LIZPACK

Statistical Analysis System

BY

William J. Gray, Ph.d.

Professor of Mathematics
The University of Alabama

Manual written in association with

Elizabeth M. Gray

Edward P. Rutledge
Manager, Alabama State Data Center
Center for Business and Economic Research
The University of Alabama

Technical assistance,

SHOWPACK and DATAPACK programming and manuals by

Michael Dudgeon

James E. Dudgeon, Ph.D.
Professor of Electrical Engineering
The University of Alabama

LIZPACK

Statistical Analysis System

by

William J. Gray, Ph.D.
Professor of Mathematics
The University of Alabama

Manual written in association with

Elizabeth M. Gray

Edward P. Rutledge
Manager, Alabama State Data Center
Center for Business and Economic Research
The University of Alabama

Technical assistance, SHOWPACK and DATAPACK programming and manuals by

Michael Dudgeon

James E. Dudgeon, Ph.D.
Professor of Electrical Engineering
The University of Alabama

LIST OF LIZPACK PROGRAMS

A. Files Administration Subsystem

1. EDITOR (all diskettes); p. 27.
2. FILEASSM.AFM: files assembler (diskette 2); p. 12.
3. FILEMOD.CFM: custom files modification; append records (diskette 2); p. 25.
4. FILEPROD.AFM: files production (diskette 2); p. 16.
5. TRANSPAK.LPK and LIZENCO.LPK: main variable transformation programs (diskette 2); p. 71.
6. MERGFILE.AFM: merging AGGREGATE files (diskette 2); p. 23.
7. TRANSPAK.REG: pre-programmed variable transformations for regression analysis and AGGREGATE files (diskette 1); p. 60.
8. DELETE.VAR: deletion of variables from AGGREGATE files (diskette 1); p. 23.
9. LIZENCO.SRS: combining time series (diskette 7); p. 134.
10. TRANSPAK.SRS: variable transformation for time series (diskette 7); p. 134.

B. Comparative Statistics

1. LIZMEAN.MCP; LIZMEAN.AGG: means comparison package (diskette 1, diskette 3); p. 33.
2. LIZ1WAY.ONE; LIZ1WAY.AGG: one-way ANOVA (diskette 1, diskette 2); p. 35.
3. LIZ2WAY.TWO; LIZ2WAY.AGG: two-way ANOVA, equal cell frequencies (diskette 1, diskette 3); p. 40.
4. LIZANOVA.NEQ; LIZ2NEQ.AGG: two-way ANOVA, unequal cell frequencies (diskette 1, diskette 3); p. 42.
5. LIZANOVA.RAN; LIZRAND.AGG: one-way randomized block ANOVA (diskette 1, diskette 3); p. 38.
6. LIZBLOCK.2WY; LIZBLOCK.AGG: two-way randomized block ANOVA (diskette 1, diskette 3); p. 43.
7. LIZANOVA.3WY; LIZ3WAY.AGG: three-way ANOVA, equal cell frequencies (diskette 3, diskette 3); p. 45.
8. LIZ3WAY.NEQ; LIZ3NEQ.AGG: three-way ANOVA, unequal cell frequencies (diskette 3, diskette 3); p. 46.
9. LIZ1WAY.ANC; LIZ1ANC.AGG: one-way ANCOVA (diskette 1, diskette 3); p. 36.
10. LIZ2WAY.ANC; LIZ2ANCV.AGG: two-way ANCOVA (diskette 1, diskette 3); p. 41.
11. LIZANCOV.RAN; LIZANRAN.AGG: one-way randomized block ANCOVA (diskette 1, diskette 3); p. 38.
12. LIZANCOV.BLK; LIZBLANC.AGG: two-way randomized block ANCOVA (diskette 1, diskette 3); p. 44.
13. LIZ1CORR.MCP; LIZ2CORR.MCP; LIZ3CORR.MCP; and LIZ4CORR.MCP: regression model ANOVA (diskette 6); p. 62.

C. Frequency Distribution

1. LIZSAMP.FRQ: random sampling; one-way frequency distribution (diskette 1); p. 31.
2. LIZCROSS.TAB: crosstabulation, contingency analysis; analysis of crossclassified data (diskette 6); p. 109.

D. Descriptive Statistics

1. LIZDESC.DES: descriptive statistics from a CUSTOM file (diskette 1); p. 33.
2. LIZDESC.AGG: descriptive statistics from an AGGREGATE file (diskette 6); p. 108.

E. Regression Analysis (also see Time Series Analysis and Forecasting)

1. LIZCORR.MUL; LIZCORR.XTM: multiple regression analysis; correlation coefficients with significance (diskette 2); p. 55.
2. LIZANAL.MUL; LIZANAL.NAM: modeling for multiple regression analysis; stepwise regression (diskette 2); p. 47.
3. LIZANAL.RES: multiple regression analysis; residual analysis (diskette 1); p. 56.
4. LIZANAL.PLY: polynomial regression (diskette 2); p. 51.
5. LIZEXPO.ANL: exponential polynomial regression (diskette 2); p. 51.
6. LIZRAD.PWR: fitting a curve of the form $y = ax^b$ (diskette 2); p. 54.
7. LIZLOG.ANL: fitting a curve of the form $y = a + b \log x$ (diskette 2); p. 54.
8. LIZEXPO.ONE: fitting a curve of the form $y = ab^x$ (diskette 2); p. 54.
9. LIZPOLY.ONE; LIZPOLY.STP: utility programs for polynomial regression (diskette 2); p. 55.
10. LIZPOLY.RES: residuals from polynomial regression and LIZEXPO.ANL (diskette 1); p. 59.

F. Factor Analysis

1. LIZCORR.MAT: preparation of the correlation matrix (diskette 4); p. 93.
2. LIZCOMP.PCL: principal component analysis (diskette 4); p. 94.
3. LIZFACTR.ANL: principal factors by iteration (diskette 4); p. 95.
4. LIZQUART.MAX: orthogonal rotation (QUARTIMAX) (diskette 4); p. 97.
5. LIZVARI.MAX: orthogonal rotation (VARIMAX) (diskette 4); p. 97.
6. LIZORTHO.MAX: orthogonal rotation (ORTHOMAX, includes EQUIMAX) (diskette 4); p. 97.
7. LIZOBLI.MIN: oblique rotation (DIRECT OBLIMIN) (diskette 4); p. 98.
8. LIZSCORE.FAC: factor score coefficients; factor scores (diskette 4); p. 98.

G. Discriminant Analysis and Other Multivariate Methods

H. Non-parametric Statistics

I. Time Series Analysis and Forecasting

J. Miscellaneous

1. Practice Data Files (diskette 8).
2. SHOWPACK (diskette 4).
3. DATAPACK (diskette 4).

LIZPACK

LIZPACK IS PUBLISHED EXCLUSIVELY BY:

PRICKLY-PEAR SOFTWARE
8532 EAST 24th STREET
TUCSON, AZ 85710
(602) 886-1505
1-800-223-5369 EXT 256

WARRANTY:
Like all Prickly-Pear Software programs, Lizpack is guaranteed to load for as long as you own it. (No time limit.) If a disk ever fails to load for any reason (even if the dog chewed it), simply return it to us for a prompt, free replacement. There is no other warranty, expressed or implied.

IMPORTANT:
To avoid loss, DO NOT use the disks supplied as your working disks. Back up the master disks and USE THE BACKUPS.

Dedicated to the memory of

Professor Charles L. Seebeck, Jr.

LIZPACK
COLOR COMPUTER™ VERSION
PREFACE

LIZPACK is a menu-driven, user-interactive statistical package for the microprocessor. It is an attempt to make those statistical procedures which are most widely used available for the personal computer at an economical cost.

LIZPACK consists of 88 statistical and file management programs. These programs are stored on seven diskettes, each of which contains two control programs. LIZPACK programs are loaded by a single keystroke from the control program LIZPACK DIRECTORY. The total amount of diskette storage space required for all LIZPACK programming is approximately 850,000 8-bit bytes.

Each LIZPACK program is designed to completely perform one or several functions. A given project, such as a factor analysis, may involve the use of several LIZPACK programs. However, each program comes into use at a logical division point where it is desirable for the user to stop and examine the current state of the project and decide what the next course of action should be.

MINIMUM SYSTEM REQUIREMENTS: LIZPACK requires 32K of memory, Extended Color BASIC™, a printer as described below, and at least one disk drive.

PRINTER: The printer must be on line at all times when LIZPACK is in memory. All printouts are designed for 8 1/2 inch paper width, but of course will work with wider paper.

One of the following printers is required:

(a) a dot matrix printer which produces 8 1/2 inch wide printouts and has normal (10 cpi), compressed (12 cpi), and condensed (16.7 cpi) print sizes, OR has normal (10 cpi) and condensed (16.7 cpi) print sizes.

(b) a dot matrix printer which uses 15 inch paper and has normal (10 cpi) print. Compressed (12 cpi) and/or condensed (16.7 cpi) print may be used if desired.

Appropriate printer codes for each size print are entered by the user at the beginning of each session.

LIZPACK may be used by individuals who have only a superficial knowledge of the use of computers. The user is not required to learn a syntax or code of any kind except in the use of transformation programs; even there, a minimal familiarity with BASIC programming will normally be sufficient.

LIZPACK does not assume that the user has an extensive knowledge of statistics. However, the user must be sufficiently familiar with the statistical processes and terminology so as to be able to provide appropriate responses to LIZPACK queries and prompts.

The LIZPACK operation manual is non-technical. Insofar as possible, we have attempted to use standard terminology throughout LIZPACK. We do not include the mathematical foundations of the statistical functions performed by the package, the formal definition of the statistical terms used, nor the guidance to the interpretation of the data. However, we do provide specific references to standard literary sources which explain the methods and terms used by the various LIZPACK programs. In many cases, less technical sources are also available and may be more appropriate for certain purposes.

LIZPACK provides for two types of data files:

1. CUSTOM files, and
2. AGGREGATE files.

CUSTOM files are load-and-run files, while AGGREGATE files are the standard data base files which are used in most statistical packages. In AGGREGATE files, data is thought of as being arranged by case in the form of a matrix. All LIZPACK data files are formatted random access files. Field formatting is handled exclusively by LIZPACK; the user is not required to participate in the process.

Like all large statistical packages with which we are familiar, LIZPACK is written in a high-level language. The early major statistical packages for mainframes, such as SPSS and BMDP, were written in FORTRAN. The more recent development of powerful BASIC languages by such software companies as MICROSOFT™ have made it possible to construct statistical packages for microprocessors in BASIC. LIZPACK is constructed within the framework of MICROSOFT™ BASIC, COMPILER, and MS-DOS™. We believe the current trend of the development and widespread availability of these systems will continue.

LIZPACK automatically makes the space normally reserved for high resolution grahics available for those programs which do not use such graphics. In addition, almost all LIZPACK programs perform dynamic dimensioning. These features serve to make efficient use of all available memory space and allows LIZPACK to handle quite large problems.

LIZPACK is divided into several packages. A brief description of each package follows.

PACKAGE A:

CORE PACKAGE

The functions of this package can be briefly summarized by categories.

DATA FILES ADMINISTRATION: Many statistical programs are equipped with their own file makers and a copy of the MINI-EDITOR. The main files administration subsystem has a complete EDITOR for editing both CUSTOM (load-and-run) and AGGREGATE (data base) files. Utility programs provide for extensive file manipulation and modification such as sorting, merging, deleting from, and adding to files.

DESCRIPTIVE STATISTICS. One-way frequency distribution, random sampling, theoretical frequency distribution, histogram, frequency polygon. Standard descriptive statistics of samples; Z-scores.

COMPARISON OF MEANS. Comparison by the t-test, Scheffe's test, and Tukey's test; examination of means for polynomial trends (equal or unequal spacing) by means of orthogonal polynomials; partitioned ANOVA tables, and standard ANOVA tables in a unique means comparison package.

One-way ANOVA, ANOVA for randomized block experiments, two-way and three-way ANOVA with equal or non-equal cell frequencies and two-way ANOVA for block design experiments.

One-way analysis of covariance, analysis of covariance for randomized block experiments, two-way ANCOVA with equal cell frequencies, and two-way ANCOVA for block design experiments.

REGRESSION ANALYSIS. Mass production of Pearson's r's. Unique packages for multiple regression allows the following: use of crossterms and dummy variables with an option to do multiple exponential regression; stepwise regression with partitioned ANOVA tables; rapid modeling by the forward and backward elimination methods; computation of partial correlation coefficients.

Residual analysis program includes: residuals and normal deviates; plot of straight line curve fit for any two variables; plots of residuals vs. variables and predicted values of dependent variable; LAG-1 serial plots; straight line predictions.

Special packages for polynomial regression and exponential regression (fitting a curve of the form $y=\exp(P(x))$ where $P(x)$ is a polynomial) feature: graphs and scattergrams in two coordinate systems; multiple graphs for comparison purposes; selection of best model by means of partial F's; curve fits with special polynomials (such as even and odd polynomials); ANOVA tables for regression analysis.

** Printer scattergram option available for compatible graphics printers.

Included is a special program for common variable transformations used in regression analysis (20 preprogrammed transformations).

Curve fits, graphs, and ANOVA tables for curves of the form $y=ab^x$, $y=ax^b$, and $y=a+b \log x$.

Utility programs for polynomials and multiple regression (some with graphics).

TRANSFORMATION OF VARIABLES: LIZPACK performs variable transformations by incorporating user written BASIC statements. The full power of BASIC is at the user's disposal for variable transformations. (See Chapter VI.)

SHOWPACK: A disk graphics package for editing and labeling LIZPACK graphs saved to diskette. High resolution graphics screen text; program to display graphs for lectures, presentations, etc.

DATAPACK: The program DATAPACK converts ASCII files to LIZPACK files or converts LIZPACK files to ASCII. This feature, together with your terminal package, may be used for interfacing with other computers.

SHOWPACK was programmed and documented by Michael Dudgeon. DATAPACK was programmed by Michael Dudgeon and documented by William J. Gray.

PACKAGE B:

FACTOR ANALYSIS

Programs in this package provide for preparation of the raw data file and the correlation matrix, extraction of the principal components (principal component analysis), and estimating the principal factors by iteration. The rotation methods available are QUARTIMAX, VARIMAX, ORTHOMAX (includes EQUAMAX), and the DIRECT OBLIMIN method. Factor score coefficients may be computed either from principal components or principal factors. Factor scores may be output to the printer and/or to a diskette data file. Graphical representation of principal factors from orthogonal rotation is available.

The user has the option to enter a correlation matrix, communalities, initial factor matrix for rotation, and to control the number of principal factors used.

PACKAGE C:

DISCRIMINANT ANALYSIS AND CANONICAL CORRELATION

The Matrix Preparation routine includes a one-way MANOVA.

Performs a complete discriminant analysis and canonical correlation analysis; Geisser classification; includes a program for one-way analysis of covariance with multiple co-variates and multiple variables (MANCOVA).

PACKAGE D:

CROSSTABULATION AND RELATED PROGRAMS

DESCRIPTIVE STATISTICS from an AGGREGATE file; Z-score transformations. Multiple diskette data files may be used.

THE ANALYSIS OF CROSS CLASSIFICATIONS - CROSSTABULATION AND CONTINGENCY ANALYSIS: Like most crosstabulation programs, this program will perform a two-way analysis or will print a multi-dimensional table as a series of two-way tables. For each two-way table the program prints ten nominal and nine ordinal measures of association. This program allows for the possibility of multi-page printouts for each two-way table. Multiple diskette data files may be used.

NON-PARAMETRIC STATISTICS: Tests and procedures include: one sample Wilcoxen signed rank test; Hodges Lehmann estimate of the median; the Friedman two-way layout; Page's test for ordered alternatives; the Kruskall-Wallis one-way ANOVA; Jonckheere's test; the Kendall and Spearman rank order correlation coefficients.

PACKAGE E:

TIME SERIES ANALYSIS AND FORECASTING

The LIZPACK forecasting programs include ANOVA for trend and seasonal effects; autocorrelations, partial autocorrelations, correlogram, and partial correlogram, differencing series, periodogram with cumulative ordinate and spectral ordinates; log periodogram (graph); normalized cumulative periodogram with white noise test and F-ratios for significance of spectral ordinates; moving average program allows user-entered weights or system computed weights for additive or multiplicative models and seasonal adjustment of data; computation of seasonal component from detrended series via harmonics; forecasting seasonal data with polynomials and trigonometric functions; forecasting seasonal data with polynomials and dummy variables; single, double, and triple exponential smoothing; a complete autoregression modeling and simulation package. Utilities: combining time series; common transformations of time series; comparative plots of time series; scattergrams from aggregate files. The forecasting package makes extensive use of graphics.

PUBLISHED BY
PRICKLY-PEAR SOFTWARE
8532 East 24th Street
Tucson, Arizona 85710

TABLE OF CONTENTS

CHAPTER I
GETTING STARTED

The following instructions must be followed for LIZPACK to function properly.

The printer must be on line at all times when LIZPACK is in memory.

With any LIZPACK program diskette in the drive, key in: LOAD "LIZPACK.SET",R <ENTER> or RUN "LIZPACK.SET" <ENTER> (include disk drive number if you are using a drive other than 0 drive). You will be asked to choose a printer baud rate; if your printer will handle a higher baud rate, you will obtain faster printer output at this rate. You will next be asked to specify the disk drive you wish LIZPACK to operate from (if you have more than one drive, it is best to use drive 0 for data diskettes and a higher drive for LIZPACK program diskettes).

You will now be asked to enter the codes which causes your printer to select standard or pica (10 cpi), compressed or elite (12 cpi), and condensed or micron (16.7 cpi) print sizes. Each set of codes must consist of two numbers separated by a comma. If your printer uses only one number to select a print size, then enter 0 (zero) as the second number. (For example, on the Radio Shack DMP 200 printer, the code for selecting standard print size is 27,19.) Consult your printer manual for the sets of codes for the selection of print size. If your printer does not have compressed print (12 cpi), enter the code for condensed print (16.7 cpi).

If your printer handles 15" paper and prints at least 130 characters per line, you may normally enter all zeros for the printer codes. Some analysis of covariance programs may need a 132 character line. In this case, you may need compressed print. Some printers have codes for COMPRESSED ON and COMPRESSED OFF only. If compressed print is to be used for such printers, enter COMPRESSED OFF codes in response to ENTER CODES FOR NORMAL PRINT. Note that the names of print sizes may vary from printer to printer, so be sure to use the number of characters per inch (cpi) to determine which printer code to enter for each size print.

The DIRECTORY (automatic program loading system) will load (the printer MUST be on line!), and the first page of the DIRECTORY will appear on the screen. You may now load any program by pressing the key under which it is listed. An exception to this is the MINI-EDITOR (remove deletion flags), which is in the same subprogram as the DIRECTORY and does not require a program load. Never attempt to load a LIZPACK program directly from a diskette; always load programs using the DIRECTORY.

If you load a program which supports the printer SCATTERGRAM option, you will be asked if you wish to exercise the option **. If so, the machine language screen dump routine will load when you enter Y. After entering Y or

N, the LIZPACK trademark will appear, followed by the sub-system ID and the menu. Before executing any program, read the operating instructions. For now, notice that one of the options reads, "TO LIZPACK" or "LIZPACK DIRECTORY;" press the key beside this option to reload the DIRECTORY.

The program LIZPACK.SET need only be run at the beginning of each session. Unless you interrupt LIZPACK operation and perform some operation which alters the contents of high RAM, you will not have to rerun LIZPACK.SET when using different LIZPACK program diskettes. If the DIRECTORY fails to load, or if the DIRECTORY fails to load a program, or if the printer does not print properly, you will need to rerun LIZPACK.SET.

The LIZPACK DIRECTORY or TO LIZPACK option on the main menus of the various LIZPACK programs will load the LIZPACK DIRECTORY on any diskette, not just that one on the diskette which contains the program in memory. This feature may be used to transfer to the DIRECTORY program on a different diskette.

CRASHES: If any LIZPACK program crashes for some reason, you may re-enter LIZPACK by typing RUN <ENTER>

REPEAT RUNS: AVOIDING DOUBLE DIMENSION ERRORS: The MAIN MENU for a LIZPACK program is the first menu which appears after the program is loaded. If a MAIN MENU option is to be repeated, you must proceed as follows to avoid a Double Dimension (DD) error:

While the MAIN MENU is listed on the screen, firmly press key Z and release it. This clears all variable tables. In virtually all cases, LIZPACK performs dimensioning dynamically; this allows the handling of a much wider variety of problems. However, it usually results in a Double Dimension error unless variable tables are cleared before a repeat run of the same menu option is made.

Pressing key Z while the MAIN MENU is on the screen will destroy the results of any and all previous menu option runs. It should not be pressed while the computer contains information which is vital for the run of another menu option.

If a program has an OPTIONS menu, none of the preceding paragraph applies. Options may be repeated as many times as desired from an OPTIONS menu.

It is recommended that you examine the DIRECTORY on all diskettes to familiarize yourself with the contents of the system. Try a few program loads, but wait until you have read Chapters II, III, and V before loading the EDITOR, a files assembly program, or exercising the MINI-EDITOR option.

MAKE AT LEAST ONE BACK-UP COPY OF EACH DISKETTE. Follow the above loading procedures; check to make sure that every program loads. Do not stop until you have a perfectly loading backup copy of each diskette.

ANSWERING QUERIES: When a LIZPACK query requires a yes or no answer, type YES<ENTER> or NO<ENTER> unless the query indicates that you are to answer with Y or N.

ERROR MESSAGES: The Color Computer version of LIZPACK is monitored by DISK EXTENDED COLOR BASIC. When your data or data file is such that LIZPACK cannot perform an operation, program execution is terminated and a standard error message is printed. The meaning of these messages may be found in the programming or disk manual provided with your computer.

