-STATPAK
1>LOAD DATAFILE
2>CURVE B,A
COLUMN OF WEIGHTS:
COLUMN FOR RESIDUALS: CFITR
COLUMN FOR COEFFICIENTS: CFITC

                 LEAST SQUARES CURVES FIT

                    INDEX  OF
 CURVE TYPE       DETERMINATION        A               B

1. Y=A+(B*X)       5.962459E-02    5.94406         -.251748
2. Y=A*(X↑B)       5.289279E-03    4.94214       -5.177359E-02
3. Y=A*EXP(B*X)    9.367934E-02    6.06018       -7.076120E-02
4. Y=A+(B/X)       3.743336E-02    5.51825         -1.42628
5. Y=1/(A+B*X)      .135872         .150591       2.230270E-02
6. Y=X/(A+B*X)     4.110713E-03   2.773789E-02     .224147
```

```
PRINT A TABLE OF RESIDUALS?YES

FOR WHAT CURVE #5

RESIDUALS FOR 5. Y=1/(A+B*X):

Y OBSERVED      Y ESTIMATED        RESIDUAL

  3.00000          5.78390         -2.78390
  4.00000          5.43345         -1.43345
  5.00000          5.12304         -.123043
  6.00000          4.84619          1.15381
  7.00000          4.59772          2.40228
  8.00000          4.37349          3.62651
  7.00000          4.17011          2.82989
  6.00000          3.98481          2.01519
  5.00000          3.81527          1.18473
  4.00000          3.65957          .340427
  3.00000          3.51608         -.516085
  2.00000          3.38342         -1.38342
ANOTHER CURVE?NO

SAVE RESIDUALS/COEFFS. FOR CURVE #5

3>SAVE DATAFIT

NEW FILE- OK? YES

4>QUIT

-
```

SECTION 7
GOODNESS-OF-FIT

A goodness-of-fit statistic is used to test the degree of conformity of a particular variable with a particular theoretical distribution. STATPAK contains two analyses to provide goodness-of-fit statistics: the CHI-SQUARE analysis, which produces a chi-square statistic; and the KOLMOGOROV-SMIRNOV analysis, which produces the Kolmogorov-Smirnov statistic. Both analyses contain a set of standard theoretical distributions including uniform, binomial, normal, exponential, and Poisson. In addition, the user may enter his own theoretical distribution, entering expected as well as observed frequencies.

DATA INPUT

For the goodness-of-fit tests, the user may enter the data prior to calling the goodness-of-fit routine, or he may enter data within the analysis. If the data is already in STATPAK, it may contain individual observations identical to the data used in all other STATPAK analyses, or it may contain frequencies of observations as one column in the matrix. If the user enters the data within the analysis, he enters frequency of observations for each interval.

The user calls the goodness-of-fit analysis and, if no data has been entered, STATPAK asks which distribution to use. If there is a data matrix, STATPAK asks which column to read and then asks if the column contains frequencies of observations. Next, STATPAK requests the selected distribution. If the user specifies a normal distribution, STATPAK requests values for the mean and standard deviation; if the user specifies a Poisson or exponential distribution, STATPAK requests the value of the mean; if the user specifies a binomial distribution, STATPAK requests the probability of success and the number of trials; and if the user specifies a uniform distribution, STATPAK requests the lower limit and the upper limit.

The next STATPAK prompt requests a column for the residuals. If the user specifies a column name, the residuals are stored in that column and may be saved with the SAVE command when control returns to the STATPAK command level. A Carriage Return response to this request instructs STATPAK to omit the column of residuals from the data matrix.

After the source and form of the data and the distribution are specified, STATPAK requests the interval specifications. The terms used to describe the desired intervals are FROM, BY, IN, TO, and the semicolon (;). A semicolon must appear as the final character in the interval specification. FROM begins the specification by identifying the lower limit; BY indicates the actual increment for the interval; IN specifies the number of intervals; and TO identifies the upper limit. Using a combination of these terms, the user may describe an interval structure as simple or as complex as desired. For example, with values from 0 to 25, and the interval format

0 1 2 3 4 5 7 9 11 13 15 20 25

can be obtained from the following interval specification:

FROM 0 BY 1 TO 5 BY 2 TO 15 IN 2 TO 25;

The STATPAK conventions for creating intervals provide that each interval contains all those values which are equal to or greater than the interval's lower limit through those which are less than its upper limit. The final interval, however, contains values equal to or greater than its lower limit and equal to or less than its upper limit. External intervals, on both sides, include all values not included previously. When running the CHI-SQUARE test, the user is asked if external intervals are to be considered and if an additional interval is desired. The additional interval is used to prevent the grouping effect in the last interval. See the example below.

The interval conventions apply as follows: In a range of data from 0 to 20 with desired intervals of 5, the interval specification is

FROM 0 BY 5 TO 20;

and the intervals created are:

External interval	<0
Interval 1	$\geqslant 0$, <5
Interval 2	$\geqslant 5$, <10
Interval 3	$\geqslant 10$, <15
Interval 4	$\geqslant 15$, $\leqslant 20$
External interval	>20

If the additional interval option is selected, then:

External interval	<0
Interval 1	$\geq 0, <5$
Interval 2	$\geq 5, <10$
Interval 3	$\geq 10, <15$
Interval 4	$\geq 15, <20$
Additional interval	$\geq 20, <21$
External interval	≥ 21

CHI-SQUARE TEST

The STATPAK user requests a CHI-SQUARE analysis by typing CHI and a Carriage Return. After he responds appropriately to each prompt, the expected and observed frequencies are printed, and the chi-square statistic is computed and printed. The system then asks if the user wishes to see a table of residuals.

Example 1

```
-STATPAK

1>CHI
```

The user calls the CHI-SQUARE analysis without previously entering any data.