SHOWPACK: LIZPACK makes extensive use of high resolution graphics as a qualitative aid to the investigator. LIZPACK does not label the graphs; however, LIZPACK graphs may be saved to diskette and used as input for the general graphics editing package SHOWPACK which can be used to label and edit the graphs with cursor control and choice of high resolution screen text. SHOWPACK can also be used for quickly displaying graphs for lectures, presentations, etc. For complete details, instructions, and other uses for SHOWPACK, see the SHOWPACK manual.

To save a graphics file for SHOWPACK input, select the SHOWPACK option from any menu on which it appears. Provide LIZPACK with the name of the graphics file. The graph will be saved to diskette. You will receive printed output which includes:

(1) minimum and maximum values of the horizontal (X) and vertical (Y) coordinates;

(2) units of measurement on each axis; and

(3) the screen coordinates of the intersection of the coordinate axis.

LIZPACK graphics files are in binary and require 3 granules each. You may wish to keep all your graphics files on a separate diskette.

We note that SHOWPACK may be used for creating graphs. In particular, it creates excellent bar graphs and pie charts. Again we refer the user to the SHOWPACK manual for a full description of its features.

DISK 1 DISK 2

A SPECIAL NOTE ON PROGRAMS LIZANAL/PLY AND LIZEXPO/ANL: The SHOWPACK option does not appear on the OPTIONS menu of these programs; this is due to lack of screen space. Nevertheless, the SHOWPACK option is available for these programs. Any time the OPTIONS menu is on the screen, press Key S and exercise the SHOWPACK option as described above.

CLOSING DOWN LIZPACK: Each LIZPACK program contains a menu option which reads END LIZPACK or EXIT LIZPACK. These are for the temporary interruption of LIZPACK operation. If you are closing down LIZPACK to turn off the computer or to use it for other purposes, it is best to exit LIZPACK from the LIZPACK DIRECTORY.

After closing down LIZPACK, a PCLEAR 0 may be in effect. If you are going to use a program which uses high resolution graphics, type PCLEAR 4 <ENTER> before loading the program.

A NOTE ON USING THE EDITOR AND MINI-EDITOR: When using either the EDITOR or the MINI-EDITOR, be sure that the data diskette which is being edited is on the default disk drive. The default disk drive is DRIVE 0 unless you have specified otherwise.

CHAPTER II
LIZPACK DATA FILES

All large statistical systems have a files or data management system. The type of files system varies from one statistical package to another. All general types of files management systems have both advantages and disadvantages. In LIZPACK every effort has been made to provide the user with the greatest possible versatility and ease of use in managing data files. However, mastering the use of the FILES ADMINISTRATION SUBSYSTEM will probably take more time than any other phase of learning to use LIZPACK.

The terminology used in data management systems varies. It is a good idea to first reach an understanding as to what the various terms mean in LIZPACK.

LIZPACK uses two types of data files. The data files are defined as follows:

AGGREGATE FILE: Any file in which data can be thought of as being arranged in rows and columns. Each row consists of measurements made on a single case or subject, and at the user's option, may contain a case number and/or an ID number or numbers. Each column contains the measurement of a specific variable or case number or ID number. For each case (i.e., row), the case number, if any, must appear in the same column, the ID's must appear in the same column, and the measurement of a given variable must appear in the same column. The AGGREGATE file allows the user to establish a database for a project.

Most of the statistical functions performed by LIZPACK may be accomplished by using an AGGREGATE file as the input file for statistical programs. However, most LIZPACK statistical programs do not accept files which occupy more than one diskette. Hence, if an AGGREGATE file requires more than one diskette, a program input file must be prepared in one of the following ways:

(1) The file may be shortened by removing extraneous variables or other information (see Chapter III).

(2) A CUSTOM file may be prepared from the AGGREGATE file.

CUSTOM FILE: Any LIZPACK data file which is in the format required by a LIZPACK statistical program and contains only the data needed by the program. Many CUSTOM files (especially CUSTOM regression analysis files) may be regarded as an AGGREGATE file for most purposes.

Custom files serve two purposes:

(1) They provide the user with a load-and-run option.

(2) CUSTOM files are compact since they contain only the information actually required by a LIZPACK statistical program.

The main disadvantages of CUSTOM files, especially those made for the ANOVA and ANCOVA programs, is that they are not compatible with other LIZPACK programs.

Some LIZPACK statistical programs which are designed specifically as load-and-run programs use CUSTOM files only. Other programs such as the regression analysis accept both CUSTOM and AGGREGATE files. For a summary of the type files used by the various LIZPACK programs, see the APPENDIX to this chapter.

Note that the term COLUMN in LIZPACK refers to the position of the observations of a variable (or case number, ID number) and has nothing to do with the term "column" as it is used with reference to cards. In some systems, what we call a COLUMN is called a FIELD.

In a LIZPACK data file, each data word consisting of an observation or case number or ID number occupies a RECORD. Each record requires 5 bytes of diskette storage space. Thus, the two observations

11

and -.179633345E-03

will occupy the same amount of disk storage space (5 bytes) and will be stored and read in from the diskette with exact accuracy. A diskette will hold somewhat more than 31,000 records.

When a listing is made of a data file on the screen or printer by means of the EDITOR or MINI-EDITOR, the contents of each record is accompanied by its record number. This convention is for your convenience. Record numbers never appear on the diskette.

LIZPACK data files created by the user must contain numeric data only.

All LIZPACK statistical programs use a disk data file. Any file which is in the specific format required by a particular LIZPACK program and contains only the information required by a specific program is called a CUSTOM file. (See Chapter VI for more information on formats.) Before running a statistical program, the user may need to prepare a CUSTOM file for it. This may be done in any of three ways:

(1) A CUSTOM file may be prepared from an AGGREGATE file. This process is discussed in Chapter III.

(2) A CUSTOM file of any type may be prepared using the EDITOR in the input mode. Long files should always be prepared in this manner because such a file does not have to be made up all at one time. The EDITOR cannot give

you assistance in preparing a file in a specific format; it merely places the information you give it in the next available file record. See Chapter V for instructions on using the EDITOR.

(3) By using the files preparation routine included with most statistical programs. These routines are specifically designed to create a file in the format required by the accompanying statistical program. These routines usually assist you in making up the file. WITH THESE ROUTINES, THE ENTIRE FILE MUST BE MADE WITHOUT INTERRUPTION. To use such a routine, first read the operating instructions for the statistical program you want to use so that you understand the required format of the data file. Then select the menu option which reads "CREATE DATA FILES". You may now give the data file any valid file name of your choice and begin entering the data.

Most programs contain a copy of the MINI-EDITOR which can be used to make minor revisions in the data file. To use the MINI- EDITOR, select the indicated menu option. Give the MINI-EDITOR your data file name; you will then be asked to enter a READ MULTIPLE, i.e., the number of records you want listed on the screen at one time (no more than 14). The MINI-EDITOR will list the specified number of records you want listed on the screen, beginning with record 1. The record number and record are listed.

To continue listing, press key C.

To edit, press key E after noting the number of the record which needs editing; the editing options will appear. You may change, delete, or insert records one at a time. If you delete records, see the note below. You must, of course, tell the EDITOR which records you want to EDIT, so make a note of these record numbers before entering the EDIT option from the LIST mode.

You may list the entire file, or you may terminate the MINI-EDITOR as follows:

(a) If you are in the LIST mode, press E and select the EXIT option;
(b) If you are in the EDIT mode, simply select the EXIT option.

You can create and edit most CUSTOM data files and make a program run with the program you have in memory unless:

(a) you must make extensive revisions. Using the MINI-EDITOR will be tedious. Use the EDITOR (See Chapter V).
(b) you must make deletions from the file. (See note below.)

NOTE: Making deletions from a data file: Deletions from a data file require special handling. If you use the DELETE RECORDS option in the MINI-EDITOR, the data in the specified record will be deleted, all subsequent data in the file is moved forward one record, and a DELETION FLAG (the number 979797) is placed in the last position of the file. These deletion flags must be removed from the file before running a program. Thus, if you must make deletions in a file,

(a) complete all editing in the file.
(b) when the file is in the desired form, (except for the offending deletion flags at the end), exit the EDITOR; the menu will reappear. Reload the directory. Now select the option which reads "MINI-EDITOR--REMOVE DELETION FLAGS." Enter your file name; the EDITOR will remove the deletion flags, and you will then have a completely edited file.

The above procedure is somewhat awkward, but it appears to be the safest way to shorten your file. What actually happens is this: The EDITOR creates a temporary file, called SCRATCH.PAD, and copies all valid records (up to the first deletion flag) from the edited file into the temporary file. The edited file is then KILLED and SCRATCH.PAD is RENAMEd to the name of your data file. For this reason, you must never make deletions from a disk which does not contain enough room for a temporary copy of your file. If you are making a very long custom file you should:

(a) make the file using the main EDITOR;

(b) make up the file in small parts (not more than 4000 records at a time). When each part is completely edited, you may add it to the longer file by using the APPEND RECORDS option in the CUSTOM FILE ASSEMBLY program. (See Chapters IV and V.)

APPENDIX TO CHAPTER II

The following programs are designed as load-and-run programs and accept CUSTOM files only:

LIZDESC.DES, LIZMEAN.MCP,
LIZ1WAY.ONE, LIZ1WAY,ANC, LIZ2WAY,TWO,
LIZANOVA.NEQ, LIZ2WAY.ANC, LIZANOVA.3WY,
LIZ3WAY.NEQ, LIZANOVA.RAN, LIZBLOCK.2WAY, LIZANCOV.BLK
and LIZANCOV.RAN.

The above programs have AGGREGATE file counterparts. These programs accept AGGREGATE files with a minimum of two to five columns.

LIZDESC.AGG, LIZMEAN.AGG,
LIZ1WAY.AGG, LIZ1ANCV.AGG, LIZBLOCK.AGG,
LIZBLANC.AGG, LIZRAND.AGG, LIZANRAN.AGG,
LIZ2WAY.AGG, LIZ2NEQ.AGG, LIZ3ANCV.AGG,
LIZ3WAY.AGG, and LIZ3NEQ.AGG.

The one-way frequency distribution program LIZSAMP.FRQ accepts either CUSTOM or AGGREGATE files for input. The Random Sampling Option is not available from AGGREGATE files. It is expected that this option will be used mainly for instructional purposes.

All the regression analysis programs and the residuals program LIZPOLY.RES of the CORE PACKAGE will accept a CUSTOM file or an AGGREGATE file. The Factor Analysis programs of Package B accept CUSTOM or AGGREGATE files. The Discriminant Analysis programs of Package C accept AGGREGATE files with a minimum of two columns. The crossclassifications analysis program LIZCROSS.TAB accepts AGGREGATE files with a minimum of three columns. The non-parametric programs of Package E accept AGGREGATE files with a minimum of two to three columns.

Because of the nature of the computations performed by the programs of the Forecasting Package (Package E) and because of the inherent nature of time series, all Forecasting programs accept two column AGGREGATE files only. The first column contains the independent variable (usually time), and the second column contains the dependent variable, i.e., the observations of the time series. The files are compatible with all the regression analysis programs of the CORE PACKAGE and all of the functions of the FILES ADMINISTRATION subsystem.

Additional details concerning LIZPACK data files are given throughout the manual.

CHAPTER III
AGGREGATE FILES

In many experiments, multiple measurements are made on a single subject or case, and a number of cases may be involved in the experiment. Analysis of the data using several different statistical processes may be desirable. In this situation, the user will probably want to record the results of the experiment in an AGGREGATE FILE. Data in an AGGREGATE file is thought of as being arranged in a matrix consisting of rows, called CASES, and COLUMNS. The user may specify the number of columns per case in an AGGREGATE file. Each column contains one numeric data word or RECORD, which may be the case number, an ID, or an observation of a variable. Each case must consist of exactly the same number of columns, and an ID number, case number, or measurement of a given variable must appear in the same column of each case. AGGREGATE files are created using the EDITOR, KEY I (INPUT RECORDS).

MISSING DATA: If an observation of one or more variables is missing for a particular case, the column containing the observations for that variable MUST be filled with the MISSING DATA symbol which is

999999

This symbol should never be used for any other purpose in an AGGREGATE file. If you actually have 999999 as a measurement you may replace it with an approximating number such as 999999.0001.

VARIABLES: The measurements of a variable occupy a single column in an AGGREGATE file. The column number for a given variable must be the same for every case. The user must keep track of which column a variable occupies in an AGGREGATE file. This may be accomplished by listing the file with the EDITOR and giving each column a label.

CASE NUMBERS: In an AGGREGATE file, the user may wish to assign each case a number. This is desirable for the purpose of identifying cases in the file and on printouts. If the user elects to use case numbers, the case number should appear in the same column for each case. This would usually be column 1. If the AGGREGATE file editing features such as LIST CASE or DELETE CASE are to be used in the EDITOR, then each case must have a unique number which appears in column 1.

ID NUMBERS: In most AGGREGATE files the user will want to use one or more columns for ID numbers. These may be used for many purposes, one of which is to enable the user to extract data from the file with the files assemblers described below.

As an example, if the data in an AGGREGATE file is to be analyzed with an ANOVA program, an ID number could be used to indicate to which cell the data for the case is to be assigned.

The user may assign as many ID numbers as desired to each case. Each ID number occupies a column and must appear in the same column for each case.

Table 3.1 is an illustration of what an AGGREGATE file might look like. This file contains 11 columns and 10 cases, and would occupy 110 records on a diskette. Each case consists of a case number, 2 ID numbers, and measurements of 8 variables.

TABLE 3.1

COL1 CASE#	COL2 ID1	COL3 ID2	COL4 VAR1	COL5 VAR2	COL6 VAR3	COL7 VAR4	COL8 VAR5	COL9 VAR6	COL10 VAR7	COL11 VAR8
1	1	1	.01	1.02	10.13	1.63	2.17	1.11	999999	.003
2	1	2	.05	1.07	9.32	.75	.32	1.71	21.1	5.02
3	1	1	.07	1.09	11.1	1.83	.4	1.001	22.2	.1
4	2	1	.09	1.12	8.7	1.91	.5	.99	13	.18
5	2	2	.04	1.07	13.5	1.27	.6	1.3	15	7.41
6	2	3	.05	1.25	5.1	1.31	.4	1.5	18	0
7	1	1	.07	1.32	999999	1.05	.3	1.2	21	.21
8	3	1	.06	1.75	6.2	1.07	.2	1.1	23	.22
9	3	2	.04	1.49	3.4	1.06	.1	1.4	27	-.18
10	2	1	.03	1.87	5.1	1.55	.7	1.8	29	1.7

An AGGREGATE file is actually stored on a diskette in the same manner as any other formatted BASIC file. The real point to an AGGREGATE file is that data is entered into the file in a certain fashion, and then the proper information is given to the LIZPACK files administration or statistical programs so that they will be able to find the data and process it properly. The following is a description of how data is entered into an AGGREGATE file.

Suppose the AGGREGATE file is to consist of N cases and each case is to contain M columns. The AGGREGATE file will then contain NM records. Records 1 through M will contain the data for the first case, records M+1 through 2M will contain the data for the second case, records 2M+1 through 3M will contain the data for the third case, etc.

As an example, for an AGGREGATE file containing the data in Table 3.1, records 1 through 11 would contain the data for case 1, records 12 through 22 would contain the data for case 2, etc. Now column 4 of the AGGREGATE file would contain the measurements for a certain variable that we refer to here as variable 1. Thus, the measurements for variable 1 would appear in records 4, 15, 26, 37, 48, 59, 70, 81, 92, and 103. The case numbers would appear in records 1, 12, 23, 34, 45, 56, 67, 78, 89, and 100.

After having read the chapter on the EDITOR (Chapter V), come back to this point and we will see how to enter the data of Table 3.1 into an AGGREGATE file.

Select the option which reads EDITOR from the LIZPACK DIRECTORY. The EDITOR will load and the query

EDITOR - ENTER FILE NAME

will appear. Type

SAMPLE.DAT <ENTER>

The EDITOR will now ask you which disk drive contains the diskette you wish the file to be written on. Type the disk drive number, usually 0, and press <ENTER>. The CONTROL KEYS list will now appear. Press Key B <ENTER> and the EDITOR will pass to the READY state. (B = Begin.) Press Key I for INPUT RECORDS. We will now enter the data from Table 3.1 into the file. On the screen you now have

1 ?

The EDITOR is asking for the data for record 1. Type

1 <ENTER>

This puts the case number for case 1 in the file. The EDITOR responds with

2 ?

Type

1 <ENTER>

to place ID1 for case 1 in record 2. The following is the list of queries and responses for the first 14 records of the AGGREGATE file. DO NOT PLACE ANY SPACES BEFORE YOUR RESPONSE. Press ENTER after each response.

QUERY	RESPONSE	NOTE
1 ?	1	Case number
2 ?	1	ID1
3 ?	1	ID2
4 ?	.01	VAR1
5 ?	1.02	VAR2
6 ?	10.13	VAR3
7 ?	1.63	VAR4
8 ?	2.17	VAR5
9 ?	1.11	VAR6
10 ?	999999	VAR7 (missing data)
11 ?	.003	VAR8
12 ?	2	Case Number
13 ?	1	ID1
14 ?	2	ID2

Continue until all 110 records have been entered into the file. Then,

111 ? 989898

and the EDITOR will return to the READY state. The number 989898 is a signal and will not actually be entered into the file. NEVER USE 989898 AS DATA, SINCE 989898 WILL TERMINATE INPUT!

Many LIZPACK statistical programs require or accept a CUSTOM file. These CUSTOM files may be created from AGGREGATE files using the programs FILEASSM.AFM and/or FILEPROD.AFM which are located on Diskette 2 of the CORE PACKAGE. We will now describe how these programs function.

FILEASSM.AFM
Diskette 2

This program may be used to produce CUSTOM files, to produce sorted or ordered AGGREGATE files, or to produce AGGREGATE sub-files, i.e., to produce a file consisting of a portion of the original file.

This program will search any number of columns for numbers which we call TARGETS. When the numbers in the columns being searched all match the preassigned targets, data from specified columns of that case will be written to a new file.

EXAMPLE 1

To illustrate how this program works, let's use the example file SAMPLE.DAT of Table 3.1. We assume that we have such a file created on a data diskette. Suppose we want to create a CUSTOM file for the multiple regression program LIZANAL.NAM. We want to use variables 1, 3, 5, and 8 from the AGGREGATE file columns 4, 6, 8, and 11, respectively. Only the data with the following ID pairs are to be used:

ID1	ID2
1	1
2	1
3	1

After loading the program, choose the option ASSEMBLE FILES from the main menu. Program queries (Q) and user responses (R) follow. We are giving the CUSTOM file we are creating the name SAMPLE.CUS. Be sure to press ENTER after each response. Do not type the R.

```
Q  ENTER NAME OF
Q    SOURCE (INPUT) FILE
R      SAMPLE.DAT
Q    DESTINATION FILE
R      SAMPLE.CUS
Q  NUMBER OF COLUMNS PER CASE IN SAMPLE.DAT
R  11
Q  NUMBER OF COLUMN SEARCHES TO BE MADE ON SAMPLE.DAT
R  2
Q  NUMBER OF TARGET RUNS MADE ON SAMPLE.DAT
R  3
Q  SEARCH COLUMN NUMBER 1 IS SAMPLE.DAT COLUMN
R  2
Q  SEARCH COLUMN NUMBER 2 IS SAMPLE.DAT COLUMN
R  3
```

NOTE: We have now informed the program that we are to make 3 target runs while searching columns 2 and 3 of the source file.

```
Q  FORMAT FOR DESTINATION FILE
Q  NUMBER OF VARIABLES (COLUMNS) TO COPY FROM SAMPLE.DAT
R  4
Q  DELETE CASES WITH MISSING DATA (Y/N)
R  Y
```

NOTE: You will usually answer <Y> here if you are making a CUSTOM file. Custom files usually do not have cases with missing data.

```
Q  VARIABLE NUMBER 1 FOR OUTPUT FILE IS FROM SAMPLE.DAT COLUMN NUMBER?
R  4
Q  VARIABLE NUMBER 2 FOR OUTPUT FILE IS FROM SAMPLE.DAT COLUMN NUMBER?
R  6
Q  VARIABLE NUMBER 3 FOR OUTPUT FILE IS FROM SAMPLE.DAT COLUMN NUMBER?
R  8
Q  VARIABLE NUMBER 4 FOR OUTPUT FILE IS FROM SAMPLE.DAT COLUMN NUMBER?
R  11
```

NOTE: The program has now been told where to find the data for the four variables. We will now tell it what to look for in the columns for TARGET RUN 1.

```
Q  TARGET RUN 1
   COLUMN 2 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R  1
Q  COLUMN 3 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R  1
```

NOTE: The program will now write all data from cases in the AGGREGATE file having 1 in column 2 and 1 in column 3 to the CUSTOM file. ← DISK HERE

```
Q  TARGET RUN 2
   COLUMN 2 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R  2
Q  COLUMN 3 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R  1
   (The second target run is made.)
Q  TARGET RUN 3
   COLUMN 2 IN SAMPLE.DAT SOULD MATCH TARGET = TO ?
R  3
Q  COLUMN 3 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R  1
```

The third run is made. We have now created a CUSTOM file named SAMPLE.CUS which contains the following data:

RECORD #	DATA	
1	.01	
2	10.13	case #1
3	2.17	
4	.003	
5	.07	
6	11.1	case #3
7	.4	
8	.1	
9	.09	
10	8.7	case #4
11	.5	
12	.18	

13	.03	
14	5.1	case #10
15	.7	
16	1.7	
17	.06	
18	6.2	case #8
19	.2	
20	.22	

Note that the data corresponding to Case 7 does not appear in the file because of the missing data symbol in column 6.