```
DISTRIBUTION: POISSON
MEAN: 10.44
COLUMN FOR RESIDUALS: CHIRES

INTERVAL SPECIFICATIONS:
FROM 0 BY 2 TO 2 BY 1 TO 22;

DO YOU WISH A SEPARATE INTERVAL FOR 22 ?YES

CONSIDER EXTERNAL INTERVALS?NO

MINIMUM EXPECTED FREQUENCY:0
```

The user enters his data within the CHI-SQUARE analysis.

Each input is the frequency for the printed interval.

```
ENTER FREQUENCY FOR THE FOLLOWING INTERVALS:

>= .00000    ,< 2.00000 :5
>= 2.00000   ,< 3.00000 :14
>= 3.00000   ,< 4.00000 :24
>= 4.00000   ,< 5.00000 :57
>= 5.00000   ,< 6.00000 :111
>= 6.00000   ,< 7.00000 :197
>= 7.00000   ,< 8.00000 :278
>= 8.00000   ,< 9.00000 :378
>= 9.00000   ,< 10.00000 :418
>= 10.00000   ,< 11.00000 :461
>= 11.00000   ,< 12.00000 :433
>= 12.00000   ,< 13.00000 :413
>= 13.00000   ,< 14.00000 :358
>= 14.00000   ,< 15.00000 :219
>= 15.00000   ,< 16.00000 :145
>= 16.00000   ,< 17.00000 :109
>= 17.00000   ,< 18.00000 :57
>= 18.00000   ,< 19.00000 :43
>= 19.00000   ,< 20.00000 :16
>= 20.00000   ,< 21.00000 :7
>= 21.00000   ,< 22.00000 :8
>= 22.00000   ,< 23.00000 :3

TOTAL OF EXPECTED FREQUENCIES DO NOT EQUAL
OBSERVED.  EXPECTED=3752.02 , OBSERVED=3754

CHI-SQUARE STATISTIC=   43.1650

PRINT A TABLE OF RESIDUALS?NO

2>QUIT
```

Example 2

```
-STATPAK

1>LOAD CHISQ
```

The user enters an existing data matrix.

```
2>CHI

COLUMN: FREQ
DOES COLUMN CONTAIN FREQUENCY OF OBSERVATIONS? YES
DISTRIBUTION: POISSON
MEAN: 10.44
COLUMN FOR RESIDUALS: CHIRES

INTERVAL SPECIFICATIONS:
FROM 0 BY 2 TO 2 BY 1 TO 22;

DO YOU WISH A SEPARATE INTERVAL FOR 22 ?YES

CONSIDER EXTERNAL INTERVALS?NO

MINIMUM EXPECTED FREQUENCY:0

TOTAL OF EXPECTED FREQUENCIES DO NOT EQUAL
OBSERVED.  EXPECTED=3752.02 , OBSERVED=3754

CHI-SQUARE STATISTIC=   43.1650

PRINT A TABLE OF RESIDUALS?NO


3>SAVE CHIANS
```

The user saves the entire data matrix, including the column of residuals, on the file CHIANS.

```
NEW FILE- OK? YES

4>QUIT

-
```

Example 3

The user enters a matrix containing a column of individual observations rather than frequencies.

```
-STATPAK

1>LOAD INF

2>CHI

COLUMN: OBS
DOES COLUMN CONTAIN FREQUENCY OF OBSERVATIONS? NO
DISTRIBUTION: UNIFORM
LOWER LIMIT: 1
UPPER LIMIT: 6
COLUMN FOR RESIDUALS: RESIDS

INTERVAL SPECIFICATIONS:
FROM 1 BY 1 TO 6;

CONSIDER EXTERNAL INTERVALS?NO

MINIMUM EXPECTED FREQUENCY:5

CHI-SQUARE STATISTIC=   7.40000

PRINT A TABLE OF RESIDUALS?YES
```

INTERVAL				OBSVD. FREQ.	EXPCTD. FREQ.	ABSOLUTE DEVIATION	WGHTD. SQD. DEVIATION
>=	1.000	,<	2.000	5	10.00	5.000	2.500
>=	2.000	,<	3.000	11	10.00	1.000	.1000
>=	3.000	,<	4.000	14	10.00	4.000	1.600
>=	4.000	,<	5.000	14	10.00	4.000	1.600
>=	5.000	,<=	6.000	6	10.00	4.000	1.600