EXAMPLE 2

We use the sample file SAMPLE.DAT of Table 3.1 to create an AGGREGATE file SAMPLE.ORD which contains all the data of SAMPLE.DAT ordered so that all data with

ID1 = 1 ID2 = 1

appears first,

ID1 = 1 ID2 = 2

appears second,

ID1 = 2 ID2 = 1

appears third, etc.

```
Q  ENTER NAME OF
      SOURCE (INPUT) FILE
R  SAMPLE.DAT
Q     DESTINATION (OUTPUT) FILE
R  SAMPLE.ORD
Q  NUMBER OF COLUMNS IN SAMPLE.DAT
R  11
Q  NUMBER OF COLUMN SEARCHES TO BE MADE ON SAMPLE.DAT
R  2
Q  NUMBER OF TARGET RUNS MADE ON SAMPLE.DAT
R  7
Q  SEARCH COLUMN 1 IS SAMPLE.DAT COLUMN
R  2
Q  SEARCH COLUMN 2 IS SAMPLE.DAT COLUMN
R  3
Q  FORMAT FOR DESTINATION FILE
   NUMBER OF VARIABLES (COLUMNS) TO COPY FROM SAMPLE.DAT
R  11
Q  DELETE CASES WITH MISSING DATA (Y/N)
R  N
Q  VARIABLE NUMBER 1 FOR OUTPUT FILE IS FROM SAMPLE.DAT COLUMN NUMBER
R  1
Q  VARIABLE NUMBER 2 FOR OUTPUT FILE IS FROM SAMPLE.DAT COLUMN NUMBER
R 2
```

(This sequence of queries and responses continues for each variable from 1 to 11. The last query and response is for variable 11.)

Q VARIABLE NUMBER 11 FOR OUTPUT FILE IS FROM SAMPLE.DAT COLUMN NUMBER
R 11
Q TARGET RUN 1
COLUMN 2 FROM SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R 1
Q COLUMN 3 FROM SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R 1
Q TARGET RUN 2
COLUMN 2 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R 1
Q COLUMN 3 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R 2
Q TARGET RUN 3
COLUMN 2 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R 2
Q COLUMN 3 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R 1
Q TARGET RUN 4
COLUMN 2 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R 2
Q COLUMN 3 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R 2
Q TARGET RUN 5
COLUMN 2 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R 2
Q COLUMN 3 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R 3
Q TARGET RUN 6
COLUMN 2 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R 3
Q COLUMN 3 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R 1
Q TARGET RUN 7
COLUMN 2 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R 3
Q COLUMN 3 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R 2

This run will create an AGGREGATE file SAMPLE.ORD as shown in Table 3.2.

TABLE 3.2

COL 1 CASE #	COL 2 ID1	COL 3 ID2	COL 4 VAR 1	COL 5 VAR 2	COL 6 VAR 3	COL 7 VAR 4	COL 8 VAR 5	COL 9 VAR 6	COL 10 VAR 7	COL 11 VAR 8
1	1	1	.01	1.02	10.13	1.63	2.17	1.11	999999	.003
3	1	1	.07	1.09	11.1	1.83	.4	1.001	22.2	.1
7	1	1	.07	1.32	999999	1.05	.3	1.2	21	.21
2	1	2	.05	1.07	9.32	.75	.32	1.71	21.1	5.02
4	2	1	.09	1.12	8.7	1.91	.5	.99	13	.18
10	2	1	.03	1.87	5.1	1.55	.7	1.8	29	1.7
5	2	2	.04	1.07	13.5	1.27	.6	1.3	15	7.41
6	2	3	.05	1.25	5.1	1.31	.4	1.5	18	0
8	3	1	.06	1.75	6.2	1.07	.2	1.1	23	.22
9	3	2	.04	1.49	3.4	1.06	.1	1.4	27	-.18

Note that there is a second option on the main menu entitled "ASSEMBLE FILES--SAMPLE COUNT". This option is similar to the one above except that after each target run the total number of cases from which data was taken is placed BEFORE the data in the file. This feature is useful for making files for some of the load-and-run programs.

In using FILEASSM.AFM, for each variable the user must specify a column number in the AGGREGATE file which contains the observations of the variable. In Example 1, the four variables come from columns 4, 6, 8, and 11. If a different order of the variables in the CUSTOM file is desired, the variables could have been taken from columns 6, 4, 11, and 8 for example. Then the first four records of the CUSTOM file would have been

Record	Data
1	10.13
2	.01
3	.003
4	2.17

and the file would continue in this fashion.

MULTIPLE DISKETTE FILES: An AGGREGATE file may occupy more than one diskette. It is important to note that when using FILEASSM.AFM that the program begins inserting records IN THE FIRST AVAILABLE RECORD of the DESTINATION file. Thus, if an AGGREGATE file occupies more than one diskette, after the first diskette is processed, make a new run with FILEASSM.AFM using the second diskette and THE SAME DESTINATION FILE. Continue this process until all diskettes of the AGGREGATE file are processed.

This same procedure applies to program FILEPROD.AFM which is described below.

In order to use multiple diskette files you must have two disk drives; the AGGREGATE file must be on one disk drive, and the destination (or CUSTOM) file must be on the other. IF ONLY ONE DISK DRIVE IS AVAILABLE, THE DESTINATION FILE WILL BE ASSEMBLED ON THE DISKETTE CONTAINING THE AGGREGATE FILE; THUS, THE AGGREGATE FILE CANNOT FILL THAT DISKETTE.

Any file which is intended as an input file for a LIZPACK statistical program must be limited to one diskette. (There are a few exceptions to this rule. See the APPENDIX to Chapter II.)

FILEPROD.AFM
Diskette 2

Using the previous program, in Example 1 we had three target runs involving three separate target sets, and in Example 2 we had seven target runs involving seven separate target sets. FILEASSM.AFM made an individual run for each target set. When using FILEPROD.AFM, all the target sets are presented to the program initially and the program makes a single run; whenever the data in the specified search columns matches any one of the target sets, the data for that case is written to a new file.

EXAMPLE 3

We illustrate the principal difference between FILEASSM.AFM by repeating the run of Example 1 using FILEPROD.AFM. All conditions here are the same as in Example 1, but the order in which the data appears in the DESTINATION file is different.

```
Q  ENTER NAME OF:
      AGGREGATE (INPUT) FILE
R  SAMPLE.DAT
Q     DESTINATION (OUTPUT) FILE
R  SAMPLE.CUS
Q  NUMBER OF COLUMNS PER CASE IN SAMPLE.DAT
R  11
Q  NUMBER OF COLUMN SEARCHES TO BE MADE ON SAMPLE.DAT
R  2
Q  NUMBER OF TARGET SETS TO APPLY TO SAMPLE.DAT
R  3
Q  SEARCH COLUMN 1 IS SAMPLE.DAT COLUMN NUMBER
R  2
Q  SEARCH COLUMN 2 IS SAMPLE.DAT COLUMN NUMBER
R  3
Q  FORMAT FOR DESTINATION FILE
   NUMBER OF VARIABLES (COLUMNS) TO COPY FROM SAMPLE.DAT
R  4
Q  DELETE CASES WITH MISSING DATA (Y/N)?
R  Y
Q  VARIABLE NUMBER 1 FOR OUTPUT FILE IS FROM SAMPLE.DAT COLUMN NUMBER
R  4
Q  VARIABLE NUMBER 2 FOR OUTPUT FILE IS FROM SAMPLE.DAT COLUMN NUMBER
R  6
Q  VARIABLE NUMBER 3 FOR OUTPUT FILE IS FROM SAMPLE.DAT COLUMN NUMBER
R  8
Q  VARIABLE NUMBER 4 FOR OUTPUT FILE IS FORM SAMPLE.DAT COLUMN NUMBER
R  11
Q  TARGET SET 1
   COLUMN 2 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R  1
Q  COLUMN 3 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R  1
Q  TARGET SET 2
   COLUMN 2 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R  2
Q  COLUMN 3 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R  3
Q  TARGET SET 3
   COLUMN 2 IN SAMPLE.DAT SHOULD MATCH TARGET = TO ?
R  3
Q  COLUMN 3 IN SAMPLE.DAT SHOULD MATCH TARGET = TO
R  1
```

The program will now assemble the following DESTINATION file SAMPLE.CUS:

RECORD	DATA	RECORD	DATA
1	.01	11	.5
2	10.13	12	.12
3	2.17	13	.05
4	.003	14	6.2
5	.07	15	.2
6	11.1	16	.22
7	.4	17	.03
8	.1	18	5.1
9	.09	19	.7
10	.5	20	1.7

Note that the data taken from the AGGREGATE file cases 8 and 10 appear in a different order in this sample. FILEPROD.AFM writes the data from a case to the DESTINATION file as soon as it makes a match with ANY target set, whereas the execution of FILEASSM.AFM creates a file ordered according to the sequence of target runs.

FILEPROD.AFM is more efficient than FILEASSM.AFM for making such files as regression analysis files or other files where the order of the cases is immaterial.

When the data in an AGGREGATE file is to be used to make CUSTOM files which are to be analyzed by an ANOVA program or another statistical program where the order of the cases is important, each case may be assigned an identifier (or identifiers) to indicate to which group or cell it belongs. Then the AGGREGATE file may be

(1) prepared with cases appearing in the natural order required by the statistical program, or

(2) prepared in any order and then arranged in the proper order using FILEASSM.AFM.

In preparing DESTINATION files from AGGREGATE files which have been prepared in either one of the above fashions, column searches are no longer necessary. FILEPROD.AFM will permit the entry of the number 0 in answer to the query

Q NUMBER OF COLUMN SEARCHES?

In this case the program will simply pull data from the specified columns for each valid case. If no column searches are made, whenever the program finishes assembling the first DESTINATION file, the following query will appear:

Q NAME OF NEXT CUSTOM FILE?

If you have no further files to make, you may enter NONE. If you need other DESTINATION files in the same format as the first, but with data taken from different columns of the AGGREGATE file, you may enter a new DESTINATION file name. You then need only answer the queries that tell the program which columns the data is to be taken from, and a second file will be assembled in the same format as the first. This process continues to allow the assembly of any number of DESTINATION files of the same format. This is called MASS PRODUCTION of CUSTOM data files and can be very useful in many situations.

EXAMPLE 4

In Table 3.3 we have a hypothetical file called ANOVA.DAT which consists of 24 cases, an ID, and 5 variables per case. Suppose we wish to analyze each variable with the ANOVA program LIZ2WAY.TWO. We will need 5 CUSTOM files, one for each variable. The following sequence of responses will produce these files, called ANOVADAT.1 through ANOVADAT.5. (NOTE: This data is for illustration only. It is NOT APPROPRIATE FOR AN ANOVA ANALYSIS.)

TABLE 3.3

COL 1 CASE	COL 2 ID1	COL 3 VAR 1	COL 4 VAR 2	COL 5 VAR 3	COL 6 VAR 4	COL 7 VAR 5
1	1	.89	6.2	.97	5.3	9.2
2	1	1.32	8.4	3.22	8.4	6.3
3	1	3.17	5.9	4.5	3.7	8.4
4	1	4.51	8.8	6.8	4.5	7.3
5	2	6.72	9.9	2.3	8.6	2.1
6	2	3.13	7.3	.5	3.3	7.6
7	2	4.15	8.8	.7	4.8	7.9
8	2	9.15	4.3	9.0	8.3	6.2
9	3	7.17	3.8	6.1	4.9	2.8
10	3	3.52	8.5	5.3	4.4	6.9
11	3	6.17	3.7	3.7	6.6	4.9
12	3	1.81	5.4	8.2	8.6	2.0
13	4	9.93	3.6	7.4	4.6	4.2
14	4	3.14	9.3	6.5	6.1	8.4
15	4	1.51	8.4	5.3	8.4	7.1
16	4	2.72	9.3	6.9	7.7	4.9
17	5	1.36	6.7	5.6	6.2	5.1
18	5	3.57	9.4	1.8	3.7	3.3
19	5	5.91	6.9	2.7	5.2	8.9
20	5	6.66	9.7	7.2	8.7	6.3
21	6	7.83	3.2	2.3	7.2	9.4
22	6	1.31	6.5	2.9	7.7	4.9
23	6	4.22	8.3	1.6	4.2	2.8
24	6	5.13	4.1	4.1	7.7	4.0

NOTES: Data with ID=1 is intended for cell 1,1
Data with ID=2 is intended for cell 1,2
Data with ID=3 is intended for cell 1,3
Data with ID=4 is intended for cell 2,1
Data with ID=5 is intended for cell 2,2 and
Data with ID=6 is intended for cell 2,3

After loading program FILEPROD.AFM, choose the option ASSEMBLE FILES from the main menu. Program queries and user responses follow: (as usual, press <ENTER> after each response).

Q ENTER NAME OF
AGGREGATE (INPUT) FILE
R ANOVA.DAT
Q DESTINATION (OUTPUT) FILE
R ANOVADAT.1

```
Q  NUMBER OF COLUMNS PER CASE IN ANOVA.DAT
R  7
Q  NUMBER OF COLUMN SEARCHES TO BE MADE ON ANOVA.DAT
R  0
Q  FORMAT FOR DESTINATION FILE
   NUMBER OF VARIABLES (COLUMNS) TO
   COPY FROM ANOVA.DAT
R  1
Q  DELETE CASES WITH MISSING DATA (Y/N)?
R  Y
Q  VARIABLE NUMBER 1 FOR OUTPUT FILE
   IS FROM ANOVA.DAT COLUMN NUMBER
R  3
```

(The first file is assembled.)

```
Q  NAME OF NEXT DESTINATION FILE
R  ANOVADAT.2
Q  VARIABLE NUMBER 1 FOR OUTPUT FILE
   IS FROM ANOVA.DAT COLUMN NUMBER
R  4
```

(The second file is assembled.)

```
Q  NAME OF NEXT DESTINATION FILE
R  ANOVADAT.3
Q  VARIABLE NUMBER 1 FOR OUTPUT FILE
   IS FROM ANOVA.DAT COLUMN NUMBER
R  5
```

(The third file is assembled.)

```
Q  NAME OF NEXT DESTINATION FILE
R  ANOVADAT.4
Q  VARIABLE NUMBER 1 FOR OUTPUT FILE
   IS ANOVA.DAT COLUMN NUMBER
R  6
```

(The fourth file is assembled.)

```
Q  NAME OF NEXT DESTINATION FILE
R  ANOVADAT.5
Q  VARIABLE NUMBER 1 FOR OUTPUT FILE
   IS ANOVA.DAT COLUMN NUMBER
R  7
```

(The last file is assembled.)

```
Q  NAME OF NEXT DESTINATION FILE
R  NONE
```

(Control returns to the MAIN MENU.)

To illustrate what the resulting files would look like, this is the list for ANOVADAT.3, where data was taken from column 5 of the AGGREGATE file ANOVA.DAT:

RECORD	DATA	RECORD	DATA
1	.97	13	7.4
2	3.22	14	6.5
3	4.5	15	5.3
4	6.8	16	6.9
5	2.3	17	5.6
6	.5	18	1.8
7	.7	19	2.7
8	9.0	20	7.2
9	6.1	21	2.3
10	5.3	22	2.9
11	3.7	23	1.6
12	8.2	24	4.1

The menu item "PRODUCE FILES--SAMPLE COUNT" should be used only when the data in the AGGREGATE file originally has been arranged in the the proper order. Its function is similar to the preceding option, except that it is designed to count the number of consecutive cases which match a given target set. When a first case is encountered which matches a new target set, the data written to the data file from cases matching the previous target set is preceded by the sample count, or number of cases which matched the target set. This option is useful for assembling CUSTOM files which must have a sample count, such as files for LIZANOVA.NEQ, LIZ1WAY.ONE, etc.

EXAMPLE 5

Suppose we have an AGGREGATE file ANOVA.DAT similar to the one of Table 3.3, except that case 24 is missing. The same responses as in Example 4 will produce 5 CUSTOM files ANOVADAT.1 through ANOVADAT.5 with sample counts. ANOVADAT.3 would appear as follows:

RECORD	DATA	RECORD	DATA
* 1	4	15	8.2
2	.97	* 16	4
3	3.22	17	7.4
4	4.5	18	6.5
5	6.8	19	5.3
* 6	4	20	6.9
7	2.3	* 21	4
8	.5	22	5.6
9	.7	23	1.8
10	9.0	24	2.7
* 11	4	25	7.2
12	6.1	* 26	3
13	5.3	27	2.3
14	3.7	28	2.9
		29	1.6

The starred records are those which contain the sample counts.

Other uses for FILEPROD.AFM are indicated below in the section on files manipulation.

The remainder of this section is devoted to the editing and manipulation of AGGREGATE files.

EDITING AGGREGATE FILES: AGGREGATE files may be edited using the EDITOR. All features of the CUSTOM FILE MODE section of the EDITOR can be used. In addition, there are some special features available in the AGGREGATE FILE MODE section.

When you enter the AGGREGATE FILE MODE section, the EDITOR immediately displays the query

Q NUMBER OF COLUMNS/CASE?

Respond with

Number of columns per case.

If you plan to use such features as LIST CASE and DELETE CASE then each case must have a unique case number which must appear in column 1.

CREATING AN AGGREGATE FILE: AGGREGATE files are created by using the EDITOR, KEY I--INPUT RECORDS. Each case in an AGGREGATE file must appear in consecutive records with column 1 appearing first, column 2 appearing second, etc.

MERGING FILES: Files are merged using the program MERGFILE.AFM, Diskette 2. The process of merging files involves three separate files:

the first file to be merged,
the second file to be merged, and
the third file which is the merged (or output) file.

The first and second files contain data to be merged. They must contain the same number of cases arranged in exactly the same order. If only one disk drive is available, all three files will be on the same diskette. If two drives are available, at least two files must be on the same diskette.

You must provide the following information initially:

(1) The name of the first file;
(2) The total number of columns per case in the first file;
(3) The name of the second file;
(4) The total number of columns per case in the second file; and
(5) The name of the merged file which will be created.

You then must specify:

(6) The number of columns which are to be merged from the first file;
(7) Which columns from the first file are to be merged; and
(8) The same information as in (6) and (7) for file 2.

If M columns are merged from the first file and N columns are merged from the second, the merged file will contain M + N columns; the first M columns come from file 1 and the second N come from file 2.

You may control the order in which columns from each file appear by noting that the order of appearance of columns in the merged file is the same as the order of their presentation to the program.

Example: Suppose 4 columns are to be merged from the first file, e.g., columns 1, 2, 5, 7. If you present the column numbers to MERGFILE in the order

1, 5, 7, 2

then these columns will appear in the new file as follows:

COL # IN FILE 1	COL # IN OUTPUT FILE
1	1
5	2
7	3
2	4

The same convention applies with respect to the second file.

DELETING CASES: If each case has a unique number which appears in column 1, then individual cases may be deleted using the DELETE CASE option of the EDITOR.

A group of cases which have an identical ID number may be deleted using the program FILEPROD.AFM. The ID numbers are used as targets. When presenting the list of target sets to the program, simply omit the ID number of the unwanted cases. You may save all columns, or any desired subset thereof, to the new file. A new file will be created in which the unwanted cases will not appear.

REORDERING VARIABLES (COLUMNS): If you wish to change the order of appearance of columns in a file, then a new file should be created with FILEPROD.AFM. Search 0 columns and specify all columns from the source file are to be written to the destination file (i.e., answer the query NUMBER OF VARIABLES TO COPY FROM [filename] with the total number of columns in the file). Then present the column numbers to the program in the order you wish them to appear in the new file. For example, if you wanted Column 7 in the old file to become Column 5 in the new file you would answer the query

```
Q  VARIABLE NUMBER 5 FOR OUTPUT FILE IS
   FROM [filename] COLUMN NUMBER?
R  7
```

DELETING VARIABLES (COLUMNS): Variable deletion is accomplished using the program DELETE.VAR, Diskette 1 of the CORE PACKAGE.

You must provide:

1. Name of source file (AGGREGATE file or CUSTOM file from which variables are to be deleted).
2. Total number of columns in the file
3. Number of variables to be deleted
4. Column numbers for variables to be deleted
5. Name of destination file (new file to be created; this file will consist of all columns of the source file except those you have specified for deletion).

OUTPUT: Destination file

ADDING VARIABLES (COLUMNS): Use the EDITOR to create an AGGREGATE file consisting only of the observations of the variables you wish added to an existing AGGREGATE file. The file you create must contain exactly the same number of cases, in the same order, as the AGGREGATE file to which it is to be added. Then follow the instructions above to merge the two files. The existing file would normally be the first file to be merged.

ADDING CASES: This may be done with the EDITOR for individual cases. Use the INPUT RECORDS option.

If two AGGREGATE files contain the same number of columns, and if corresponding columns contain observations of the same variable (or ID number or case), then one file may be added to the other using the APPEND RECORDS option of the CUSTOM FILE ASSEMBLY program. There is a limit of 4000 records which may be appended at one time.

LISTING AGGREGATE FILES (PRINTER): AGGREGATE files may be listed by case using KEY P of the AGGREGATE FILE MODE section of the EDITOR. You may enter and have printed a list of labels which identifies the contents of each column.

COPYING FILES: CUSTOM or AGGREGATE files may be copied using the APPEND RECORDS option of the CUSTOM FILE ASSEMBLY program.

AGGREGATE files could be copied using FILEPROD.AFM. Search 0 columns, answer the query

Q NUMBER OF VARIABLES TO BE COPIED FROM [FILENAME]?

with the number of columns in the AGGREGATE file, and then present the column numbers to the program in the order 1, 2, . . . n where n is the total number of columns.

If you have more than one disk drive, it is usually possible to COPY files using the COPY feature of the Disk Operating System. (The COPY feature works with only one disk drive on some computers.)

SORTING AND ORDERING FILES: If cases have proper ID numbers, sorting and ordering may be done with FILEASSM.AFM which is described above.

ADDING VARIABLES TO REGRESSION FILES: A CUSTOM regression analysis file of K variables may be regarded as an AGGREGATE file of K columns for the purpose of adding, deleting, or transforming variables. Remember that in a regression file, the variable in the last column is regarded as the dependent variable. Although this convention may be changed by certain options in some multiple regression programs, it is best to always keep the dependent variable in the last column. Thus, when adding independent variables to CUSTOM regression files with program MERGFILE, always name the file containing the variables to be added as the first file. This will cause the added variables to appear in the first column of the destination file.

CHAPTER IV
CUSTOM FILE ASSEMBLY

CUSTOM files made for one program may be modified to some extent to make them compatible with other programs. For example, files made for LIZ2WAY.TWO, which contain no cell (or sample) count numbers, can be made compatible with LIZMEAN.MCP or similar programs by inserting the sample numbers with the EDITOR or using the APPEND RECORDS--SAMPLE COUNT feature described below. Conversely, sample counts may be removed with the EDITOR or with CUSTOM FILE ASSEMBLY (LIZMOD.CFM). There is, however, a definite limit to the amount of modification which can be made on CUSTOM files, and this is their most undesirable feature. If a set of data is to be used in several statistical programs, it is usually advisable to construct an AGGREGATE file for the data (see Chapter V).

The CUSTOM FILE ASSEMBLY subprogram is named LIZMOD.CFM.

APPEND RECORDS

Records from a specified file are appended onto the file you specify. Not more than 4000 records may be appended at once. You must specify the first and last record of a consecutive block of records which are to be appended.

As an example, suppose file PERSPRO.DAT contains 1,820 records and file PERSPRO.BIG contains 18,237 records. If you specify that you want records 1 to 1820 of files PERSPRO.DAT appended to file PERSPRO.BIG, then all the records in the file PERSPRO.DAT will be transferred, in order, to file PERSPRO.BIG beginning at record 18,238. File PERSPRO.BIG will then contain 20,057 records.

If a very large CUSTOM or AGGREGATE file is being made up, then the file should be made up in small parts. Each smaller part can be completely edited, then added to the large file using this feature.

This feature can also be used to break large files down into smaller parts, and to copy files.

NOTE: If different diskettes are being used, then the two filenames involved in this operation may be the same. You must keep up with the order of the diskettes. You will be asked first to insert the diskette containing the records you want appended (source file), and then you will be asked to insert the diskette containing the file to which the records are to be appended (destination file). After inserting the diskettes, press <ENTER> or any alphanumeric key.

APPEND RECORDS--SAMPLE COUNT

This feature works just like APPEND RECORDS except the records are appended preceded by the sample count. The sample count is the number of records appended.

CHAPTER V
USING THE EDITOR

When using the EDITOR, the data file to be edited must be on the default drive (usually drive 0).

You have been provided with a practice data disk which contains several data files. One of these is for use in learning how to use the EDITOR. The data in it is meaningless otherwise. The file is aptly named BIGMESS.GOB. You may use this file when practicing with the EDITOR.

To load the EDITOR, place any LIZPACK program diskette in the drive and follow the instructions for loading LIZPACK. When the DIRECTORY appears, use the C key to advance to the DIRECTORY page where the EDITOR is listed. Press the proper key to load the EDITOR. You will immediately be asked for the name of the file you wish to edit. After entering the filename, the list of CUSTOM FILE MODE Control Keys will appear. Notice that key H is labeled "HELP!" Remember this; anytime you are using the EDITOR you can get the control key list by pressing H. (The EDITOR must be in the READY state for Key H to be enabled. To return to the EDITOR from HELP, press key B.)

Press key B. The EDITOR will enable all control keys and pass to the READY state. You may now begin editing by pressing a Control Key. The Control Keys cannot be used until the EDITOR is in the READY state, i.e., after pressing key B. (The one exception to this rule is key A; see the description below.)

<u>CUSTOM FILE CONTROL KEYS:</u>

KEY I (INPUT RECORDS)

Never input records into a file which contains a deletion flag! If records are input past a deletion flag, the deletion flag (979797) will be treated the same as any other piece of data. In addition, each record which is input after the deletion flag will be located one record location past where it should have been. This could lead to some very interesting results. If, however, you should find yourself in this situation, the file can be repaired. Find the deletion flag (979797) by listing the records (see instructions below, key L) and delete the deletion flag just as if it were any other record (using key D). This removes the deletion flag and places a deletion flag at the END of the file, which must then be removed by using key R, or is removed automatically when you exit the EDITOR.

The INPUT RECORDS option allows you to make a CUSTOM or AGGREGATE file. The EDITOR will display the first available record number in the file. This record number will be 1 if the file is empty and will be the next available record otherwise. Begin entering records (press <ENTER> after each record).

See Chapter III for more specific information on creating AGGREGATE files. You may terminate the process of entering records by entering the number

989898

which is a signal to EDITOR that you wish to return to the READY state. (NEVER use 989898 as a data word!)

KEY C (CHANGE RECORDS)

You may change any number of consecutive records. You must specify the first and last record you wish to change. You must change all the records you specify. When all records are changed, the EDITOR returns to the READY state.

KEY D (DELETE RECORDS).

This key deletes a block of consecutive records. Specify the first and last record you wish to delete. (If you wish to delete only one record, type the record number, a comma, and repeat the record number.) The same conventions apply here as with the MINI-EDITOR. See Key R below.

KEY R (REMOVE DELETION FLAGS)

You must use this key if you have made deletions and plan to input more records into your file. Deletions are handled in the same manner as with the MINI-EDITOR (See Chapter II).

You may use this key at any time to remove deletion flags; however, when you leave the EDITOR, either by an EXIT LIZPACK or by reloading the DIRECTORY, all deletion flags are automatically removed. If you input data past a deletion flag, the automatic removal process will either not work or will function incorrectly.

KEY J (INSERT RECORDS)

You must specify the first record number of an insertion, and the number of consecutive records to be inserted. You must insert the specified number of records before control returns to the READY state.

KEY L (LIST RECORDS)

You must specify the first and last number of a block of consecutive records you wish displayed. The READ MULTIPLE controls the number of records to be displayed at one time (no more than 14). After a block of records has been displayed, press L to continue listing. After all records are listed, press L to return to the READY state. In the LIST mode the following EDIT keys are enabled: KEY C, KEY D, and KEY I. After performing any editing, you will return to the LIST mode. Pressing KEY X at any time when records are displayed terminates listing, and returns the EDITOR to the READY state. Note that this is NOT the normal function of key X.

KEY P (PRINT FILE)

The entire data file is sent to the printer by record number and record contents.

KEY X (RELOAD DIRECTORY)

If KEY X is pressed when in the READY state, deletion flags are removed, files are closed, and the DIRECTORY is reloaded.

If you press key X while in the LIST mode, control is returned to the READY state.

KEY E (EXIT LIZPACK)

Deletion flags are removed, files are closed, and LIZPACK operation is ended.

KEY S (INPUT-AUTO SAMPLE COUNT)

For some files, such as those for one-way ANOVAs, two-way ANOVAs with non-equal cell frequencies, and the Means Analysis package LIZMEAN.MCP, the observation for the sample must be preceded by the sample count. This key performs this task automatically; use this key just like KEY I except when all data for a sample (or cell) has been keyed in, enter 999999, and the sample count will automatically be entered in the proper record. KEY I instructions apply otherwise.

KEY A (LIST AGGREGATE CONTROL KEYS)
(ENABLED ONLY WHEN CUSTOM CONTROL KEY LIST IS DISPLAYED)

When you press this key you will be asked for the number of columns per case (see Chapter III). If you want to use the control keys without editing an AGGREGATE FILE, enter any number. The AGGREGATE FILE Control Key List is displayed.

<u>AGGREGATE FILE CONTROL KEYS:</u>

KEY B (BEGIN)

The control keys are enabled and the EDITOR enters the READY state.

KEY C (DISPLAY CUSTOM FILE CONTROL KEYS)
(ENABLED ONLY WHEN AGGREGATE CONTROL KEYLIST IS DISPLAYED)

The CUSTOM file control keys are displayed.

KEY L (LIST CASE)

The case you specify is located and is listed on the screen. If the case contains more than 14 records, press KEY L to continue listing. KEYS C, D, J, and X are enabled. When all records have been listed, KEY L returns the EDITOR to the READY state. KEY X terminates listing at any time records are displayed.

If this option is used, each case in the AGGREGATE file must have a unique case number which must be located in column 1.

KEY D (DELETE CASE)

The case you specify is located and deleted. Deletion flags are inserted at the end of the file.

If this option is used, each case in the AGGREGATE file must have a unique case number which must be located in column 1.

KEY P (PRINT FILE)

Sends contents of the AGGREGATE file to the printer. After pressing Key P, the EDITOR will print

(1) The name of the file,
(2) The number of columns in the file, and
(3) The total number of records in the file.

You will then be asked if you wish to enter labels for the columns. If you answer "NO" to this query, the file will be printed immediately. If you answer "YES" to this query, you will then be asked to enter a label for each column. This label could be CASE NUMBER, the name of the variable which occupies the column, or any other description which will serve to remind you of the contents of the column. You may use up to 50 characters per label. You may use any characters EXCEPT A COMMA OR A COLON. This option allows you to make a log of the contents of the columns of an AGGREGATE file. Each column number and the label for that column will be printed on a separate line.

The file is printed by case, record number, and record contents.

KEY N (EDIT NEW FILE)

Deletion flags are removed from the file you are editing and the file is closed. You will then be asked to enter the name of the new file which you wish to edit.

CHAPTER VI
THE LIZPACK CORE PACKAGE PROGRAMS

Throughout this manual, you will be given a list of information required by LIZPACK programs. This information is entered in response to LIZPACK prompts and queries. Virtually all LIZPACK statistical programs ask for the name of the study, name of investigator, date of run, and the name of the data file to be used. In the program descriptions which follow, these items may be omitted from the lists of information you are required to provide.

1. FREQUENCY DISTRIBUTIONS; RANDOM SAMPLING
(Program LIZSAMP.FREQ, DISKETTE 1)

FREQUENCY DISTRIBUTION OPTION: MENU OPTION 3. The frequencies of the sample over a set of class intervals are measured. You will be asked to specify the number of class intervals on which you want frequencies counted (see instructions below). Choose this number to best suit the type of distribution you desire (continuous data, integer data, etc.). If you have integer data and you wish to count the number of times a given integer appears, the chosen number of class intervals should be (range of the sample + 1).

After choosing the menu options FREQUENCY DISTRIBUTION or RANDOM SAMPLING, you must answer the standard questions for the header. If you answer the query

IS THIS RUN FROM A CUSTOM FILE?

with NO, you must provide

(1) the total number of columns in the AGGREGATE file, and

(2) the file column number which contains the variable on which the frequency distribution is to be made.

If you have chosen the FREQUENCY DISTRIBUTION option, the STEPS list will appear. If you know the maximum and minimum values of the observations in your example, you may save time by choosing STEP 2 and entering them; otherwise choose STEP 1 and let the system find them.

Press Key 3 for the third step. You will be given the maximum, minimum, and range (maximum minus minimum). Enter your choice of the number of class intervals to be used. The length of the class interval will appear. The program will set up the class intervals so that the minimum value is the midpoint of the first class interval. An observation is considered to be in a class interval if it is at least equal to the left endpoint (ENDPOINT A) of the interval and is less than the right endpoint (ENDPOINT B) of the interval.

After STEPS reappears, press Key 4. A frequency distribution will be performed, a table will be printed and the menu will reappear.

OUTPUT: The Frequency Distribution Table lists class number, endpoints of the class interval, and frequency on the class interval for each class interval. Also given is the number of cases, mean, variance, and standard deviation of the sample.

RANDOM SAMPLING OPTION: This option is not available from an AGGREGATE file. In this option, random samples, or more properly, random sub-samples, are drawn directly from the disk data file (the disk drive will operate continuously during this procedure). The distribution of random sample means is measured exactly as in the FREQUENCY DISTRIBUTION option. In general, you need some estimate of the mean and standard deviation of the parent population. These estimates could, for example, be obtained by doing a frequency distribution as above. You must specify the maximum and minimum mean you expect to measure. In general, the probability of a mean occurring more than 4 standard units (on either side) away from the population mean is near zero.

In general, the data file will need to contain a moderately large sample from the parent population (at least 100 observations). To obtain a good distribution, a rather large number of random samples must be drawn. It is suggested that you consult a good applied statistics textbook for more detailed information on this procedure.

This routine will draw and process about 50 random samples consisting of 10 observations each in 1 minute.

After choosing this option and going through program initialization, the STEPS list will appear. The steps must be done in order.

STEP 1: Specify the maximum and minimum sample mean expected.
STEP 2: Specify the number of class intervals for the frequency distribution.
STEP 3: Specify the number of samples to be drawn, and the number of observations in each sample.

OUTPUT: The initial output for this option is the same as for the FREQUENCY DISTRIBUTION option. The remaining output is the "Estimated Population Sigma" or estimated standard deviation of the population. This is computed as the square root of the mean of the sample variances. The population mean is estimated as the mean of the random sample means.

THEORETICAL FREQUENCY DISTRIBUTION: This option must be exercised immediately AFTER doing a FREQUENCY DISTRIBUTION or RANDOM SAMPLING. By using the estimates of the population parameters obtained from either process, it computes the actual frequency distribution which will occur if your sample is drawn from a normal population. By comparing the results of this process with those of either of the options above, you should have a good estimate of whether or not your data follows the normal distribution.

OUTPUT: Same format as for the FREQUENCY DISTRIBUTION option.

HISTOGRAM and FREQUENCY POLYGON options may be exercised after doing any frequency distribution. Either graph will remain on the screen for your inspection until an alphanumeric key is pressed; control is then returned to the menu.

CUSTOM DATA FILE: A CUSTOM file is a one-column AGGREGATE file which contains the observations arranged in consecutive records.

MISSING DATA: A CUSTOM file may not contain missing data symbols. If the run is from an AGGREGATE file, the cases containing a missing data symbol will be ignored.

2. DESCRIPTIVE STATISTICS, Z-SCORES
Program LIZDESC.DES, Diskette 1

DESCRIPTIVE STATISTICS: Computes descriptive statistics for each of an arbitrary number of samples. You must provide:
(1) name of data file, and
(2) number of samples

OUTPUT: Sample number, number of cases, minimum, maximum, range, mean, variance, standard deviation, coefficient of variability, standard error, skewness, and kurtosis for each sample.

Z-SCORES: This option must be exercised after DESCRIPTIVE STATISTICS. You do not have to answer any questions.

OUTPUT: Sample number, sample mean, raw score, and Z-score for each observation in each sample.

DATA FILE: Observations for each sample must appear in consecutive records. The observations for each sample must be preceded by the sample count, i.e., the number of observations in the sample. The file maker included with this program automatically inserts the sample count, so you need only enter the observations from each sample. Follow the instructions given by the file maker; keying in 999999 after all data for a sample has been entered causes the sample count to be entered.

3. MEANS COMPARISON PACKAGE
Program LIZMEAN.MCP, Diskette 1
Program LIZMEAN.AGG, Diskette 3

You must provide the following information:
(1) name of population unit being studied,
(2) name of variable measured,
(3) data file name, and
(4) number of samples in the data file.

If you are using the program LIZMEAN.AGG, you must also provide:
(5) the total number of columns in the AGGREGATE file,
(6) the file column number for the independent variable,
(7) the low and high values for the independent variable (see Chapter X), and
(8) the file column number for the dependent variable.

After computing the appropriate information, the list of samples and the OPTIONS MENU will appear. The options are described below. AFTER MANY OF THESE OPTIONS ARE EXERCISED, DATA WILL BE DISPLAYED ON THE SCREEN UNTIL YOU PRESS ANY ALPHANUMERIC KEY; control is then returned to the menu.

DISPLAY MEANS: Means are displayed on the screen. Means are arranged in order of decreasing magnitude.

COMPARE MEANS: All possible pairs of means of samples or SAMPLES AVAILABLE list are compared. The sample numbers, denominator degrees of freedom, and F-ratios (square of Student's T) are displayed for each pair of samples.

ANOVA (SCREEN): An ANOVA for all available samples is displayed on the screen. This is a one-way ANOVA for all samples on the SAMPLES AVAILABLE list.

DELETE SAMPLE: The sample number you specify is deleted from the SAMPLES AVAILABLE LIST.

RESTORE ALL SAMPLES: All sample numbers are restored to the SAMPLES AVAILABLE list.

ENTER CONTRASTS: Suppose there are N samples on the SAMPLES AVAILABLE list. Then a contrast is an N-tuple

$$C = (C_1, C_2, . . ., C_n)$$

of numbers with $C_1+C_2+ . . . +C_n = 1$. Each C_i is called a "coefficient." If $D = (D_1, D_2, ... D_n)$ is another contrast, then C and D are ORTHOGONAL if $C_1D_1+C_2D_2+...+C_nD_n = 0$.

The ENTER CONTRASTS option allows the user to enter up to N-1 contrasts, where N is the number of samples on the SAMPLES AVAILABLE list. The user must specify the number of contrasts to be entered.

Contrasts entered for the option ANOVA W/CONSTRASTS must be orthogonal.

Contrasts entered for Scheffe's test in the option MULTIPLE COMPARISON TESTS need not be orthogonal. Contrasts remain in the contrast table until new contrasts are entered, or until the POLYNOMIAL TRENDS option is exercised.

POLYNOMIAL TRENDS: provides a test to determine if the means follow the trend of a polynomial of degree M, with M <= N-1, where N is the number of samples available. Each mean is assumed to correspond to an abscissa. The abscissas are assumed to be evenly spaced. The sample pointers are set so that the samples means will correspond to the natural ordering of their abscissas. For example, if sample mean number 3 corresponds to the smallest abscissa, set sample point 1 equal to 3. This procedure must be followed for each sample pointer when the program asks you to SET SAMPLE POINTERS. You must also specify the degree of the polynomial you wish to use.

The polynomials used are polynomials which are orthogonal over a discrete range. The contribution of individual terms partition the sum of squares. An ANOVA table is printed which shows the contribution of each term and the F-ratios for test of significance of each term. The symbol REM stands for "remainder." This option will destroy any user entered contrasts.

POLYNOMIAL TRENDS--UNEQUAL SPACING: Same as above except you must enter the abscissa to which each mean corresponds. Abscissas need not be equally spaced. This option will destroy any user entered contrasts.

MULTIPLE COMPARISON TESTS:

(a) TUKEY'S TEST: You must enter

(1) The percent significance level at which you wish to test.

(2) The appropriate Q-value (this is obtained from a table of studentized ranges).

This option computes confidence intervals for each pair of means on the MEANS AVAILABLE list.

(b) SCHEFFE'S TEST: You must enter appropriate contrasts by means of the ENTER CONTRASTS option before using this option. Failure to do so may terminate execution of the program with an accompanying ERROR message. You must also enter

(1) Percent significance level you are working at.
(2) The F-value requested (this is obtained from an F-table).

The confidence interval is computed for each contrast previously entered by means of the ENTER CONTRASTS option.

CUSTOM DATA FILE: Same arrangement as for LIZDESC.DES.

4. ONE-WAY ANALYSIS OF VARIANCE
Program LIZ1WAY.ONE, Diskette 1
Program LIZ1WAY.AGG, Diskette 3

This program outputs to the printer a standard one-way ANOVA for an arbitrary number of samples. You must provide

(1) name of population unit being studied,
(2) name of variable on which observations were made,
(3) name of data file, and
(4) number of samples.

If you are using the program LIZ1WAY.AGG, you must also provide:
(5) the total number of columns in the AGGREGATE file,
(6) the file column number for the independent variable,
(7) the low and high values for the independent variable (see Chapter X), and
(8) the file column number for the dependent variable.

This is a standard one-way ANOVA program which provides statistics for the test of the hypothesis that the mean of K samples are equal. The sum of squares is decomposed in the classical manner. Computations are for a fixed effect model.

OUTPUT: A standard one-way ANOVA table which contains the among, within, and total sum of squares (SS) and their degrees of freedom (DF); mean sum of squares (MSS or MS) for among and total; F-ratios; means for each sample and the grand mean.

CUSTOM DATA FILE: Same arrangement of data as LIZDESC.DES.

5. ONE-WAY ANALYSIS OF COVARIANCE
Program LIZ1WAY.ANC, Diskette 1
Program LIZ1ANC.AGG, Diskette 3

This program outputs to the printer a standard one-way analysis of covariance table for an arbitrary number of samples with one covariate. You must provide the same information as for LIZ1WAY.ONE or LIZ1WAY.AGG, and in addition, you must provide
(9) the name of the covariate.

If you are using LIZ1ANC.AGG, then you must provide
(10) the file column number for the covariate.

OUTPUT: Analysis of covariance table containing:
(1) total sum of squares of covariate, SS(X,X);
(2) total sum of squares of dependent variable, SS(Y,Y);
(3) total sum of products of dependent variable and covariate, SS(X,Y);
(4) adjusted total sum of squares for dependent variable, ADJ.SS(Y,Y);
(5) sum of squares for treatment effect and error (residual);
(6) adjusted sum of square for treatment effect and error (residual);
(7) adjusted mean sum of square for dependent variable, ADJ. MSS(Y,Y);
(8) appropriate degrees of freedom; and
(9) regression coefficients.

CUSTOM DATA FILE: Suppose X is the covariate, Y is the dependent variable, and N_i is the number of observations of each of X and Y for sample i. The data must appear in the data file as follows:

N_1

X_1

Y_1

X_2

.

.

.

X_{N1}

Y_{N1}

If you use the file maker included with the program, you must provide
(1) number of samples, and
(2) number of observations for each sample.

6. ANALYSIS OF VARIANCE: RANDOMIZED BLOCK EXPERIMENTS
Program LIZANOVA.RAN, Diskette 1
Program LIZRAND.AGG, Diskette 3

The randomized block experiment is a frequently used special case of the two-way ANOVA, and in fact, could be analyzed as such. A special program is included for this type of ANOVA because of its ease of use and the difference in the type of means tables which are output to the printer.

This program outputs to the printer a standard ANOVA table for randomized block experiments. You must provide
(1) name of treatments being studied,
(2) name of blocks,
(3) name of variable,
(4) name of data file,
(5) number of treatments, and
(6) number of blocks.