```
3>SAVE OBS,RESIDS ON CHIRES

NEW FILE- OK? YES

4>QUIT

-
```

KOLMOGOROV-SMIRNOV TEST

The KOLMOGOROV-SMIRNOV analysis is a goodness-of-fit test used for the same purpose as the CHI-SQUARE test: to test observed data against some theoretical distribution.

The Kolmogorov-Smirnov test compares the observed cumulative probability with the theoretical cumulative probability and finds the maximum deviation, D, between the two. The value D can then be used with a table of D-values to find the confidence level of the fit.

The user specifies intervals for the test, as described on page 92. External intervals are automatically considered since the Kolmogorov-Smirnov test is cumulative. At the end points of each interval, the deviation between the observed cumulative probability and the theoretical cumulative probability is calculated. The maximum deviation is chosen from the deviation at each interval.

Example

```
-STATPAK

1>LOAD INF

2>KOLMOGOROV

COLUMN: OBS
DOES COLUMN CONTAIN FREQUENCY OF OBSERVATIONS? NO
DISTRIBUTION: UNIFORM
LOWER LIMIT: 1
UPPER LIMIT: 6
COLUMN FOR RESIDUALS: KOLRES

INTERVAL SPECIFICATIONS:
FROM 1 BY 1 TO 6;

KOLMOGOROV-SMIRNOV STATISTIC:     .100000

PRINT A TABLE OF RESIDUALS?YES
```

INTERVAL				OBS. CUM. FREQ.	OBS. CUM. PROB.	EXP. CUM. PROB.	ABSOLUTE DEVIATION
<	1.000			0	.00000	.00000	.00000
>=	1.000	,<	2.000	5	.10000	.20000	.10000
>=	2.000	,<	3.000	16	.32000	.40000	.08000
>=	3.000	,<	4.000	30	.60000	.60000	.00000
>=	4.000	,<	5.000	44	.88000	.80000	.08000
>=	5.000	,<=	6.000	50	1.00000	1.00000	.00000
>	6.000			50	1.00000	1.00000	.00000

```
3>SAVE KOLMOG

NEW FILE- OK? YES

4>QUIT

-
```

SECTION 8
CONFIDENCE LIMITS

The CONFIDENCE analysis computes the confidence limits on the mean of a set of values. For example, in a particular column of the data matrix, one could say that with a 95% confidence level, the mean lies between 6.7 and 8.3. This is a two-sided confidence interval. A one-sided confidence interval predicts, with the specified confidence, that the mean is less than or greater than some specified value.

The user calls the CONFIDENCE analysis by typing:

p><u>CONFIDENCE variable</u> ↵

STATPAK computes and prints the mean, standard deviation, and standard error for that variable.

The user then specifies the type of confidence interval desired, one-sided or two-sided, and enters the confidence level or levels to be computed. After the bounds for the requested confidence levels are printed, control is returned to STATPAK command level.

Example

```
-STATPAK

1>LOAD EXAM

2>LIST


TITLE- EXAFIT
         C1       C2       C3
1#      5.5     7.80      9.2
2#     -6.3     8.99      4.0
3#     62.7    34.10    -10.0
4#     50.0    30.00     20.0
5#     12.0   -34.20     32.1
6#      9.0     4.00      1.0
7#     -3.0     4.00     51.0
8#     71.0    17.00    -14.0
9#      1.0     5.00     51.0
10#     2.0     4.00      6.0
11#    81.0    32.00    -41.0
12#     6.0     8.00     10.0
13#    32.0    55.00     61.0
14#    12.0    54.00     71.0
15#     3.0     5.00     91.0


3>CONFIDENCE C2


MEAN:    15.6460
STANDARD DEVIATION:    22.5486
STANDARD ERROR OF MEAN:    5.82202

ONE-SIDED OR TWO-SIDED TEST:TWO

CONFIDENCE LEVEL(S) (%):95,90


CONFIDENCE LEVEL 95.00 %, TWO-SIDED TEST
LOWER BOUND:    3.15894
UPPER BOUND:    28.1331


CONFIDENCE LEVEL 90.00 %, TWO-SIDED TEST
LOWER BOUND:    5.39159
UPPER BOUND:    25.9004
```

```
4>CONFIDENCE C2

MEAN:    15.6460
STANDARD DEVIATION:    22.5486
STANDARD FRROR OF MEAN:    5.82202

ONE-SIDFD OR TWO-SIDED TEST:ONE

CONFIDENCE LEVEL(S) (%):90


CONFIDENCE LEVEL 90.00 %, ONE-SIDED TEST
LOWER BOUND:    7.81520
UPPER BOUND:    23.4768

5>QUIT

-
```

SECTION 9
THE XPOS TIME SERIES ANALYSIS

The XPOS program forecasts the future values of a variable, based on as many as 100 past observations, by calculating a set of forecasting parameters and smoothing coefficients. The XPOS program uses the technique of exponentially weighted moving averages and incorporates linear trends and seasonal factors.[1]

The user initiates the XPOS analysis by typing

p> XPOS column name ↵

at STATPAK command level. STATPAK prompts for the initial periods, and the user enters the number of periods for XPOS to use in calculating the initial values for the forecasting parameters and smoothing coefficients. The number of periods may be as many as 60, must be greater than the number of periods in a year, must be less than the total number of periods, and must be a multiple of the number of periods per year. Thus, if there are 12 periods in a year, the number of periods or observations used to calculate the initial parameters may be 24, 36, 48, or 60.

Initial values for the forecasting parameters and smoothing coefficients are based on part of the past observations. Using these initial values, XPOS calculates a forecast for the next period, compares the computed value with the observed value, and updates the forecasting parameters accordingly. This process is repeated until the series of past observations is exhausted.

---

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

The next STATPAK prompt is:

OF SEASONS:

The user defines a time period by entering the number of periods per year. He enters 4 to indicate quarterly periods, 12 to indicate monthly periods, and so on.

STATPAK then requests:

COLUMN FOR PARAMETERS:

If the user wishes to save the parameters calculated by the XPOS analysis, he must enter a column name; otherwise he types only a Carriage Return. Then STATPAK prompts:

COLUMN FOR FORECASTS:

The user must enter a column name if he wishes XPOS to calculate forecasts. When the user enters the column name, XPOS asks for the number of forecasts the user wants.

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

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

This sum is computed for all possible combinations of the three smoothing coefficients between 0 and 1, in intervals of .1. The combination of the three coefficients which minimizes the sum of squared deviations is termed the optimal set. XPOS then calculates forecasts based on the smoothed average, trend factor, and seasonal factor calculated with the optimal smoothing coefficients.

The next two STATPAK prompts ask the user what he wishes printed. He may request a listing of the past smoothed series and/or a listing of computed parameters. If the user types YES and a Carriage Return in response to the STATPAK prompt

PRINT PAST SMOOTHED SERIES:

XPOS prints the past observations with the current trend and seasonal factors, deseasonalized data, and forecasts. If the user types YES and a Carriage Return after STATPAK prompts

PRINT COMPUTED PARAMETERS:

XPOS prints the following:

SO	The most recent deseasonalized and smoothed average.
R	The most recent estimate of the trend factor. R reflects a weighted rate of increase or decrease of deseasonalized observations. The trend factor is a linear contributor in the forecasting equation.
A,B,C	The smoothing constants with values between 0 and 1. A is a coefficient in the equation used to calculate SO; B is a coefficient in the equation used to calculate the seasonal factors; and C is a coefficient in the equation used to calculate the trend factor.
Standard error of forecast	

If the user supplied column names for them, the parameters calculated by the XPOS analysis and any requested forecasts are in core as additional columns of the data matrix and may be listed and/or saved.

Example

```
-STATPAK
1>LOAD XP
2>LIST

TITLE- XPDATA
       PRODA   PRODB
1#     343.9   101.0
2#     366.3   102.0
3#     355.8   103.5
4#     395.4   105.4
5#     398.7   104.0
6#     531.4   104.5
7#     271.3   106.0
8#     256.7   106.0
9#     301.1   107.3
10#    377.8   106.2
11#    354.8   103.0
12#    376.1   108.2
13#    349.6   109.0
14#    380.1   109.2
15#    364.7   109.3
16#    398.3   109.7
17#    399.7   107.6
18#    546.6   108.7
19#    276.2   109.2
20#    253.3   105.6
21#    334.5   107.0
22#    337.6   107.2
23#    374.6   107.1
24#    390.2   108.0
25#    362.0   102.9
26#    395.3   104.5
27#    392.9   106.2
28#    401.0   107.0
29#    429.8   106.5
30#    561.4   108.0
31#    303.5   108.0
32#    270.2   108.2
33#    349.2   106.5
34#    380.9   106.3
35#    419.0   107.8
36#    409.9   109.0
```

3>XPOS PRODA

INITIAL PERIODS: 24
OF SEASONS: 12
COLUMN FOR PARAMETERS: XPPAR
COLUMN FOR FORECASTS:FORECS
OF FORECASTS: 12

PRINT PAST SMOOTHED SERIES : YES

PRINT COMPUTED PARAMETERS : YES

PERIOD T	SEASON J	AVERAGE SO	TREND R	SEASONAL F(J)	ACTUAL S(T)	FORC BASED ON PREV PER S(T-1,1)
1	1	360.7	.4014	.9549	343.9	.0000
2	2	359.8	.1573	1.0207	366.3	.0000
3	3	359.5	.6469E-01	.9906	355.8	.0000
4	4	360.0	.1340	1.0976	395.4	.0000
5	5	360.6	.2397	1.1043	398.7	.0000
6	6	360.4	.1505	1.4759	531.4	.0000
7	7	360.7	.1683	.7521	271.3	.0000
8	8	362.1	.4227	.7070	256.7	.0000
9	9	359.3	-.2275	.8444	301.1	.0000
10	10	364.5	.8589	1.0249	377.8	.0000
11	11	363.6	.5056	.9798	354.8	.0000
12	12	363.3	.3505	1.0370	376.1	.0000
13	1	364.1	.4500	.9590	349.6	.0000
14	2	366.1	.7618	1.0346	380.1	.0000
15	3	367.2	.8117	.9928	364.7	.0000
16	4	366.9	.6078	1.0879	398.3	.0000
17	5	366.4	.3833	1.0935	399.7	.0000
18	6	367.5	.5246	1.4850	546.6	.0000
19	7	367.9	.4921	.7510	276.2	.0000
20	8	366.4	.8843E-01	.6945	253.3	.0000
21	9	372.4	1.276	.8875	334.5	.0000
22	10	364.8	-.4948	.9453	337.6	.0000
23	11	367.9	.2259	1.0105	374.6	.0000
24	12	369.8	.5506	1.0516	390.2	.0000
25	1	371.7	.8364	.9708	362.0	355.1
26	2	374.5	1.216	1.0514	395.3	385.5
27	3	379.7	2.018	1.0263	392.9	373.0
28	4	379.1	1.494	1.0638	401.0	415.3
29	5	383.1	1.992	1.1162	429.8	416.2
30	6	383.7	1.711	1.4676	561.4	571.8
31	7	389.1	2.460	.7742	303.5	289.4
32	8	391.1	2.358	.6916	270.2	272.0
33	9	393.4	2.359	.8875	349.2	349.2
34	10	397.2	2.644	.9562	380.9	374.2
35	11	402.8	3.235	1.0342	419.0	404.1
36	12	402.8	2.584	1.0244	409.9	427.0

SO = 402.8
R= 2.584
A= .2000
B= .8000
C= .2000

SEASONAL FACTORS