For program LIZRAND.AGG, you must also provide
(7) the file column numbers for two independent variables (the row and column on block variables), and the low and high values for each (see Chapter X).
(8) the file column numbers for the dependent variable.

OUTPUT: ANOVA table with sum of squares for total, treatment, blocks and errors with their degrees of freedom. Mean sum of squares for treatment, blocks, and error. F-ratios for treatment and block effects. Table of means for each level of treatment and blocks. Grand mean and standard deviation.

CUSTOM DATA FILE: The data is thought of as being arranged in a matrix with observations for blocks as columns and treatments as rows:

TREATMENT NUMBER	BLOCK NUMBER 1	2	3	. . .	c
1	Y_{11}	Y_{12}	Y_{13}	. . .	Y_{1c}
2	Y_{21}	Y_{22}	Y_{23}	. . .	Y_{2c}
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
r	Y_{r1}	Y_{r2}	Y_{r3}	. . .	Y_{rc}

Here Y_{ij} is the observation for treatment i and block j. The data must appear in the data file in the order

Y_{11} (record 1 of file)
Y_{12} (record 2 of file)
.
.
.
Y_{1c}
Y_{21}
Y_{22}
.
.
.
Y_{2c}
.
.
.
Y_{r1}
.
.
.
Y_{rc} (last record of file)

7. ANALYSIS OF COVARIANCE: RANDOMIZED BLOCK EXPERIMENTS

Program LIZANCOV.RAN, Diskette 1
Program LIZANRAN.AGG, Diskette 3

Outputs to the printer a standard analysis of covariance table for randomized block experiments with one covariate. You must provide all the information required in LIZANOVA.RAN or LIZRAND.AGG and
(9) the name of the covariate.

If you are using program LIZANRAN.AGG, you must also provide
(10) the file column for the covariate.

OUTPUT: This program does an ANOVA on both the covariate and dependent variable and outputs the same statistics for each as described under LIZANOVA.RAN above.

The analysis of covariance table contains

(1) total sum of squares for the covariate and dependent variable and the total sum of products for the covariate and dependent variable,

(2) sum of squares and products for treatment, blocks, and error,

(3) adjusted sum of squares for treatment and error,

(4) all appropriate degrees of freedom,

(5) F-ratios for treatment effect.

CUSTOM DATA FILE: Let X denote the covariate and Y be the variable. Data is thought of as being arranged as follows:

TREATMENT NUMBER	BLOCK NUMBER 1	2	3	. . .	c
1	X_{11} Y_{11}	X_{12} Y_{12}	X_{13} Y_{13}	. . . . . .	X_{1c} Y_{1c}
2	X_{21} Y_{21}	X_{22} Y_{22}	X_{23} Y_{23}	. . . . . .	X_{2c} Y_{2c}
. . .	. . .	. . .	. . .		. . .
r	X_{r1} Y_{r1}	X_{r2} Y_{r2}	X_{r3} Y_{r3}	. . . . . .	X_{rc} Y_{rc}

The data is arranged in the file as follows:

X_{11} (record 1)
Y_{11} (record 2)
X_{12}
Y_{12}

.
.
.

X_{1c}
Y_{1c}
X_{21}
Y_{21}

.
.
.

X_{2c}
Y_{2c}

.
.
.

X_{r1}
Y_{r1}

.
.
.

X_{rc}
Y_{rc} (last record)

8. TWO-WAY ANALYSIS OF VARIANCE--EQUAL CELL FREQUENCIES
Program LIZ2WAY.TWO, Diskette 1
Program LIZ2WAY.AGG, Diskette 3

This is a standard 2-way ANOVA with equal cell frequencies. You must provide

(1) name of row factor,
(2) name of column factor,
(3) name of variable,
(4) observations per cell,
(5) number of rows, and
(6) number of columns.

For LIZ2WAY.AGG, you must also provide

(7) the file column numbers for the two independent variables with the low and high values for each (see Chapter X).
(8) the file column number for the dependent variable.

OUTPUT: A two-way ANOVA table for fixed, random, and mixed effects models. The format for this table is suggested by Kleinbaum and Kupper, Applied Regression Analysis and Other Multivariable Methods, p. 326. Consult this reference for interpretation of F-ratios for random and mixed effects models. The following abbreviations are used:

CR (columns random)
CF (columns fixed)
RR (rows random)
RF (rows fixed)

The ANOVA table contains appropriate sum of square, mean sum of squares, and degrees of freedom for main effects, interaction, error and total.

Also printed are tables of means for each level of the row and column factor and for each cell together with the grand mean and standard deviation.

CUSTOM DATA FILE: Data for a 2-way ANOVA is arranged in cells. For example, for a 2 x 3 run we would have

ROW FACTOR	COLUMN FACTOR 1	2	3
1	Y_{111}, Y_{112}, ...,Y_{11n}	Y_{121}, Y_{122}, ...,Y_{12n}	Y_{131}, Y_{132}, ...,Y_{13n}
2	Y_{211}, Y_{212}, ...,Y_{21n}	Y_{221}, Y_{222}, ...,Y_{22n}	Y_{231}, Y_{232}, ...,Y_{23n}

where N is the number of observations per cell. Y_{ijk} is the K^{th} observation of the variable Y corresponding to row factor i and column factor j. Arrangement of data for a general P x M ANOVA is similar. The first two subscripts are the cell numbers. The data must be arranged in the file so that

data from cell 1,1 appears first,
data from cell 1,2 appears next,
. . .
data from cell 1,m appears next,
data from cell 2,1 appears next,
. . .
. . .
data from cell 2,m appears next,
. . .
data from cell p,m appears last.

9. TWO-WAY ANALYSIS OF COVARIANCE: EQUAL CELL FREQUENCIES
Program LIZ2WAY.ANC, Diskette 1
Program LIZ2ANCV.AGG, Diskette 3

You must provide the information required by LIZ2WAY.TWO and LIZ2WAY.AGG, and

(9) the name of the covariate.

If you are using LIZ2ANCV.AGG, you must provide

(10) the file column number for the covariate.

OUTPUT: This program prints two-way ANOVA tables for both the covariate and independent variable together with tables of means for both. The output is in the same format as that for LIZ2WAY.TWO above.

The analysis of covariance table contains:

(1) sum of squares and products for the covariate and dependent variable for total, row and column effects, interaction, and error,

(2) sum of square for row + error, column + error, and interaction + error effects,

(3) adjusted sum of squares and adjusted mean sum of squares for row + error, column + error, interaction + error, and error,

(4) appropriate degrees of freedom,

(5) F-ratios for adjusted row, column, and interaction effects,

(6) Regression coefficients and standard errors for error, row, column, and interaction effects.

CUSTOM DATA FILE: Data for both the covariate, X, and variable Y, are arranged in cells as with the two-way ANOVA. Let X_{ijk} denote the k^{th} observation of the covariate for cell i,j, and N be the number of observations per cell. Arrangement of data in the file is as follows:

X_{111} (Record 1)
Y_{111}
X_{112}
Y_{112}
.
.
.
X_{11N}
Y_{11N}
X_{121}
Y_{121}
.
.
.

Thus, data for row 1 is read in first, data for row 2 is read in second, etc.

10. TWO-WAY ANALYSIS OF VARIANCE: NON-EQUAL CELL FREQUENCIES
Program LIZANOVA.NEQ, Diskette 1
Program LIZ2NEQ.AGG, Diskette 3

This is a standard program for a two-way analysis of variance with non-equal cell frequencies. A fixed effects model is assumed. When cell frequencies are non-equal, the sum of squares does not decompose orthogonally, hence an adjustment must be made. The method of adjustment used by this program is that given by Kleinbaum and Kupper, Applied Regression Analysis and Other Multivariable Methods, pp. 357-361. If the method of unweighted means does not apply to your data, you must use the regression model ANOVA (see description 23 below).

You must provide the following information:
(1) name of row factor,
(2) name of column factor,
(3) name of variable,
(4) number of rows, and
(5) number of columns.

For LIZ2NEQ.AGG, you must also provide
(6) the file column numbers for the two independent variables with the low and high values for each (see Chapter X), and
(7) The file column number for the dependent variable.

OUTPUT: The program prints a standard two-way ANOVA table containing appropriate sum of squares, mean sum of squares, degrees of freedom, and F-ratio.

The program also prints tables of means in the same format as those printed by LIZ2WAY.TWO.

CUSTOM DATA FILE: same as LIZ2WAY.TWO except that the data for each cell must be preceded by the number of observations for that cell. The file maker will automatically insert the number of observations per cell at the proper place, so that you will need only enter the observations for each cell. Data files for this program may be mass produced from AGGREGATE files--see the section on AGGREGATE FILES.

11. TWO-WAY ANALYSIS OF VARIANCE WITH BLOCK DESIGN
Program LIZBLOCK.2WY, Diskette 1
Program LIZBLOCK.AGG, Diskette 3

This is a special case of a three-way ANOVA.

Information you must provide is
(1) name of row factor,
(2) name of column factor,
(3) name of blocks,
(4) name of variable,
(5) observations per cell,
(6) number of rows, and
(7) number of columns.

If you are using LIZBLOCK.AGG, you must provide
(8) the file column numbers for three independent variables. These are the row, column, and block variables. You must give the low and high values for each.
(9) the file column number for the dependent variable.

This program is similar to LIZ2WAY.TWO with the exception that it assumes that the data within each cell contains exactly one observation from a block or group. An example of an experimental situation which gives rise to a block design is as follows:

Suppose an investigator wishes to measure the effect of rations on the amount of weight gained by bacon pigs. The number of rations being tested is three. The investigator has five pens available in each of which he places three pairs of bacon pigs. Each pair consists of one male and one female. The amount of weight gained by each pig is measured after a certain length of time and a 3 x 2 analysis of variance is performed (row factor: ration; column factor: sex). It is noticed that the means for different pens vary significantly. Thus each pen becomes a BLOCK. The precision of the two-way ANOVA will be increased if the effect of blocking is removed before performing the analysis.

The arrangement of data in a CUSTOM file for this program is very important. For the example above, data should be arranged as follows:

ROW (RATION)	COLUMN (SEX) 1-M	2-F
	Wt. gain, pen 1	Wt. gain, pen 1
	Wt. gain, pen 2	Wt. gain, pen 2
1	Wt. gain, pen 3	Wt. gain, pen 3
	Wt. gain, pen 4	Wt. gain, pen 4
	Wt. gain, pen 5	Wt. gain, pen 5
	Wt. gain, pen 1	Wt. gain, pen 1
	.	.
2	.	.
	.	.
	Wt. gain, pen 5	Wt. gain, pen 5
	Wt. gain, pen 1	Wt. gain, pen 1
	.	.
3	.	.
	.	.
	Wt. gain, pen 5	Wt. gain, pen 5

Thus the data must be arranged so that the data for a given block occupies the same relative position in each cell.

OUTPUT: The output for this program is the same as for LIZANOVA.TWO except that in the ANOVA table, the sum of squares and degrees of freedom for the block effect are printed.

CUSTOM DATA FILE: Same as LIZ2WAY.TWO subject to the arrangement of data by blocks, as discussed above.

12. TWO-WAY ANALYSIS OF COVARIANCE WITH BLOCK DESIGN

Program LIZANCOV.BLK, Diskette 1
Program LIZBLANC.AGG, Diskette 3

Information you must provide is
(1) name of row factor,
(2) name of column factor,
(3) name of blocks,
(4) name of variable,
(5) observations per cell,
(6) number of rows,
(7) number of columns,
(8) the name of the covariate, and
(9) the name of the variable.

If you are using LIZBLANC.AGG, you must provide the same information regarding the AGGREGATE file as for LIZBLOCK.AGG, and
(10) the file column number for the covariate. You will note that you must provide all the information required for LIZBLOCK.2WY, plus number (8).

A situation where the program might be used is as follows: consider the example in number 11 (LIZBLOCK.2WY) above. Not surprisingly, the weight gain in each bacon pig was found to depend on the initial weight of the pig. Thus, initial weight is used as the covariate with weight gain the variable.

The data for both the variable and covariate for each cell is subject to arrangement for LIZBLOCK.2WY. For example, for Ration 1 the scheme is

ROW (RATION)	COLUMN 1 (M)	2(F)
	INITWT,WTGAIN (PEN 1)	INITWT,WTGAIN (PEN 1)
1	INITWT,WTGAIN (PEN 2)	INITWT,WTGAIN (PEN 2)
	. . .	. . .
	INITWT,WTGAIN (PEN 5)	INITWT,WTGAIN (PEN 5)

OUTPUT: The output is the same as for the program LIZ2WAY.ANC except that in the analysis of covariance table, the sum of squares and products for the covariate and dependent variable for the block effect is printed along with their degrees of freedom.

NOTE: The ANOVA tables which are printed do not contain adjustments for the block effect.

CUSTOM DATA FILE: Same as LIZ2WAY.ANC subject to the arrangement of data by blocks as discussed above.

13. THREE-WAY ANOVA, EQUAL CELL FREQUENCIES
Program LIZANOVA.3WY, Diskette 3
Program LIZ3WAY.AGG, Diskette 3

This is a standard three-way ANOVA procedure. You must provide
(1) the name of factor A,
(2) the name of factor B,
(3) the name of factor C,
(4) the name of the dependent variable,
(5) the number of levels for each of Factors A, B, and C, and
(6) the number of observations per cell.

If you are using LIZ3WAY.AGG, you must also provide
(7) the total number of columns in the AGGREGATE file,
(8) the file column numbers for the three independent variables and the low and high values for each (see Chapter X), and
(9) the file column number for the dependent variable.

OUTPUT: The program prints a three-way ANOVA table containing
(1) the sum of squares total, error (residual), each factor, each two-way interaction, and the three way interaction together with the degrees of freedom of each,
(2) appropriate mean sum of squares, and
(3) F-ratios for the fixed and random effects models.

The program prints a separate table giving the F-ratios for the 6 mixed effects models. The F-ratios are computed according to the formulas given by Kleinbaum and Kupper, Applied Regression Analysis and other Multivariable Methods, page 367, Table 20.8.

The programs print three tables of means for
(1) each level of each factor,
(2) each level of each two-way interaction, and
(3) each cell.

CUSTOM DATA FILE: The data is arranged in cells, as in the case of a two-way ANOVA. The file-maker included in program LIZANOVA.3WY will guide you in the construction of the file. The data for all cells at level 1 of factor A is entered first, then all data for level 2 of factor A is entered, etc.

Within each level of factor A, the data for level 1 of factor B is entered first, then all data for level 2 of factor B is entered, etc.

Within each level of factor A and factor B, the data for each level of factor C is entered beginning with the lowest level and proceding to the highest level.

14. THREE-WAY ANOVA, NON-EQUAL CELL FREQUENCIES
Program LIZ3WAY.NEQ, Diskette 3
Program LIZ3NEQ.AGG, Diskette 3

The operation of these programs is the same as those of 13, with the following exception: these programs do not ask for the number of observations per cell.

OUTPUT: The output is the same as for LIZANOVA.3WY except that in the table of means, these programs print the number of observations on which each mean is based.

CUSTOM DATA FILE: The data file is the same as for LIZANOVA.3WY except that the data for each cell must be preceded by the cell frequency. This is done automatically by the file maker included with the program. When you have entered all the data for a given cell, type

999999<ENTER>

The program will insert the cell frequency in the proper record.

15. MULTIPLE REGRESSION ANALYSIS: MODELING
Programs LIZANAL.NAM and LIZANAL.MUL, Diskette 2

These two programs are similar except in LIZANAL.NAM variables are referred to by name, and in LIZANAL.MUL they are referred to by number. When the variables represent terms of a polynomial in several variables, for example, it might be more convenient not to have to name them; in this case LIZANAL.MUL would be used. For many purposes, it is easier to refer to the variables by name; in these cases, use LIZANAL.NAM.

In both programs the dependent variable is assumed to be the last variable in the data file.

These programs operate in two modes. The FORWARD MODE is obtained by selecting menu item 2 and the BACKWARD MODE is entered from menu item 4.

The FOREWARD MODE is more useful when
(1) stepwise regression is to be performed,
(2) partial correlation coefficients of low order are to be computed, or
(3) modeling by the forward stepwise method is to be done.

The BACKWARD MODE is more useful when
(1) high order partial correlation coefficients are to be computed, or
(2) modeling by the backward elimination procedure is to be used.

In any event, you may pass back and forth between the forward and backward mode options menus so that you may use the options in either. See Chapter VIII for more information on using these programs.

The printer prints a heading for partial correlation coefficients, so it is recommended that all partials be computed before regression analysis tables are printed.

Initial information you must provide is
(1) the number of variables on the data file.

CROSSTERM

In addition to using the file variables, you have the option of defining new variables called cross terms which are products of the file variables.

You are asked for the number of cross terms you wish to use. (If you do not want any, enter 0.)

If you are using LIZANAL.NAM you will be asked to provide
(1) names of file variables,
(2) names of cross terms.
(3) You will then be asked to define the cross terms you wish to use.

A cross term is defined by entering two variable names separated by a comma (or variable numbers separated by a comma if you are using LIZANAL.MUL). Note that once a cross term has been defined, it may be used in the definition of other cross terms.

EXAMPLE: Suppose the file variables are

QUET, AGE, SMK, SBP

You are to create three cross terms named

QTSMK, AGSMK, AQS

which are to represent

QUET*SMK, AGE*SMK, and AGE*QUET*SMK

respectively.

You would define the cross terms as follows:

Crossterm 1:	QUET, SMK
Crossterm 2:	AGE, SMK
Crossterm 3:	AGE, QTSMK

NOTE: Henceforth, the term variable will mean file variable or crossterm.

LOG OPTION

It frequently happens that an investigator wishes to replace the dependent variable Y by its natural logarithm, LN(Y) (or in BASIC, LOG(Y)). You may choose this option by answering "YES" to the LOG OPTION query. If you exercise this option, then all values of the dependent variable must be positive. This option results in a curve fit of the form

$$Y=\exp(a_1x_1+a_2x_2+...+a_nx_n).$$

The last file variable is assumed to be the dependent variable.

The next query is

IS THIS RUN FROM A CUSTOM FILE?

If the data file is a custom regression file, if you are using all the variables in the file, and if a data file COVAR.MAT has not been created by a previous run on LIZCORR.MUL, LIZCORR.XTM, LIZANAL.MUL, or LIZANAL.NAM (see next query below), you must answer YES. If you answer NO, the next query will be

INPUT COVARIANCE MATRIX?

You may answer this query by YES only if a data file COVAR.MAT has been created by a previous run of LIZCORR.MUL, LIZCORR.XTM, LIZANAL.MUL, or LIZANAL.NAM with the SAME SET OF VARIABLES AND CROSSTERMS. Otherwise, answer NO and you must then provide

(1) the total number of columns in the AGGREGATE file, and
(2) a data file column number for each variable.

If you have answered the query INPUT COVARIANCE MATRIX? with NO, then a covarance matrix will be created with the current set of variables and crossterms and will be output to diskette as as data file COVAR.MAT. This data file may be used for subsequent runs (with the SAME SET OF VARIABLES AND CROSSTERMS) with programs LIZANAL.NAM, LIZANAL,MUL, or LIZANAL.RES. This feature may save a considerable amount of time; the computation of the covariance matrix is a relatively slow process.

FORWARD MODE: DESCRIPTION

If you choose the forward mode, then after input and initialization, you will be asked to enter the name (or number) of two variables. THE SECOND VARIABLE ENTERED WILL BECOME THE DEPENDENT VARIABLE.

The VARIABLES AVAILABLE list and the menu options will appear. In either the forward or backward mode, the last variable on the VARIABLES AVAILABLE list is assumed to be the dependent variable.

You may now exercise the options.

PRINT CORRELATION COEFFICIENT

Compute and prints a partial correlation coefficient. You must specify two correlation variables from the VARIABLES AVAILABLE list. All other available variables become the control variables. Thus the order of the correlation coefficient is one less than the variables available.

To return to the OPTIONS menu, press ENTER when you are asked for a correlation variable.

ADD VARIABLE

Any variable you specify will be added to the VARIABLES AVAILABLE list. After specifying a variable to be deleted, you will be asked if you wish to

DELETE ANOTHER VARIABLE?

If you answer YES, you may specify another variable to be deleted. You may continue deleting variables until the above query is answered by NO. New parameters for the model are then computed.

DELETE VARIABLE

Any variable on the VARIABLES AVAILABLE list may be deleted. You will then be asked if you wish to

ADD ANOTHER VARIABLE?

If you answer YES, you may specify another variable to be added. This process is repeated until you answer the above query with NO. New parameters for the model are then computed.

DELETE ALL VARIABLES

You will be asked to enter the names (or numbers) of two variables. This option returns you to the initial conditions which existed when you first entered the FORWARD MODE.

PARTIAL F'S

The partial F's of all variables not on the VARIABLES AVAILABLE list are sent to the screen. The indicated Degrees of Freedom are for the denominator. Numerator DF is always 1 for partial F's.

TO BACKWARD MODE

Control is transferred to the BACKWARD MODE OPTIONS menu. All parameters are unchanged.

BACKWARD MODE: DESCRIPTION

In this mode, all file variables and crossterms will appear on the VARIABLES AVAILABLE list when the OPTIONS menu appears.

OPTIONS 1 AND 2

These perform the same function as in the FORWARD MODE.

RESTORE ALL VARIABLES

All file variables and cross terms are placed on the VARIABLES AVAILABLE list. (New parameters are computed.)

SET VARIABLE POINTERS

This option allows you to specify the number and order of the variables on the VARIABLES AVAILABLE list. In particular, you may change the dependent variable.

TO FORWARD MODE

Transfers control to the FORWARD MODE OPTIONS menu. All parameters are unchanged.

REGRESSION RUN

A regression analysis is performed for all variables which appear on the VARIABLES AVAILABLE list. OUTPUT: An ANOVA table for regression analysis including the contribution to the sum of squares and the partial F for each variable. The partial F for each variable is computed as if the variable were the last to be entered into the model. The program also prints the regression coefficient, standard error, degrees of freedom, and Student's T for each variable. The constant term, labeled as "Y-INTERCEPT", and the standard error of estimate for the regression are printed.