```
F(  1 ) =     .9708
F(  2 ) =     1.051
F(  3 ) =     1.026
F(  4 ) =     1.064
F(  5 ) =     1.116
F(  6 ) =     1.468
F(  7 ) =     .7742
F(  8 ) =     .6916
F(  9 ) =     .8875
F( 10 ) =     .9562
F( 11 ) =     1.034
F( 12 ) =     1.024
```

STANDARD ERROR OF FORECAST IS 12.79

4>LIST

TITLE- XPDATA

	PRODA	PRODB	XPPAR	FORECS
1#	343.9	101.0	402.80916557647	393.56885972
2#	366.3	102.0	2.58408285063	428.94920618
3#	355.8	103.5	.20000000000	421.37787544
4#	395.4	105.4	.80000000000	439.49638394
5#	398.7	104.0	.20000000000	464.05569129
6#	531.4	104.5	.97083229987	613.90411848
7#	271.3	106.0	1.05140451024	325.84386789
8#	256.7	106.0	1.02634553794	292.89074770
9#	301.1	107.3	1.06378113075	378.14684433
10#	377.8	106.2	1.11624410148	409.86388677
11#	354.8	103.0	1.46756889208	445.98592589
12#	376.1	108.2	.77416396584	444.40596119
13#	349.6	109.0	.69162530261	.00000000
14#	380.1	109.2	.88753132874	.00000000
15#	364.7	109.3	.95617378380	.00000000
16#	398.3	109.7	1.03420844908	.00000000
17#	399.7	107.6	1.02440608160	.00000000
18#	546.6	108.7	.00000000000	.00000000
19#	276.2	109.2	.00000000000	.00000000
20#	253.3	105.6	.00000000000	.00000000
21#	334.5	107.0	.00000000000	.00000000
22#	337.6	107.2	.00000000000	.00000000
23#	374.6	107.1	.00000000000	.00000000
24#	390.2	108.0	.00000000000	.00000000
25#	362.0	102.9	.00000000000	.00000000
26#	395.3	104.5	.00000000000	.00000000
27#	392.9	106.2	.00000000000	.00000000
28#	401.0	107.0	.00000000000	.00000000
29#	429.8	106.5	.00000000000	.00000000
30#	561.4	108.0	.00000000000	.00000000
31#	303.5	108.0	.00000000000	.00000000
32#	270.2	108.2	.00000000000	.00000000
33#	349.2	106.5	.00000000000	.00000000
34#	380.9	106.3	.00000000000	.00000000
35#	419.0	107.8	.00000000000	.00000000
36#	409.9	109.0	.00000000000	.00000000

5>SAVE XPANS

NEW FILE- OK? YES

6>QUIT

-

After the user saves the parameters calculated in the XPOS analysis, he may wish to update the parameters by entering the actual value for an additional period. He may do this by typing

p> XPOS WITH column name ↵

where the column name corresponds to the column containing the previously computed parameters. STATPAK prompts for the number of seasons; the number of the current season, that is, the season for which the user wishes to enter the observed value of the variable; and the observed value for the current season. Next, STATPAK prompts for a column name for the updated parameters, a column name for new forecasts, and the number of forecasts the user wishes calculated. These prompts are printed as follows:

```
# OF SEASONS:
SEASON FOR CURRENT PERIOD:
VALUE FOR CURRENT PERIOD:
COLUMN FOR PARAMETERS:
COLUMN FOR FORECASTS:
# OF FORECASTS:
```

If the user types a Carriage Return rather than a column name in response to the STATPAK prompt for a column for the forecasts, the number of forecasts is not requested and no forecasts are computed.

Finally, STATPAK asks if the user wishes the updated parameters printed. The user responds with YES, or NO, and a Carriage Return. The newly created columns may be explicitly saved with the SAVE command.

Example

```
-STATPAK

1>LOAD XPANS

2>XPOS WITH XPPAR

# OF SEASONS: 12
SEASON FOR CURRENT PERIOD: 1
VALUE FOR CURRENT PERIOD: 380.4
COLUMN FOR PARAMETERS: UPPARS
COLUMN FOR FORECASTS:UPFORE
# OF FORECASTS: 6

PRINT UPDATED PARAMETERS :YES

 SO =      402.7
 R =      2.042
 A =      .2000
 B =      .8000
 C =      .2000


SEASONAL FACTORS

F(  1 ) =      .9499
F(  2 ) =      1.051
F(  3 ) =      1.026
F(  4 ) =      1.064
F(  5 ) =      1.116
F(  6 ) =      1.468
F(  7 ) =      .7742
F(  8 ) =      .6916
F(  9 ) =      .8875
F( 10 ) =      .9562
F( 11 ) =      1.034
F( 12 ) =      1.024
```

```
3>LIST UPFORE 1#:6# 


            UPFORE
1#   425.52637852
2#   417.47975202
3#   434.87 889126
4#   458.60482370
5#   605.94138016
6#   321.22336141


4>SAVE UPDXP 

NEW FILE- OK? YES 

5>QUIT 

-
```

SECTION 10
DATA GENERATION

The data generation module permits the user to employ the analytical power of STATPAK to examine historical data, define a model from his past observations, and forecast future values based on this model. The user may then save the data for direct use in Tymshare's TYMTAB and FINPAK programs. STATPAK permits the user to create columns of variable values which may be forecasts based on any of the following:

- A linear function of one independent variable.
- A linear function of several independent variables.
- A polynomial function of one independent variable.
- The XPOS time series forecast function.
- User-defined functions.
- A user-defined step function.

Variable forecasts may be calculated from saved coefficients generated by the linear regression analysis, the multiple regression analysis, the polynomial regression analysis, or the XPOS analysis. Alternatively, the user may enter coefficients for a linear or polynomial function directly, or he may create a column of values for use with any of the forecasting methods.

The user requests this module in STATPAK by typing the FORECAST command. Additional information typed by the user following the word FORECAST depends on the data generation method he wishes to use. As in all STATPAK analysis, the system prompts for required information omitted by the user; that is, the user may type

p> FORECAST ↵

and STATPAK prompts for each specification.

If a data matrix is currently in STATPAK, the maximum number of forecasts which may be saved in a new column is equal to the number of rows in that matrix. If no data is in STATPAK, the maximum number of forecasted values which may be saved is equal to the number of forecasts created by the first FORECAST command. For example, if the matrix contains 20 rows of data and the user types

p> FORECAST 25 SALES LINEAR ↵

the new column SALES contains only 20 values.

STEP FUNCTION

The user may create a column of variables for a step function by entering the data in a simplified format. He types

p> FORECAST n variable name STEP ↵

where n is the number of values for the function, and the variable name corresponds to the new column name for the values. STATPAK prompts for successive values, and the user may enter each value separately or enter a value and the number of times it is to appear in succession in the new column. The value and the frequency of that number may be separated by a comma (,), an asterisk (*), or a space. For example,

```
-STATPAK

1>FORECAST 25 SALES STEP
```

The user wishes to enter a column of 25 values and save them in a column named SALES.

```
TYPE VALUE FOR
 SALES(1): 250
 SALES(2): 240 4
 SALES(6): 310 7
 SALES(13): 295 3
 SALES(16): 350
 SALES(17): 325 7
 SALES(24): 255 2
```

The second number in each entry indicates the number of times the value appears.

```
2>LIST SALES
```

The user lists the newly created column of values.

```
      SALES
1#      250
2#      240
3#      240
4#      240
5#      240
6#      310
7#      310
8#      310
9#      310
10#     310
11#     310
12#     310
13#     295
14#     295
15#     295
16#     350
17#     325
18#     325
19#     325
20#     325
21#     325
22#     325
23#     325
24#     255
25#     255
```

```
3>SAVE STALL
```

The user saves the data on a new file.

```
TITLE: PRODX
NEW FILE- OK? YES
```

STATPAK prints this prompt because each data file must have a title, and no title was previously entered.

```
4>QUIT

-
```

LINEAR FUNCTION OF A SINGLE VARIABLE

The user may obtain forecasts of a variable based on a linear function of the form y=A+Bx, where x is the independent variable. The coefficients A and B are entered directly by the user or exist in a column of the data matrix. The values in the data matrix may be coefficients saved from a linear regression analysis or explicitly entered by the user. The general form of the FORECAST command for this function is

p>FORECAST n variable name LINEAR ↵

where n is the number of values the user wishes STATPAK to compute, and the variable name specified is the column name for the computed forecasts.

The system prompts for the independent variable, the coefficients, and the starting value and step size for the independent variable. The first prompt is

IND. VARIABLES:

and the user enters the name of the independent variable. Next, STATPAK prompts for the coefficients by printing

CONSTANT & COEFFICIENTS:

and the user enters values for A and B, or names the column containing the values for A and B. Finally, the system prompts for the starting value and interval size for the independent variable, and the user enters the initial value and the step size for successive values of the independent variable.

STATPAK computes the requested variable values and saves them in a column with the name of the variable to be forecasted. These values may be saved and/or listed.

Example

```
-STATPAK

1>LOAD LINREG
```

The user loads his data file which contains the results of a linear regression analysis.