PARTIAL F'S

YOU WILL NEED AN F-TABLE WHEN USING THIS OPTION FOR MODELING.

The partial F's of all variables on the VARIABLES AVAILABLE list are sent to the screen. Indicated DF's are for the denominator. Numerators' DF is 1.

See Chapter VIII, "Performing Regression Analysis with LIZPACK" for instructions for using these programs for modeling and stepwise regression analysis.

CUSTOM DATA FILE: Suppose that the total number of variables (including the dependent variable) is K. Then a custom file is an AGGREGATE file with K columns, where of course each variable occupies one column. THE DEPENDENT VARIABLE MUST BE IN THE LAST COLUMN.

16. POLYNOMIAL AND EXPONENTIAL REGRESSION
Program LIZANAL.PLY, Diskette 2
Program LIZEXPO.ANL, Diskette 2

Program LIZANAL.PLY fits a curve of the form y = p(x) where p(x) is a polynomial of degree specified by the user.

Program LIZEXPO.ANL fits a curve of the form y = exp (p(x)), where p(x) is a polynomial, by means of the linearization Ln y = p(x). Thus, all values of the dependent variable must be positive.

These programs are quite similar; they differ mostly in the graphics section. While LIZEXPO.ANL fits the curve Ln y = p(x), it plots the curve y = exp (p(x)) against the actual observed values of y. The operation of both the programs is the same.

Upon entering the program from the menu option POLYNOMIAL REGRESSION, you must provide the degree of polynomial with which you wish to begin. Note that this degree can be reduced later, so you should specify the highest degree of a polynomial in which you might be interested. Amount of computing time required increases rapidly with the degree of the polynomial.

If you answer the query

IS THIS RUN FROM A CUSTOM FILE?

with NO, then you must provide

(1) the total number of columns in the data file,
(2) the file column number for the independent variable, and
(3) the file column number for the dependent variable.

After initialization, the TERMS AVAILABLE list and the OPTIONS menu will appear. You may now exercise the options.

OPTIONS

PARTIAL F'S (OPTION F)

THIS OPTION MUST BE EXERCISED IMMEDIATELY AFTER THE OPTION MENU APPEARS if it is to be used at all. If other options, with the exception of graphing, are used first, this option may yield incorrect results.

This option helps to select the proper polynomial model for your data. It lists the degrees of freedom and partial F ratio for a least square regression for polynomials of each degree up to the highest degree available. The proper polynomial model is the highest degree polynomial with a significant partial F. If all available polynomials have significant partial F's, you may not have started with a polynomial of sufficiently high degree.

GRAPHS

Graphs may be drawn in two coordinate systems:

STANDARD SYSTEM: the origin of the coordinate system is at the mean of the independent variable and mean of the dependent variable.

NONSTANDARD SYSTEM: the data is shown in its natural or "raw" form.

Upon entry into the program the system is set for drawing graphs on the STANDARD coordinate system. You may draw a graph of the existing curve fit by exercising OPTION 2, "DRAW GRAPH, STANDARD". After drawing a graph, press any alphanumeric key to return to the OPTIONS menu. To switch systems, select OPTION 7, "CLEAR GRAPH SCREEN, N.S." The new coordinate system and the scattergram for your data will appear. You may return to the OPTIONS menu after exercising this or any other graphing option by pressing any alphanumeric key. To return to the STANDARD system, you must select OPTION 3.

If terms are deleted, you may draw graphs of the new curve fit on the same screen with previous graphs. Old graphs may be removed from the screen by selecting OPTION 3 or OPTION 7.

At any time, selecting a DRAW GRAPH option will result in the graph of the curve fit for the current TERMS AVAILABLE list.

You may look at the graph screen at any time by selection of OPTION 1, "INSPECT SCREEN."

DELETE TERM (OPTION 5)

Any term (power of the independent variable) may be deleted. A new curve fit is performed. A new TERMS AVAILABLE list appears when control returns to the OPTIONS menu.

This option allows you to

(1) reduce the degree of the polynomial, or

(2) perform curve fits for special types of polynomials, such as even or odd polynomials.

PRINT CORRELATION COEFFICIENT (OPTION 1)

Upon entering the program, the printer sets up to print total correlation coefficients for polynomial fits. Thus, you should use the option, if it is desired at all, before doing a REGRESSION RUN.

OUTPUT:

Printer: Terms available, total correlation coefficient for existing curve fit, degrees of freedom for F-ratios, F-ratio for significance of curve fit.

Screen: Total correlation coefficient.

Control is returned to the OPTIONS menu by pressing any alphanumeric key.

RESTORE ALL TERMS

The curve fit for the original polynimial is restored to the system.

REGRESSION RUN (OPTION 9)

Regression analysis for the curve fit of the current TERMS AVAILABLE list is sent to the printer.

OUTPUT: ANOVA table for regression; regression coefficient, standard error, degrees of freedom for Student's T, and Student's T for significance for each term; Y-intercept (constant term); means and standard deviation for each term; standard error of estimate; squared correlation coefficient.

SCREEN DUMP (OPTION 0)

The current contents of the graphics screen is sent to the printer. To exercise this option you must

(1) have a compatible printer with graphics capability, and

(2) you must have answered "Y" to the query "IS SCATTERGRAM OPTION DESIRED?" before beginning execution of the program.

CUSTOM DATA FILE: Let X_i and Y_i denote the i^{th} observation of the independent and dependent variable, respectively. The arrangement of the data is

RECORD	1	2	3	4	5	6	. . .
DATA	X_1	Y_1	X_2	Y_2	X_3	Y_3	. . .

The order of the observation is not important as long as

(1) corresponding observations of the independent and dependent variable are adjacent, and

(2) the independent variable appears first.

17. FREQUENTLY USED REGRESSION METHODS
Programs LIZRAD.PWR, LIZEXPO.ONE, and LIZLOG.ANL, Diskette 2

The operation of all these programs is the same. These programs perform the following special frequently used least-squares regressions:

Program	Curve Fit
LIZRAD.PWR	$y = ax^b$
LIZEXPO.ONE	$y = ab^x$
LIZLOG.ANL	$y = a + b\text{Ln}(x)$

All programs do linearized regressions, but the graphs are of the actual curve fits.

After selecting the regression option from the MAIN MENU and answering the standard header queries, you must provide information concerning the data file.

If you answer the query IS THIS RUN FROM A CUSTOM FILE? with NO, then you must provide

(1) the total number of columns in the data file,
(2) the file column number for the independent variable, and
(3) the file column number for the dependent variable.

OUTPUT: ANOVA table for linearized regression. The coefficients A and B; mean and standardized deviation of each variable; squared correlation coefficients.

After doing a regression analysis you may exercise the RESIDUALS option from the main menu. You must specify

(1) a name for the residual file, and
(2) if you answer YES to the query ARE RESIDUALS TO BE COMPUTED BY SUBTRACTION? then residuals are computed by subtracting computed values from observed values. Otherwise, residuals are computed by division and no computed value of the independent variable may be zero. (Residuals are normally computed by subtraction. In certain models, such as "forecasting" models, the residuals are assumed to appear multiplicatively.)

OUTPUT: table containing

1. values of the independent variable
2. observed value of Y (dependent variable)
3. computed value of Y
4. residual for each observation.

DISKETTE file containing values of independent variable in first column; residuals in second column.

The diskette file is in the standard form of a two variable LIZPACK regression file (or a TIME SERIES file in the forecasting package).

CUSTOM DATA FILE: Same as LIZANAL.PLY

18. POLYNOMIAL REGRESSION: UTILITY PROGRAMS
Programs LIZPOLY.ONE and LIZPOLY.STP, Diskette 2

LIZPOLY.ONE does a single regression analysis for the polynomial of the degree you specify and produces a scattergram and graph.

OUTPUT: Same format as from REGRESSION RUN from LIZANAL.PLY.

LIZPOLY.STP does a regression analysis, accompanying ANOVA table, and regression coefficients with data for significance tests for the polynomial of the degree you specify AND ALL POLYNOMIALS OF A LESSER DEGREE. Otherwise, OUTPUT is the same format as for a REGRESSION RUN from LIZANAL.PLY. Scattergram and multiple graphs are produced.

After selecting the regression option from the MAIN MENU and answering the standard header queries, you must provide information concerning the data file.

If you answer the query IS THIS RUN FROM A CUSTOM FILE? with NO, then you must provide
(1) the total number of columns in the data file,
(2) the file column number for the independent variable, and
(3) the file column number for the dependent variable.

19. CORRELATION MATRICES: UTILITY MULTIPLE REGRESSION PROGRAMS
Programs LIZCORR.MUL and LIZCORR.XTM, Diskette 2.

Both programs will compute and print correlation matrices. LIZCORR.XTM has the cross term and log options (see LIZEXPO.ANL) while LIZCORR.MUL is a straight multiple regression program. Both programs will do a regression analysis with a standard ANOVA table. Variables are referred to by number. You must provide
(1) the number of file variables, and
(2) for LIZCORR.XTM, you must follow the same directions as for LIZANAL.MUL for defining cross terms.

If you answer the query IS THIS RUN FROM A CUSTOM FILE? with NO, then you must provide
(1) the total number of columns in the aggregate file, and
(2) the file column number for each variable.

CAUTION: You must exercise the option BEGIN RUN - COMPUTE CORR. MAT. on the MAIN MENU before printing correlation coefficients or doing a regression analysis. You must exercise the option "CORRELATION COEFFICIENTS" before doing a MULTIPLE REGRESSION ANALYSIS if you want a print-out of the correlation matrix. A regression run alters the correlation matrix.

OUTPUT: Same format as from a REGRESSION RUN from LIZANAL.MUL.

CUSTOM DATA FILE: Same as LIZANAL.MUL.

20. RESIDUAL ANALYSIS - MULTIPLE REGRESSION PROGRAM LIZANAL.RES, DISKETTE 1

This program performs a residual analysis from a multiple regression analysis. It is intended to be used after the proper regression model has been determined by means of one of the modeling programs LIZANAL.MUL or LIZANAL.NAM, and is considered the final step in a multiple regression analysis. The statistics related to the regression analysis which are printed by this program are more comprehensive than those printed by the other multiple regression analysis programs. Like LIZANAL.NAM or LIZANAL.MUL, this program may be entered from the main menu by the FORWARD MODE or the BACKWARD MODE. All variables are named.

Follow the instructions for LIZANAL.NAM for the initial operation of this program. The operation of this program and LIZANAL.NAM are exactly the same up to the point where the queries concerning the data file begin. This program, like LIZANAL.NAM, has the provision for using a previously prepared covariance matrix; however, the present program must also read the raw data file. If the present run is from a CUSTOM file, answer the query

IS THIS RUN FROM A CUSTOM FILE?

with YES; otherwise answer no. If a covarance matrix has previously been created, answer

INPUT COVARIANCE MATRIX?

with YES, else answer NO. If you answered the CUSTOM file query with NO, you must now provide the same information for the AGGREGATE file as for LIZANAL.NAM. Notice that information concerning the raw data file must be entered whether or not you input the covariance matrix from file COVAR.MAT.

The descriptions of the options follow.

OPTIONS--BACKWARD MODE

OPTION 1: STRAIGHT LINE PLOT

Does a straight line least square fit for any two variables on the VARIABLES AVAILABLE list and plots the line along with the scattergram of the observations. You must specify the names of the independent and dependent variables by answering the LIZPACK queries. The computer will pause after drawing the graph. Press key C to return control to the OPTIONS menu.

OPTION 2: DELETE VARIABLE

Allows you to delete any variable on the VARIABLES AVAILABLE list.

OPTION 3: RESTORE ALL VARIABLES

Returns all variables to the VARIABLES AVAILABLE list.

OPTION 4: STRAIGNT LINE PREDICTIONS

CAUTION: THIS OPTION MUST BE EXERCISED ONLY AFTER EXERCISING OPTION 8: REGRESSION RUN, WITH EXACTLY TWO VARIABLES ON THE VARIABLES AVAILABLE LIST.

In this option, "predicted" values of the dependent variable are computed for specified values of the independent variable. You must enter the values of the independent variable (one at a time) for each predicted value of the dependent variable you wish to compute. When you have finished computing predicted values, answer the query

VALUE OF INDEPENDENT VAR

with

999999

and control will be returned to the OPTIONS menu. Upon entering this option you will be asked to enter the value of Student's T for probability level at which you are working. The number of degrees of freedom will be provided. This value of T is used in computing confidence intervals for the predicted Y's, and the values of T may be found in a T-table. (If you enter 0, confidence intervals <0,0> are printed.)

OUTPUT: values of independent variable, predicted Y's, and confidence intervals.

OPTION 5: SET VARIABLE POINTERS

See instructions for this option under LIZANAL.NAM (Section 15 of this chapter).

OPTION 6: RESIDUAL ANALYSIS

CAUTION: THIS OPTION MUST BE EXERCISED ONLY AFTER EXERCISING OPTION 8: REGRESSION RUN.

Does a residual analysis for the least squares curve fit of the variables on the VARIABLES AVAILABLE list.

OUTPUT:

PRINTER: Table containing the

1. Case (observation number),
2. Observed Y (Y being the dependent variable),
3. Computed Y,
4. Residual,
5. Normal deviate,
6. Studentized residual (Weisberg, p. 105),
7. V(I,I) (Weisberg, p. 104),
8. Cook's distance (Weisberg, p. 108), and
9. T (for outlier), (Weisberg, p. 115)

for each observation.

See Weisberg, pp. 101-117 for details concerning the nature and use of the Statistics 6 - 9. These statistics are used to single out cases which may be influential or may be outliers. The statistic V(I,I) is related to the Mahalanobis D; see Weisberg, p. 105.

The following statistics are also printed:

t(2,1),t(1,2) statistics (see Draper and Smith, Applied Regression Analysis, Wiley, 1981, page 150).

Darbin-Watson Statistic

Sum of squares of residuals from direct computation.

DISKETTE: File containing
1. Computed Y's (first column), and
2. Residuals (second column).
This file is named "RESID.DAT".

OPTION 7: TO FORWARD MODE

Control is transferred to the FORWARD MODE options menu.

OPTION 8: REGRESSION RUN

Does a regression analysis for all variables on the VARIABLES AVAILABLE list.

OUTPUT:
Standard ANOVA table for regression analysis
Table of partial regression coefficients which contains for each variable:
1. variable number
2. regression coefficient
3. standard error
4. mean of the variable
5. Student's T for significance test. (The square of this values is the PARTIAL F for the variable.)

Other statistics:
DF for t-test on regression coefficients
Standard error of estimate
Mean of dependent variable
Standard error of estimate as percent of mean of dependent variable.
Percent of variance explained by regression.
Number of observations.

OPTIONS: FORWARD MODE

CAUTION: Options 1, 7, and 8 must be exercised ONLY after exercising OPTION 6: RESIDUAL ANALYSIS in the backward mode.

NOTE ON THE SHOWPACK OPTION: Although the SHOWPACK option does not appear on the OPTION menu, any graphics page may be saved to diskette by pressing key S while the FORWARD MODE menu is on the screen. Then follow the instructions for SHOWPACK.

Also see the instructions for OPTION 8 for saving residual plots to diskette.

OPTION 1: PLOT RESIDUALS VS. PREDICTED Y

A plot is made of the residual vs. predicted Y for each observation. The vertical coordinate is the residual. SEE CAUTION ABOVE. After plotting is finished, press key C to return to the OPTIONS menu.

OPTION 2: ADD VARIABLE

The variable you specify is added to the VARIABLES AVAILABLE list.

OPTION 3: DELETE VARIABLE

The variable you specify is deleted from the VARIABLES AVAILABLE list. (This option cannot be exercised when only two variables are on the list.)

OPTION 4: DELETE ALL VARIABLES

All variables on the VARIABLES AVAILABLE list are deleted.

Control is sent to the beginning of the forward mode; you will be asked to specify two variables to be entered on the variable list.

OPTION 5: END RUN

Control is transferred to the main menu.

OPTION 6: TO BACKWARD MODE

Control is transferred to the BACKWARD MODE options menu.

OPTION 7: LAG-1 SERIAL PLOT

Let E_i be the residual computed for observation i, i=1, . . . , N. Then E_i is plotted vs. E_{i-1}, i=2, . . . , N. This is a check for simple serial correlation among the residuals. After plotting is finished, press key C to return to the OPTIONS menu.

OPTION 8: PLOT RESIDUALS VS. VARIABLES

Plots are made of the residuals vs. each variable on the VARIABLES AVAILABLE list. After plotting is finished, pressing the following keys will result in the indicated action:

C - The plot of the residuals vs. the next available variable is made if any variables remain. Otherwise control is returned to the OPTIONS menu.

S - The SHOWPACK option is exercised. See instructions for SHOWPACK.

X - Control is returned to the OPTIONS menu.

CUSTOM DATA FILES: same format as for LIZANAL.MUL or LIZANAL.NAM.

PRACTICE DATA FILES: same as for LIZANAL.NAM.

21. RESIDUALS FROM POLYNOMIAL AND EXPONENTIAL REGRESSION PROGRAM LIZPOLY.RES, Diskette 1

This program computes the residuals for a polynomial regression from programs LIZANAL.PLY, LIZPOLY,STP, or LIZPOLY.ONE, or from an exponential regression from program LIZEXPO.ANL. Residuals may be computed by subtraction or by division. Residuals are output to a diskette file which may be used as input to any one of the LIZPACK time series analysis or forecasting programs (assuming the independent variable is time and/or is equally spaced). Residual plots may be obtained by using program LIZPLOT.SRS in the forecasting package. If the independent variable is time, other analyses may be made with LIZ1TIME.SRS and LIZ2TIME.SRS.

You must provide:

(1) the name of the source data file, i.e., the data file used for the regression,

(2) name for the residual file,

(3) degree of the polynomial,

(4) Answer <YES> or <NO> to the query, "IS THIS AN EXPONENTIAL REGRESSION?", and

(5) you will then enter the coefficients of the polynomials as fitted by the regression program. These coefficients are printed with the output from the regression run. The constant term is entered first (it is printed out as the Y-INTERCEPT by the regression programs). Then enter the coefficients of each term as you are asked to do so.

OUTPUT:

Printed table containing values of the independent variable, observed values of dependent variable Y, and

Computed Y

Residuals

Diskette file in the standard form of a 2-variable regression analysis file; the independent variable is the first variable and the residuals are the second variable.

22. COMMON VARIABLE TRANSFORMATIONS FOR REGRESSION ANALYSIS AND AGGREGATE FILES

Program TRANSPAK.REG, Diskette 1

This program performs some of the common transformations used in regression analysis. It accepts CUSTOM regression or AGGREGATE files as input. A CUSTOM regression analysis file containing K variables is an AGGREGATE file containing K columns, with variable 1 in column 1, variable 2 in column 2, etc.

This program is very useful in many situations where standard one-variable transformations are to be performed on an AGGREGATE file or CUSTOM regression file. It is not as versatile as the variable transformation programs discussed in Chapter 7, but it is extremely easy to use in situations where it is applicable.

After selecting TRANSFORMATION OF VARIABLES on the MAIN MENU, you must provide

(1) name of source file,

(2) total number of columns (in source file; for regression file this is the total number of variables),

(3) name of destination file which will contain the transformed variables. Your source file will be unaltered.

TRANSPAK.REG will now make a copy of the source file into the destination file. All variable transformations will now be made on the variables in the destination file.

The OPTIONS menu will appear. You may select any transformation (see list below). You will then be asked to enter an appropriate constant (if applicable) and

(4) the column number of the variable being transformed.

The appropriate transformation will be made on all observations of the variable specified and control is returned to the options menu where you may select another transformation or EXIT to the MAIN MENU.

You may make as many transformations as desired on as many variables as desired. You may inspect the transformed file with the MINI-EDITOR, which may be entered from the MAIN MENU.

It is your responsibility to examine your data and make sure that the designated transformation may be performed. If you attempt to perform an impossible transformation, an error message (and termination of program execution) will result.

The following transformations are available: (X is an observation of the variable being transformed)

TRANSFORMATION	NOTES, LIMITATIONS
SIN (X)	X in radians
COS (X)	X in radians
TAN (X)	X in radians
ARCSIN (X)	-1<X<1
ARCCOS (X)	-1<X<1
ATN (X)	
LN (X)	Natural log of X, X>0
EXP (X)	-84<x<84
LOG (X)	Base 10 logarithm, X>0
10^X (10^X)	-37<X<37
EXPONENTIATE	X^a; you must enter a
SQRT(X)	Square root of X; X>=0
ADDITION OF CONSTANT	The same constant is added to each observation of the variable. You must enter the constant, which may be positive or negative.

ABS (X)	Each observation is replaced by its absolute value.
MULTIPLY BY A CONSTANT	You must enter the constant when asked to do so.
INT (X)	Truncate X to the nearest integer without rounding.
RECIPROCAL	1/X; X<>0
ROUND	Round X to the nearest integer
LN((1/X)-1)	0<X<1: This is usually used as a transformation of the dependent variable when fitting a curve of the form $Y=1/(1+EXP(a_1X_1+...+a_nX_n)$
(Y^L-1)/L,L<>0 LN(y),L=0	See Draper and Smith, Applied Regression Analysis, Wiley, 1981, p. 225, for a discussion of this family of transformations. You will be asked to enter L as "LAMBDA".

23. ANALYSIS OF VARIANCE OR COVARIANCE: REGRESSION MODEL

Program LIZ1CORR.MCP, Diskette 6
Program LIZ2CORR.MCP, Diskette 6
Program LIZ3CORR.MCP, Diskette 6
Program LIZ4CORR.MCP, Diskette 6

These programs will perform a wide variety of one-, two-, or three-way ANOVAs or ANCOVAs (multiple covariates), using appropriate regression models. They will, in fact, perform any type ANOVA or ANCOVA which is provided for by the individual programs described earlier in this chapter. However, when one of the individual programs is applicable, it should normally be used because

1. The regression model ANOVA or ANCOVA requires more memory and hence can process much fewer groups or cells than one of the specialized programs.
2. The regression model may be much more time consuming than one of the individual programs.