```
2>FORECAST 6 NET LINEAR

IND. VARIABLES: TIME
CONSTANT & COEFFICIENTS: LINC
 START & STEP FOR TIME: 1,1
```

The user names the column containing the values for A and B.

```
3>LIST 1#:6# NET
```

He lists the computed values.

```
              NET
1#    7.4934628926
2#    8.6588280536
3#    9.8241932147
4#   10.9895583757
5#   12.1549235368
6#   13.3202886978


4>SAVE NETFOR
```

The new data file contains all the data in LINREG plus the new column NET.

```
NEW FILE- OK? YES

5>QUIT

-
```

LINEAR FUNCTION OF SEVERAL VARIABLES

The user may obtain forecasts for a variable which is a function of several other variables. The equation for the variable is of the form $y=B_0+B_1x_1+B_2x_2+\cdots+B_kx_k$, where $x_1, x_2, \ldots, x_k$ are the independent variables, and the coefficients $B_0, B_1, B_2, \ldots, B_k$ are values directly entered by the user or values stored in a column of the loaded data matrix. The column values may be calculated and saved in the multiple regression analysis or entered by the user.

The form of the FORECAST command for a linear function of several variables is the same as the command for a linear function of a single variable, that is,

p> FORECAST n variable name LINEAR ↵

STATPAK prompts for the independent variables, constant and coefficients, and starting value and step size for each independent variable. For example,

IND. VARIABLES: A,B,C,D ↵

instructs STATPAK to use the independent variables A, B, C, and D to calculate forecasts. The next STATPAK prompt and user-typed values might be:

CONSTANT & COEFFICIENTS: 23.5,44,15.6,119.02,87.59 ↵

The user may enter the name of the column which contains the constant and coefficients. Each linear function requires a single constant and one coefficient per independent variable. Thus, 23.5 is the constant, 44 is the coefficient of A, 15.6 is the coefficient of B, 119.02 is the coefficient of C, and 87.59 is the coefficient of D.

STATPAK prompts for a starting value and the step size for each independent variable by printing successive prompts:

```
START & STEP FOR A:
START & STEP FOR B:
START & STEP FOR C:
START & STEP FOR D:
```

Example

```
-STATPAK

1>LOAD MULREG

2>COLS

INDEP1,INDEP2,INDEP3,INDEP4,INDEP5,DEPVAR,MULTR,MULTC

3>LIST 1#:6# MULTC


               MULTC
1#   105.22304314421
2#      -.64995771712
3#      -.47173536878
4#      1.94162483826
5#      -.00717678248
6#      -.23857415682


4>FORECAST 12 FORDEP LINEAR

IND. VARIABLES: INDEP1,INDEP2,INDEP3,INDEP4,INDEP5
CONSTANT & COEFFICIENTS: MULTC
 START & STEP FOR INDEP1: 1,1
 START & STEP FOR INDEP2: 3,.5
 START & STEP FOR INDEP3: 2,.1
 START & STEP FOR INDEP4: 1,1.5
 START & STEP FOR INDEP5: 5,2
```

```
5>LIST FORDEP 1#:12#

              FORDEP
1#     105.841081429
2#     104.661505024
3#     103.481928619
4#     102.302352215
5#     101.122775809
6#      99.943199404
7#      98.763622999
8#      97.584046594
9#      96.404470189
10#     95.224893785
11#     94.045317379
12#     92.865740974

6>SAVE FORE

NEW FILE- OK? YES

7>QUIT

-
```

POLYNOMIAL FUNCTION

The user may request forecasts of a variable based on a polynomial function of the form $y=B_0+B_1x+B_2x^2+\cdots+B_kx^k$, where the values of the coefficients are entered directly by the user or are stored in a column of the data matrix. The values of the coefficients may be calculated in the polynomial regression analysis and saved for use in this analysis. The user specifies the degree of the polynomial and the starting value and value increments for the independent variable. The user requests this analysis by typing

p> FORECAST n variable name POLYNOMIAL

where n is the desired number of forecasted values, and the variable name is the column name for the computed forecasts. The system prompts

DEGREE:

and the user enters the degree of the polynomial he wants STATPAK to use. STATPAK then prompts for a constant and the coefficients by printing:

CONSTANT & COEFFICIENTS:

The user either enters values for the coefficients or the name of a column containing the coefficients. STATPAK then prompts

START & STEP:

and the user enters the initial value and the increment for successive values of the independent variable.

Example

```
-STATPAK

1>LOAD POLREG
```

This data file contains the results of a polynomial regression analysis.

```
2>FORECAST 6 FORES POLYNOMIAL

DEGREE: 3
CONSTANT & COEFFICIENTS: PCOEFS
START & STEP: 15,5

3>LIST FORES
```

The user wishes to compute values for a polynomial of degree 3, using coefficients from the column PCOEFS. STATPAK computes the equation for independent variable values 15, 20, 25, 30, 35, and 40.