A regression model must be used in any of the following situations:

1. For a two- or three-way ANOVA when there are empty cells or when the method of unweighted means is not applicable (see Kleinbaum and Kupper, Applied Regression Analysis, p. 358).
2. For a one- or two-way randomized block experiment when observations are missing.
3. For a two- or three-way analysis of covariance with unequal cell frequencies.
4. For any type of analysis of covariance with multiple covariates.
5. For performing a three-way randomized block analysis.
6. For performing an analysis with a latin square design or cross-over design.

Since the regression model may require a lot of memory for matrix storage space, the program LIZ1CORR.MCP prepares the design matrix (i.e., for each case it constructs a vector consisting of the covariates (if any), the dummy variables, and the dependent variable), and then computes the covariance matrix based on the design matrix. Program LIZ2CORR.MCP or LIZ3CORR.MCP performs the actual regression with a hierarchical or symmetrical model. Program LIZ4CORR.MCP allows you to modify the covariance matrix.

Program LIZ1CORR.MCP requires the use of one to three independent variables corresponding to the main effects, which are called Factor A, Factor B, and Factor C. Only Factor A is used for a one-way analysis; Factors A and B are used for a two-way analysis, and all three factors are used in a three-way analysis. The user must specify the number of levels for each applicable factor. In the following discussion, let us suppose that

(1) the number of levels for Factor A is a, where $a > 1$;
(2) the number of levels for Factor B is b, where $b = 0$ or $b > 1$;
(3) the number of levels for Factor C is c, where $c = 0$ or $c > 1$; and
(4) the number of covariates is d, where $d >= 0$.

The program LIZ1CORR.MCP then constructs dummy variables for each factor as follows:

(1) For level A it constructs $a - 1$ dummy variables X(1), X(2), ... X(a-1) where

$X(I) = -1$ for level a,
$X(I) = 1$ for level I, $I<a$, and
$X(I) = 0$ otherwise.

(2) If $b > 1$ it constructs $b - 1$ dummy variables for Factor B, and
(3) If $c > 1$ it constructs $c - 1$ dummy variables for Factor C.

The dummy variables of (2) and (3) are similar to those of (1).

The two-way interactions between Factors A and B are referred to as the AB interactions. Then interactions AC, BC, and the three-way interaction ABC have obvious meanings.

If interaction AB is to be used in a two-way ANOVA, then LIZ1CORR.MCP will construct (a-1)(b-1) variables consisting of all products of the dummy variables for Factor A and Factor B taken two at a time. If interactions AC and/or BC are used in a three-way ANOVA, the corresponding procedure will be followed for each of these interactions. If the three-way interaction ABC is used in a three-way ANOVA, then LIZ1CORR.MCP will construct (a-1)(b-1)(c-1) variables consisting of all products of the dummy variables for Factors A, B, and C taken three at a time.

Program LIZ1CORR.MCP provides the user with the option to eliminate any two-way or three-way interaction from the analysis prior to constructing the covariance matrix. After constructing a covariance matrix with interactions, the interactions may be removed by LIZ4CORR.MCP, if desired.

For a one-way analysis with d covariates, the total number of variables required is $a + d$ (this includes the dependent variable). For a pure analysis of variance, we have $d = 0$.

For a two-way analysis with all interactions, the number of variables required is ab + d. You may eliminate (a-1)(b-1) interaction terms.

For a three-way analysis with all interactions, the number of variables required is abc + d. You may eliminate up to (a-1)(b-1) + (a-1)(c-1) + (b-1)(c-1) + (a-1)(b-1)(c-1) interaction terms.

There is no programmed limit to the number of covariates or levels for the factors. However, the amount of memory available will naturally limit the number of variables which may be used. The following are the estimated limits for the number of variables which may be used in an analysis:

COLOR COMPUTER and 64K IBM-PC: 64 variables
128K IBM-PC: 80+ variables.

For program LIZ1CORR.MCP, you must provide the following information:

(1) NUMBER OF COVARIATES?
For an analysis of variance only, enter 0. Otherwise, enter the total number of covariates.

(2) NUMBER OF LEVELS FOR FACTOR A?
This must always be a positive number, a.

(3) NUMBER OF LEVELS FOR FACTOR B?
For a one-way analysis, enter 0. If you enter a positive number, b, then you will be asked for

(4) NUMBER OF LEVELS FOR FACTOR C?
For a two-way analysis, enter 0. Otherwise, enter the appropriate positive number, c.
If you entered b > 0 at step (3), you will be asked if you wish to

(5) ELIMINATE AB INTERACTIONS?
If you answer YES, the crossterms for interaction AB will not be placed in the regression model.
If you entered c > 0 at step (4), you will be asked if you wish to

(6) ELIMINATE AC INTERACTIONS?
(7) ELIMINATE BC INTERACTIONS?
(8) ELIMINATE ABC INTERACTIONS?
If you answer YES to any one of these questions, the crossterms for the corresponding interaction will not be placed in the model. (See the discussion below concerning randomized block experiments.)

You must now provide
(9) NAME OF DATA FILE?
(10) TOTAL NUMBER OF COLUMNS IN DATA FILE?
If you are using covariates, you will be asked to provide a column number for each covariate.

You must now provide
(11) COLUMN NUMBER FOR IND. VAR., FACTOR A?
(12) LOW VALUE?
(13) HIGH VALUE?

If you entered b > 0 at step (3), you will be asked to enter
(14) COLUMN NUMBER FOR IND. VAR., FACTOR B?
(15) LOW VALUE?
(16) HIGH VALUE?

If you entered c > 0 at step (4), you will be asked to enter
(17) COLUMN NUMBER FOR IND. VAR., FACTOR C?
(18) LOW VALUE?
(19) HIGH VALUE?
(20) COLUMN NUMBER FOR DEP. VAR.?

The values of the independent variable must be integers between -999999 and 979796. They serve to identify the levels or groups to which each case belongs. The low value corresponds to the lowest level of the factor and the high value corresponds to the highest level. HIGH VALUE must be greater than LOW VALUE (LIZPACK does not check this). There must be no gaps between LOW VALUE and HIGH VALUE.

EXAMPLE: For a one-way analysis, suppose the column number for the independent variable, Factor A, is 5; suppose the low value is -3 and the high value is 4. The analysis will then be based on 8 groups (levels). Any case which has the value -3 in column 5 will be placed in group 1; a case which has value -2 in column 5 will be placed in group 2, etc.

EXAMPLE: For a two-way analysis, suppose
(a) the column number for the independent variable, Factor A, is 2, LOW VALUE is 0, and HIGH VALUE is 3 (the number of levels for Factor A is 4).
(b) the column number for the independent variable, Factor B, is 3, LOW VALUE is 2, and HIGH VALUE is 3 (the number of levels for Factor B is 2).
Then a case with the indicated values in columns 2 and 3 are assigned cells as follows:

VALUE IN COL. 2	VALUE IN COL. 3	LEVELS (OR CELLS)
0	2	1,1
0	3	1,2
1	2	2,1
1	3	2,2
2	2	3,1
2	3	3,2
3	2	4,1
3	3	4,2

For a three-way analysis, the LEVELS assignment procedure is similar.

As we observed previously, there must be no gaps between LOW VALUE and HIGH VALUE. This means for example, that if LOW VALUE is 0 and HIGH VALUE is 3, then there must be at least one case in the data file for which the value of the independent variable is 0, one for which it is 1, one for which it is 2, and one for which it is 3. (This does <u>not</u> preclude empty or non-equal cells.)

If a case has a value of the independent variable which does not lie between LOW VALUE and HIGH VALUE, then the case will be ignored.

MISSING DATA: If the missing data symbol 999999 is encountered in any column containing a covariate, independent variable, or the dependent variable, then the case will be ignored.

RANDOMIZED BLOCK EXPERIMENTS: You may perform a one- or two-way randomized block experiment with these programs. Observations may be missing, i.e., there may be empty cells.

A one-way randomized block experiment is performed as a two-way analysis with at most one observation per cell. YOU MUST ELIMINATE THE TWO-WAY INTERACTION (AB INTERACTION).

A two-way randomized block experiment is performed as a three-way analysis with at most one observation per cell. YOU MUST ELIMINATE THE THREE-WAY INTERACTION (ABC INTERACTION). <u>It is also customary to eliminate the two-way interactions involving the blocks variable</u>, but this does not have to be done. The blocks variable may correspond to either Factor A, B, or C (user's option).

You may also perform a three-way randomized block experiment with b blocks by providing b-1 dummy variables X(1), ... X(b-1) as covariates. Define X(I) as follows:

X(I) = -1 if the case belongs to block b,
X(I) = 1 if the case belongs to block I, and
X(I) = 0 otherwise.

Introduce all other covariates into the model before introducing the dummy variables as covariates.

LATIN SQUARE AND CROSS OVER DESIGNS: You may perform a latin square analysis as a three-way ANOVA (with or without covariates). Use one independent variable as the row variable, and one as the treatment variable. All three variables must have the same number of levels. The levels assign each case to a cell, and there must be at most one observation per cell. YOU MUST ELIMINATE ALL 2-WAY AND 3-WAY INTERACTIONS.

A cross over design is similar to a latin square design except replications occur in the row or column variable. Such a design may be handled in a manner similar to a latin square design, with the number of levels for the row or column variable adjusted for replication. See W. G. Cochran and G. M. Cox, <u>Experimental Design</u>, Wiley, New York, 1957, for more information on these designs.

OUTPUT: The output consists of the diskette data file COVAR.MAT which may be used ONLY for input for the program LIZ2CORR.MCP or LIZ3CORR.MCP. <u>Do not try to use this file for input for any other regression analysis program.</u>

After running LIZ1CORR.MCP, you may load LIZ2CORR.MCP, LIZ3CORR.MCP, or LIZ4CORR.MCP from the main menu of LIZ1CORR.MCP.

For program LIZ2CORR.MCP or LIZ3CORR.MCP you must provide

(1) The name of each covariate (if any),
(2) the name of each independent variable, and
(3) the name of the dependent variable.

There is some flexibility in the number of characters used in naming the variables, but until you gain experience, we suggest that you use at most five characters for the names of covariates and independent variables, and at most 30 characters for the name of the dependent variable.

LIZ3CORR.MCP performs the analysis using a hierarchical approach. The model is referred to as Model 1 or 2 by Kleinbaum and Kupper, p. 363, and Model 3 by Overall and Klett, p. 448.

OUTPUT: The output for LIZ3CORR.MCP is an ANOVA table which we now describe.

The program enters the covariates and dummy variables into the model one at a time. It prints a line for each variable, containing

(1) the degrees of freedom (DF),
(2) the regression sum of squares (SS),
(3) the mean sum of squares (MS),
(4) F-ratio (Partial F for the variable), and
(5) the partial regression coefficient.

The program prints a total for the covariates, each main effect, and each applicable two- or three-way interaction consisting of

(1) the degrees of freedom (DF),
(2) the regression sum of squares (SS),
(3) the mean sum of squres (MS), and
(4) F-ratios. The first F-ratio is the multiple-partial F for the covariates or dummy variables for the effect. The numerator DF is the DF printed on the line, and the denominator DF is

total DF minus numerator DF.

(Total DF is obtained from the TOTAL line at the bottom of the table.) The second F-ratio is printed in parentheses and is the MSS divided by the MS for ERROR (usual ANOVA statistic). The denominator DF is

DF for error minus numerator DF

The first partial F is usually used for the case of unequal cell frequencies, and the second for the case of equal cell frequencies.

The numbers following variable names refer not to the level of the factor, but to the number of the dummy variables.

NOTE: In this program, the covariates are always entered into the model first. In the case of equal cell frequencies, the order of entry of main effects into the model is immaterial. This is not the case when all frequencies are unequal. In all cases, the order of entry of the effects is controlled by the user's specifications of the independent variable for the main effect in LIZ1CORR.MCP. The order of entry may be changed for different runs. More than one run, with different orders of entry of the main effects, may be desirable (see Kleinbaum and Kupper, pp. 362-366, and Overall and Klett, Chapter 18). Alternately, the first analysis may show that certain main effects and/or interactions are insignificant, and it may be desirable to make a new run with these effects eliminated. Since the construction of the design and covariance matrix for a large problem may require an enormous amount of time, the program LIZ4CORR.MCP is provided to allow the modification of the covariance matrix for more than one analysis.

Program LIZ3CORR.MCP uses a symmetric regression model, i.e., the order of entry of the main effects does not affect the analysis. The model used is referred to as Model 3 by Kleinbaum and Kupper, p. 363, and as Model 2 by Overall and Klett, p. 448.

NOTE: There is a third model referred to in Kleinbaum and Kupper or Overall and Klett. This model is no longer recommended as a general regression model for ANOVA and is not implemented in LIZPACK.

OUTPUT: The output for LIZ3CORR.MCP is an ANOVA table which lists the
(1) degrees of freedom,
(2) sum of squares,
(3) mean sum of squares, and
(4) F-ratios

for the covariates, each main effect, each interaction which is present in the model. The ERROR degrees of freedom, sum of squares, and mean sum of squares is given. Finally, the program prints the total degrees of freedom and total sum of squares. The conventions for F-ratios printed in parentheses are the same as for LIZ2CORR.MCP.

The denominator DF for a multiple partial F for a main effect is computed by
(1) adding the DFs for the covariates and all main effects, and
(2) subtracting the result of (1) from the DF for the TOTAL SS.

We recommend that you read Kleinbaum and Kupper, p. 362, last paragraph, or the reference to Applebaum and Cramer which is cited there, before interpreting the partial F's of LIZ3CORR.MCP.

THE OUTPUT FOR LIZ4CORR.MCP IS ALWAYS THE DISKETTE FILE COVAR.MAT, and any run of LIZ4CORR.MCP will result in the loss of any previous file named COVAR.MAT. For this reason, we suggest that immediately after obtaining the OPTIONS menu in LIZ4CORR.MCP you exercise OPTION 8 and RENAME the file COVAR.MAT to another file with name of your choice. On all future runs, the new filename may be used and the original file will not be altered.

To modify the covariance matrix with LIZ4CORR.MCP, select main menu option MODIFY COVARIANCE MATRIX. The prompt NAME OF FILE FOR COVARIANCE MATRIX? will appear. If the file COVAR.MAT has not been renamed by a previous run, press <ENTER>. Otherwise, enter the new file name.

The covariance matrix will load and the OPTIONS menu will appear.

OPTION 1: COVARIATES allows you to reduce the number of covariates in the model, or to change their order of entry into the model. You must specify the number of covariates to be retained and you must give the number (from the old model) for each of the covariates to be retained. You may eliminate all covariates by specifying that 0 covariates are to be retained.

OPTIONS 2, 3, and 4 allow you to change the order of entry of the main effects into the model. When you interchange main effects, appropriate two-way interactions are also interchanged. Thus, if main effects A and B are interchanged, the program automatically interchanges AC and BC effects (if they are present). Main effects are referred to on the menu by number. The numbers refer to the order in which the effects appear in the EFFECTS IN MODEL list at the top of the screen.

OPTION 5: ELIMINATE EFFECTS transfers control to a new options menu. The options are

1. ELIMINATE MAIN EFFECT 1
2. ELIMINATE MAIN EFFECT 2
3. ELIMINATE MAIN EFFECT 3
4. ELIMINATE 2-WAY INTERACTION 1X2
5. ELIMINATE 2-WAY INTERACTION 1X3
6. ELIMINATE 2-WAY INTERACTION 2X3
7. ELIMINATE 3-WAY INTERACTION
8. ESCAPE TO MAIN MENU.

Main effect 1 indicates the first effect which appears on the EFFECTS IN MODEL list at the top of the screen. It could be Factor A, B, or C. Similar conventions apply to other effects. Note that if you eliminate a main effect, any two- or three-way effects derived from that main effect which are present will also be eliminated. OPTION 8 returns control to the previous OPTIONS menu.

OPTION 6: END RUN causes the covariance matrix for the new model to be written to the file COVAR.MAT. Control is returned to the main menu.

OPTION 8: CHANGE COVARIATE TO DEP. VAR.: This option is used only if you wish to create a covariance matrix for more than one dependent variable. To process more than one dependent variable, you may introduce additional dependent variables as covariates on the initial run of LIZ1CORR.MCP. Before making a run with LIZ2CORR.MCP or LIZ3CORR.MCP, you must first use OPTION 7 of LIZ4CORR.MCP to RENAME the file COVAR.MAT, thus saving the original covariance matrix, and then use OPTION 1 to eliminate from the model all covariates which are actually dependent variables. For subsequent runs on LIZ2CORR.MCP or LIZ3CORR.MCP, use LIZ4CORR.MCP, OPTION 8, with the original covariance matrix, to

(1) eliminate covariates which are dependent variables, and
(2) specify which covariate (in the original covariance matrix) is to become the dependent variable. The old dependent variable will be removed from the model and replaced by the new one.

REFERENCES

1. F. J. Anscombe, "Examination of Residuals," Proc. Fourth Berkeley Symp. Math. Statist. Prob., I, (1961), pp. 1-36.

2. W. G. Cochran and G. M. Cox, Experimental Designs, Wiley, New York, 1957.

3. C. Daniel and F. S. Wood, Fitting Equations to Data, Wiley, New York, 1980.

4. N. R. Draper and H. Smith, Applied Regression Analysis, Wiley, New York, 1981.

5. D. G. Kleinbaum and L. L. Kupper, Applied Regression Analysis, Duxbury Press, North Scituate, Mass., 1978.

6. N. Nie, et al., SPSS, McGraw-Hill, New York, 1975.

7. J. E. Overall and C. J. Klett, Applied Mulivariate Analysis, McGraw-Hill, New York, 1972.

8. F. M. Speed and R. R. Hocking, "The Use of the R( )-notation with Unbalanced Data," Amer. Stat., 30(1), pp. 30-33.

9. R. G. D. Steel and J. H. Torrie, Principles and Procedures of Statistics, McGraw-Hill, New York, 1960.

10. S. Weisberg, Applied Linear Regression, Wiley, New York, 1980.

11. G. O. Wesolowsky, Multiple Regression and Analysis of Variance, Wiley, New York, 1976.

CHAPTER VII
VARIABLE TRANSFORMATIONS

LIZPACK incorporates user written BASIC statements into a subroutine for transforming variables. This opens the way for the user to exercise the full computational power of BASIC for variable transformations. Variables may be transformed from either an AGGREGATE FILE or a CUSTOM FILE. For CUSTOM files, it is assumed that the data is arranged in rows and columns, as in the case of an AGGREGATE FILE. The observation of each variable occupies one column. The number of columns in the file is arbitrary. The transformations may alter one or more file variables and/or create new variables.

Variable transformations are carried out by LIZPACK program CODEPACK. CODEPACK is actually two programs, TRANSPAK.LPK and LIZENCO.LPK. In both programs an existing file of variables is used to create a new file which consists of some combination of transformed variables, non-transformed variables, and newly created variables. In both programs, you will be asked to specify

(1) NAME OF SOURCE FILE, and
(2) NAME OF DESTINATION FILE.

TRANSPAK will save all columns from the source file and all newly created variables to the modified file; of course, if some of the variables from the source file are being transformed, then the transformed variables will appear in the modified file. For example, suppose the source file contains 11 columns, and that 3 new variables are being created. The modified file will contain 14 columns. The first 11 columns correspond to the same variables (or case number, ID, etc.) as they did in the source file; the last 3 columns contain the newly created variables. If columns 4, 7, and 10 of the source file contain variables which are to be transformed, then the values of the transformed variables will appear in columns 4, 7, ad 10 of the modified file. Thus, TRANSPAK always creates a modified file which is at least as large as the source file.

LIZENCO is similar to TRANSPAK except you may specify which columns, if any, from the source file are to be saved to the modified file. For example, the source file might contain 11 columns. You may want to save 6 of these columns to the modified file and create 2 new variables with LIZENCO. Then the modified file will contain 8 columns; the first 6 will correspond to variables of the source file (but perhaps transformed by LIZENCO), and the last 2 correspond to the newly created variables.

In planning for variable tranformations, first decide which variables in the file will actually be used in the computations. Of course some variables might be used to create new variables or alter other variables without actually being transformed themselves. Nevertheless, ALL file variables which can have as many transformations in one step as you like.

are used in the computation must be specified. Make a list of the names of the variables to be used and the file column in which they appear. Such a list might be

VARIABLE NAME	FILE COLUMN
WT	4
HGT	6
AGE	7
SEX	3
IQ	8

Notice that you have given a name to each file variable. This name will be used by CODEPACK.

Now decide how many new variables are to be created (of course, there might be none). Give these variables a name and decide the order in which you want them to appear in the modified file.

You must follow these conventions in naming file variables:

* (1) File variable names must consist of at least two letters or a letter and a number. The first character must always be a letter. REMEMBER that BASIC only looks at the first TWO characters of a variable name, so your variables must have a unique name when only the first two characters are considered. You may use additional characters for convenience.

* (2) Variable names may not begin with X, Y, or Z.

Now write the BASIC coding to transform existing file variables and/or create new variables. Give your statements numbers between 10000 and 19990. Line 10000 MUST contain a statement or an UNDEFINED LINE ERROR will result. Your coding is subject to all the rules of BASIC and will not be edited by LIZPACK. Your coding will become a subroutine and will be directly executed by BASIC. YOU MUST NOT USE VARIABLES BEGINNING WITH X, Y, OR Z in your coding. Variables beginning with X, Y, and Z are reserved for LIZPACK use.

NOTE: If your file contains missing data symbols, you must either ask CODEPACK to DELETE CASES, MISSING DATA, or test for missing data. If you perform a transformation on the missing data symbol 999999, it will appear in the new file modified to non-missing data.