```
            FORES
1#   2971.9999994
2#   7261.9999986
3#  14451.9999971
4#  25291.9999950
5#  40531.9999919
6#  60921.9999881
7#       .0000000
```

4>SAVE PFORES ↵

NEW FILE- OK? YES ↵

5>QUIT ↵

-

THE XPOS FORECAST FUNCTION

The user may obtain forecasts for a variable based on the XPOS forecasting model. The forecasting parameters must be stored in a column of the data matrix. The parameters may be computed in the XPOS analysis and saved, or the user may enter data for the column. The equation for the XPOS model forecasts is

$$FOR(t) = (SO + tR)\ F(t \bmod L)$$

where FOR(t) is the forecast number t, SO is the most recent smoothed average, R is the current trend factor, L is the number of seasons per year, and F(t mod L) is the seasonal factor corresponding to the season being forecasted. The parameters required for input in this analysis are the same parameters computed by the XPOS analysis described on page 00.

The user requests XPOS model forecasts by typing

p>FORECAST n variable name XPOS ↵

and STATPAK requests the name of the column containing the forecasting parameters by printing:

PARAMETERS:

The user enters the name of the column containing the parameters, and STATPAK prompts for the number of seasons per year and the season for the next period with the prompts:

OF SEASONS:
SEASON FOR NEXT PERIOD:

The season for the next period must be the season corresponding to the season following the last observation used to calculate the parameters. STATPAK then calculates the requested number of forecasted values and stores them in a column with the variable name typed by the user in the FORECAST command.

Example

```
-STATPAK

1>LOAD XPANS

2>FORECAST 18 SALES XPOS

PARAMETERS: XPPAR
# OF SEASONS: 12
SEASON FOR NEXT PERIOD: 1

3>LIST 1#:18# FORECS,SALES


            FORECS          SALES
1#     393.56885972    393.56885972
2#     428.94920618    428.94920618
3#     421.37787544    421.37787544
4#     439.49638394    439.49638394
5#     464.05569129    464.05569129
6#     613.90411848    613.90411848
7#     325.84386789    325.84386789
8#     292.89074770    292.89074770
9#     378.14684433    378.14684433
10#    409.86388677    409.86388677
11#    445.98592589    445.98592589
12#    444.40596119    444.40596119
13#        .00000000   423.67339288
14#        .00000000   461.55220255
15#        .00000000   453.20381828
16#        .00000000   472.48316686
17#        .00000000   498.66929817
18#        .00000000   659.41195375


4>QUIT


-
```

USER-DEFINED FUNCTION

The user may request variable forecasts based on a user-defined function comprised of the built-in functions and operators described on page 42. Any combination of functions and operators constitute a valid expression. The user types

p> FORECAST n variable name=expression

where n is the number of values to be calculated; the variable name is the name of a new data column to contain the computed values; and the expression is a combination of constants, independent variables, operators, and/or functions as discussed on page 41.

STATPAK prompts for the initial value and the value of the increment for successive values of each independent variable. The user enters these values, and STATPAK calculates the new column of data which the user may save on a file with the SAVE command.

Example

```
-STATPAK

1>FORECAST 15 F1=EXP(X)+(Y↑2)

 START & STEP FOR X: 23.1,.25
 START & STEP FOR Y: 15,.5
```

The number of forecasts computed by this command sets the number of rows which may be saved in successive commands.

```
2>FORECAST 12 F2=SIN(X)-(2.4*COS(Y))

 START & STEP FOR X: .27,.04
 START & STEP FOR Y: .5,.01
```

The user may request no more than 15 values saved.

```
3>FORECAST 14 F3=SQR((A+B)/C)

 START & STEP FOR A: 131.2,.8
 START & STEP FOR B: 129.5,1.5
 START & STEP FOR C: 20.4,2.1

4>SAVE FFORE

TITLE: FVLS
NEW FILE- OK? YES

5>LIST


TITLE- FVLS
                       F1             F2             F3
1#      1.0769673596E+10  -1.8394667119   3.5748303127
2#      1.3828534576E+10  -1.7895281819   3.4189017080
3#      1.7756189816E+10  -1.7398682238   3.2839842943
4#      2.2799398972E+10  -1.6905485541   3.1658287872
5#      2.9275007708E+10  -1.6416300328   3.0612951144
6#      3.7589853900E+10  -1.5931725676   2.9680063154
7#      4.8266327751E+10  -1.5452350196   2.8841258326
8#      6.1975191531E+10  -1.4978751115   2.8082093736
9#      7.9577721044E+10  -1.4511493369   2.7391035344
10#     1.0217981635E+11  -1.4051128719   2.6758746875
11#     1.3120148118E+11  -1.3598194888   2.6177580111
12#     1.6846603645E+11  -1.3153214718   2.5641202475
13#     2.1631467254E+11    .0000000000   2.5144320275
14#     2.7775353744E+11    .0000000000   2.4682469865
15#     3.5664260156E+11    .0000000000    .0000000000


6>QUIT


-
```