You will be provided with data from the file one case at a time. This is the data available for variable transformation.

Next, write the DUMMY STATEMENTS by simply setting source file names and newly created variable names equal to zero with statement numbers from 20000-21000. Line 20000 MUST contain a dummy statement, or an UNDEFINED LINE ERROR will result. For example, if the variables from the source file have been named WT, HGT, AGE, SEX, and IQ, your DUMMY STATEMENTS might be

```
20000    WT=0
20005    HGT=0
20010    AGE=0
20015    SEX=0
20020    IQ=0
```

If you are creating a new variable named QUET, then the DUMMY STATEMENT for it would appear next as

```
20025     QUET=0
```

The order in which you set variables equal to zero must be exactly the same as the order in which you will present their names to CODEPACK.

Now load TRANSPAK or LIZENCO from the appropriate diskette. Select Option 2, USER CODING, from the menu. The run will end so that you may enter your coding and DUMMY STATEMENTS. Then type RUN <ENTER>.

Select menu option 2, VARIABLE TRANSFORMATIONS.

You must provide

(1) NAME OF SOURCE FILE.

(2) NAME OF DESTINATION FILE

(3) TOTAL COLUMNS PER CASE IN THE SOURCE FILE.

(4) NUMBER OF VARIABLES IN THE SOURCE FILE TO BE REFERENCED IN USER CODING.

(5) NUMBER OF NEW VARIABLES TO BE CREATED.

(6) If you specify that N variables from the source file are to be referenced by the user's coding, these variables will be referred to as variable 1 to variable N. For each variable you must specify:

(a) the column number of the source file which contains the variable, and

(b) the name of the variable.

(7) You will be asked to provide a name for each of the variables which are to be created by user coding.

IMPORTANT: You must enter variable names in exactly the same order as they are set equal to zero in the DUMMY STATEMENTS.

Now CODEPACK will execute your DUMMY STATEMENTS; this will force BASIC to assign locations to all variables. CODEPACK will find these locations in the BASIC variables tables so that CODEPACK can provide the data needed by USER CODING for variable transformations.

You must next specify whether or not you want a case deleted if missing data appears in any source file variable you have named.

If you are using LIZENCO, you will be asked to specify COLS SAVED FROM SOURCE FILE TO DESTINATION FILE. Then you must specify which columns you want saved from the source file, and the column each is to occupy in the destination file (Note: if, for example, you save 7 columns from the source file, these columns must occupy the first seven columns of the modified file in some order; you dictate the order in which they appear.)

EXAMPLE

Suppose an AGGREGATE file contains the chemical analysis of samples of certain substances created at different chemical plants. These substances consist of water, chlorine, nitrogen, and potassium. Suppose the arrangement of the data in any row is as follows:

COL 1	CASE NUMBER (or sample number)
COL 2	PLANT NUMBER
COL 3	Total wt. of sample
COL 4	Wt. of water in sample
COL 5	Wt. of chlorine in sample
COL 6	Wt. of nitrogen in sample
Col 7	Wt. of potassium in sample

The name of the file is CHCOMP.AGG.

Suppose we wish to create a new file named CHCOMP.MOD which is to consist of

COL 1	Case Number
COL 2	Plant Number
COL 3	Total wt. of sample
COL 4	Wt. of chlorine in sample
COL 5	Wt. of nitrogen in sample
COL 6	Wt. of potassium in sample
COL 7	Combined percentage of chlorine, nitrogen, and potassium in sample

The following are the details of variable names, USER CODING, DUMMY STATEMENTS, and answers to queries for this example. Use LIZENCO for this example.

(1) Variable names

NAME	MEANING	COL. NUMBER (source file)
WT	Total Wt. of sample	3
CLWT	Wt. of chlorine	5
NWT	Wt. of nitrogen	6
KWT	Wt. of potassium	7
CMBPCT	Combined percentage of CL, N, and K in the sample (new variable)	

(2) USER CODING

```
10000 CMBPCT = CLWT + NWT + KWT
10005 CMBPCT = CMBPCT/WT
10010 CMBPCT = CMBPCT * 100
```

order of terms in the formulas does not have to be same as sequence in Dummy statements

(3) DUMMY STATEMENTS

```
20000 WT = 0
20005 CLWT = 0
20010 NWT = 0
20015 KWT = 0
20020 CMBPCT = 0
```

(4) RESPONSES (R) TO QUERIES (Q)

Q NAME OF:
SOURCE FILE
R CHCOMP.AGG
Q DESTINATION FILE
R CHCOMP.MOD
Q TOTAL NUMBER OF COLUMNS PER CASE FOR FILE CHCOMP.AGG
R 7

```
Q  NUMBER OF VARIABLES (COLUMNS) FROM FILE CHCOMP.AGG TO BE
           REFERENCED IN USER CODING
R  4
Q  NUMBER OF NEW VARIABLES TO BE CREATED BY USER CODING
R  1
Q  ENTER THE NAMES OF THE VARIABLES FROM CHCOMP.AGG AS THEY
           ARE TO BE REFERENCED BY USER CODING
   VARIABLE NUMBER 1
   ENTER COLUMN NUMBER FROM FILE CHCOMP.AGG
R  3
Q  VARIABLE NAME?
R  WT
Q  VARIABLE NUMBER 2
   ENTER COLUMN NUMBER FROM FILE CHCOMP.AGG
R  5
Q  VARIABLE NAME?
R  CLWT
Q  VARIABLE NUMBER 3
   ENTER COLUMN NUMBER FROM FILE CHCOMP.AGG
R  6
Q  VARIABLE NAME?
R  NWT
Q  VARIABLE NUMBER 4
   ENTER COLUMN NUMBER FROM FILE CHCOMP.AGG
R  7
Q  VARIABLE NAME?
R  KWT
Q  ENTER NAMES OF VARIABLES ADDED (i.e. variables to be created by
           user coding)
   VARIABLE NUMBER 1
   VARIABLE NAME?
R  CMBPCT
Q  DELETE CASES, MISSING DATA? (Y/N)
R  Y
Q  ENTER NUMBER OF COLUMNS TO BE SAVED FROM FILE CHCOMP.AGG
           TO FILE CHCOMP.MOD
R  6
Q  COLUMN 1 IN FILE CHCOMP.MOD CORRESPONDS TO WHAT FILE COLUMN NUMBER
           IN FILE CHCOMP.AGG?
R  1
Q  COLUMN 2 IN FILE CHCOMP.MOD CORRESPONDS TO WHAT FILE COLUMN NUMBER
           IN FILE CHCOMP.AGG?
R  2
Q  COLUMN 3 IN FILE CHCOMP.MOD CORRESPONDS TO WHAT FILE COLUMN NUMBER
           IN FILE CHCOMP.AGG?
R  3
Q  COLUMN 4 IN FILE CHCOMP.MOD CORRESPONDS TO WHAT FILE COLUMN NUMBER
           IN FILE CHCOMP.AGG?
R  5
Q  COLUMN 5 IN FILE CHCOMP.MOD CORRESPONDS TO WHAT FILE COLUMN NUMBER
           IN FILE CHCOMP.AGG?
R  6
```

Q COLUMN 6 IN FILE CHCOMP.MOD CORRESPONDS TO WHAT FILE COLUMN NUMBER
IN FILE CHCOMP.AGG?

R 7

If in the USER CODING section above we add the statements

```
10015 CLWT = (CLWT/WT)*100
10020 NWT = (NWT/WT)*100
10025 KWT = (KWT/WT)*100
```

then in the modified file CHCOMP.MOD, columns 4, 5, and 6 will contain the PERCENTAGE of chlorine, nitrogen, and potassium in the sample.

You may use all BASIC functions for variable transformations. These functions are listed in your computer manual.

TRANSFORMATION OF VARIABLES IN A CUSTOM REGRESSION ANALYSIS FILE: The last variable in a regression file is always treated as the dependent variable by LIZPACK. If a regression file is being transformed AND NEW VARIABLES ARE BEING CREATED, these variables appear as the last variables (columns) in the file. Thus, to maintain the dependent variable's position as last in the file, proceed as follows:

(1) Use the program LIZENCO for the transformations.

(2) Include the dependent variable as one of the CURRENT VARIABLES to be used and give it a name, e.g., call it VA.

(3) Include a newly created variable name for the dependent variable (e.g., call the new name VB). Present the new variable name for the dependent variable LAST when naming newly created variables. This will reserve the last column in the modified file for variable VB.

(4) In your subroutine for transforming variables, you would then include a statement

VB=VA

if the dependent variable is not being transformed. This would transfer the dependent variable to the last column of the modified file.

(5) Do not save the column in the source file containing the dependent variable to the destination file.

If the dependent variable is being transformed, assign its transformed value to the variable VB in your transformation subroutine.

It should be noted that LIZENCO reserves columns for newly created variables in the order in which their names are presented to the program. This allows the user to specify the order in which such variables appear in the modified file. The above example also illustrates how variables from the source file may be transferred to the modified file in any desired order.

CHAPTER VIII
PERFORMING REGRESSION ANALYSIS WITH LIZPACK

DATA FILES FOR REGRESSION ANALYSIS: A LIZPACK CUSTOM regression analysis data file which contains K variables (including the dependent variable) is an AGGREGATE file containing K columns. The file obviously cannot contain case numbers or ID numbers. However, the file is in a convenient form for the following types of processing:

(1) Addition of variables using the program MERGEFILE (see Chapter III).

(2) Deletion of variables using the program DELETE.VAR (see Chapter III).

(3) Transformation of variables using programs TRANSPAK or LIZENCO (see Chapter III).

(4) If it is necessary to eliminate some of the variables in the file from the regression analysis, the CUSTOM file may be treated as an AGGREGATE file.

MULTIPLE REGRESSION ANALYSIS: The LIZPACK multiple regression analysis package is quite comprehensive and allows a complete analysis of correlated data. A complete program for a regression analysis is as follows:

STEP I: A correlation matrix is prepared for all variables suspected of contributing to the analysis. This matrix should be prepared by either of the programs LIZCORR.XTM or LIZCORR.MUL by selecting the option

3. BEGIN RUN - COMPUTE CORR. MAT.

on the main menu. To print the correlation matrix, select option

4. CORRELATION COEFFICIENTS

This must be done BEFORE doing a multiple regression analysis with these programs.

Correlation coefficients will be printed in a table with

a. numbers of the variables,
b. the correlation coefficient,
c. degrees of freedom for Student's T, and
d. Student's T for testing the significance of the correlation coefficient.

The correlation coefficient should be examined, and meaningless variables should be deleted from the analysis before proceeding to Step II. This is important because if a regression run is not executed with LIZCOR.MUL or LIZCORR.XTM, these programs will have space to compute the correlation matrix of considerably more variables than can be handled by LIZANAL.NAM or LIZANAL.MUL.

STEP II: The proper model should now be selected using programs LIZANAL.NAM or LIZANAL.MUL. These programs allow the user complete control of the number and arrangement of the variables in the model at any time. They are especially well adapted to performing Forward Stepwise Regression and Backward Elimination. Their design follows the philosophy that a statistical package for the personal computer should allow a maximum amount of meaningful user interaction; this philosophy is in contrast to the fully automated batch procedures which are necessary for the efficient utilization of large-scale computers.

FORWARD STEPWISE REGRESSION: If a large number of variables are involved, this is usually the only efficient method of modeling. The following algorithm is used to select the proper model; regression statistics at each step may be printed if they are needed. The proper model may be selected much more rapidly if regression statistics are not printed, but in any event a note should be made of the order in which the variables enter the model.

Select the FORWARD MODE option on the main menu. After a considerable amount of computation you will be asked to enter the name (or number) of two variables. In the following algorithm, substitute "NUMBER" if you are using LIZANAL.MUL; this algorithm is given for program LIZANAL.NAM. YOU WILL NEED AN F-TABLE.

ALGORITHM

1. Enter the name of the independent variable which is most correlated with the dependent variable, then enter the name of the dependent variable.
2. If regression statistics are desired, select option BACKWARD MODE, do a REGRESSION RUN, then select option FORWARD MODE.
3. Select the PARTIAL F'S option. Partial F's of all variables NOT in the model are listed on the screen with their denominator degree of freedom (DF) (numerator DF is always 1). If no partial F is significant, then the variables on the VARIABLES AVAILABLE list form the proper model.
4. If one or more of the partial F's are significant, add the variable with the most significant partial F using the option ADD VARIABLES.
5. Select the BACKWARD MODE option.
6. Select the PARTIAL F'S option. The partial F's of all variables in the model with their denominator DF will be listed on the screen.
7. If all the partial F's are significant, go to 8. Otherwise, delete the variable with the least significant partial F using the DELETE VARIABLE option. Then return to step 6.
8. If regression statistics are desired, print them by performing a REGRESSION RUN.
9. Select option FORWARD MODE. Go to step 3.

BACKWARD ELIMINATION: Select the BACKWARD MODE option on the main menu, then follow this algorithm.

ALGORITHM

1. If regression statistics are desired, select the REGRESSION RUN option.
2. Select option PARTIAL F'S. The partial F's and denominator DF of each variable in the model is listed on the screen. If all partial F's are significant, the variables on the VARIABLES AVAILABLE list form the proper model.

3. If some partial F is not significant, delete the variable with the least significant F by using the option DELETE VARIABLE.
4. Go to step 1.

NOTE: While the Backward Elimination procedure appears to be simpler, the Forward Stepwise procedure is usually much faster. The Backward Elimination procedure may be quite expensive as far as computing time is concerned.

FORCING VARIABLES

If certain variables are to be forced into the model, the best procedure will usually be to select the FORWARD MODE option from the main menu. The variables to be forced into the model may be introduced one at a time using the ADD VARIABLE option or all at once using the option SET VARIABLE POINTERS on the BACKWARD MODE options menu. Then follow the algorithm for the Forward Stepwise procedure using all remaining variables, but do not delete any forced variables. REMEMBER THAT THE LAST VARIABLE ON THE VARIABLES AVAILABLE LIST MUST BE THE DEPENDENT VARIABLE.

After the proper model has been selected, all insignificant variables should be deleted from the analysis before proceding to Step III. This will save time, and in addition, because of the graphics capability of the residual analysis program LIZANAL.RES, it will handle fewer variables than either of the modeling programs.

<u>STEP III</u>: After the proper model has been selected, a residual analysis may be performed using program LIZANAL.RES which operates in both the FORWARD MODE and BACKWARD MODE. As in the modeling programs, variables may be deleted, added, rearranged by setting the variable pointers, and all variables may be restored to the variables available list.

The regression statistics printed by this program are more comprehensive than those printed by the other multiple regression programs. The correlation between any pair of variables may be examined graphically. Residual plots may be performed to allow graphical examination of the residuals. See the program description for LIZANAL.RES for operating instructions.

If it is so desired, the user may follow the same procedures for introducing and removing variables as dictated by the modeling programs and print regression statistics at each stage. At the user's option, a residual analysis may be performed for each stage.

RESIDUAL STATISTICS: The residual statistics printed by this program are discussed in Chapter 3 of Draper and Smith's <u>Applied Regression Analysis</u>, 2nd Edition, Wiley, 1981. The Durbin-Watson test for serial correlation among the residuals is discussed beginning on page 162. See page 150 for a discussion and definition of the statistics T(1,1), T(1,2), and T(2,1) which measure the banding effects of the residuals. A discussion of the interpretation of residual plots is to be found on pages 147-149.

TERMINOLOGY: The standard term "Standard Error of Estimate" is sometimes called the "Standard Deviation of Residuals."

The term "Error" in ANOVA tables under "Source" is also called "Residual." In any event, it is the sum of squares of the residuals.

Draper and Smith (op.cit.) refer to the dependent variable as the "Response."

PARTIAL CORRELATION COEFFICIENTS are computed and printed by programs LIZANAL.NAM and LIZANAL.MUL. Appropriate statistics for tests of significance are also given.

STRAIGHT LINE PREDICTIONS and their confidence intervals are performed by LIZANAL.RES. Residuals for the straight line fit may be computed, printed, and saved to diskette. Graphical examination of the curve fit is available.

POLYNOMIAL AND EXPONENTIAL REGRESSION: The program LIZANAL.PLY (described in Chapter VI, Section 16) fits a curve of the form y=P(x), where P(x) is a polynomial, by the least squares method, and LIZEXPO.ANL (Chapter IV, Section 16) fits a curve y=exp(P(x)) where P(x) is a polynomial.

One should strive to fit the polynomial of lowest degree which is statistically significant. Polynomials of high degree are undesirable for many extrapolation, or even interpolation purposes. Attempting to fit a polynomial of very high degree may cause an overflow error if the independent variable contains values with large absolute values.

These programs allow considerable user control and plot graphs in two coordinate systems. Modeling is done by entering the program with a polynomial of degree which is an estimate of the upper bound for a polynomial fit for the data. The degree may then be reduced as needed.

If LIZEXPO.ANL is used, then all values of the dependent variable must be positive; the actual curve fit is for Ln(y)=P(x). Graphs are drawn for the curve y=exp(P(x)).

The utility programs LIZPOLY.ONE and LIZPOLY.STP are described in Chapter VI, Section 18.

Residuals for all these programs are computed by the program LIZPOLY.RES. Residuals are printed and saved to diskette for possible further analysis.

USING AGGREGATE FILES AS INPUT FILES: The CORE PACKAGE regression programs accept CUSTOM regression analysis files or AGGREGATE files as input files. The use of custom files as input files is illustrated in Chapter IX. To use an AGGREGATE file as an input file, the procedures of Chapter IX are modified as follows:

Answer the query

IS THIS RUN FROM A CUSTOM FILE?

by

NO<ENTER>

You will then be asked to provide

(1) the total number of columns in the AGGREGATE file, and

(2) the file column number for each variable. For example, if you have answered the query

NUMBER OF VARIABLES

with 10 (as the total number of variables to be entered into a multiple regression analysis), you must enter 10 file column numbers, one for each variable. ENTER THE FILE COLUMN NUMBER FOR THE DEPENDENT VARIABLE LAST.

For the bivariate regression analysis programs such as LIZANAL.PLY, you will be asked to enter the file column number for the independent and dependent variables.

MISSING DATA: If LIZPACK encounters a missing data symbol in any column which has been specified as a column containing one of the regression variables, then the entire case will be ignored. CAUTION: If the regression program has graphics capabilities, then the first case of the file must not contain a missing data symbol in any column containing a regression variable.

OTHER REGRESSION PROGRAMS: The special regression programs LIZRAD.PWR, LIZLOG.ANL and LIZEXPO.ONE are described in Chapter VI, Section 17 of this manual.

Further regression programs are to be found in the forecasting package (PACKAGE E). While these programs are used primarily for forecasting purposes, some of them could prove useful in other contexts.

DESCRIPTIVE STATISTICS: You may obtain full descriptive statistics for each variable in a regression analysis data file by using the program LIZDESC.AGG which produces descriptive statistics from an AGGREGATE file. A regression analysis file for K variables may be regarded as an AGGREGATE file with K columns. The program LIZDESC.AGG appears on diskette 6.

REFERENCES

1. Dixon, W. J., Ed., BMD Biomedical Computer Programs, Berkeley, University of California Press, 1979.

2. N. R. Draper and H. Smith, Applied Regression Analysis, John Wiley and Sons, New York, 1981.

3. D. G. Kleinbaum and L. L. Kupper, Applied Regression Analysis and Other Multivariable Methods, Duxbury Press, North Scituate, Mass., 1978.

4. N. H. Nie, et.al., SPSS, Statistical Package For the Social Sciences, McGraw-Hill, New York, 1975.

5. R. G. D. Steel and James H. Torrie, Principles and Procedures of Statistics, McGraw-Hill, New York, 1960.

CHAPTER IX
PRACTICE DATA FILES

DISK 8

A diskette containing a number of practice data files is included with LIZPACK. These data files are described below. The information needed to answer LIZPACK queries is provided along with the programs for which the data files correspond.

The user should read the instructions for individual programs before running them.

After reading the instructions for the individual programs, the data files may be used for practice runs. Data files may be listed to the screen or the printer using the editor; this should help to further familiarize the user with the format of LIZPACK custom files.

To make a practice run on a program, you should answer the query
NAME OF DATA FILE
or
NAME OF OBSERVATION FILE
by giving the name of the data file listed. Appropriate information for other queries is given below.

After the name of each data file a reference and page number appears. The reference refers to one of the text books listed below from which the data is taken. You may refer to the appropriate text for a complete description of the problem to which the data applies. Some of the older texts may be out of print; however, they should be available at many libraries.

REF. A: D. G. Kleinbaum and L. L. Kupper, Applied Regression Analysis and Other Multivariate Methods, Duxbury Press, 1978.

REF. B: F. C. Mills, Introduction to Statistics, Henry Holt and Co., 1956.

REF. C: R. G. D. Steel and James H. Torrie, Principles and Procedures of Statistics, McGraw-Hill, 1960.

1. PROGRAM: LIZSAMP.FRQ
   DATA FILE: BUTTRFAT.NEW
   REF. C, p. 50, Table 4.1
   FREQUENCY DISTRIBUTION option
   Let the system find the maximum and minimum.
   Number of class intervals: 13
   RANDOM SAMPLING option

Step 1 MAXIMUM: 53.5
MINIMUM: 26.5

Step 2 Number of class intervals: 19
Step 3 Number of samples: 500
Observations per sample: 10

THEORETICAL FREQUENCY DISTRIBUTION OPTION

No queries.

2. PROGRAM: LIZDESC.DES
DATA FILE: DOSAGE.GMS
REF. A, p. 266, Table 17.1
NUMBER OF SAMPLES: 4

3. PROGRAM: LIZMEAN.MCP
DATA FILE: DOSAGE.GMS
REF. A, p. 266, Table 17.7
Name of population unit: SUBST
Name of variable: POTENCY
Number of samples: 4

4. PROGRAM: LIZ1WAY.ONE
DATA FILE: DOSAGE.GMS
REF. A, p. 266, Table 17.7
NAME OF POPULATION UNIT: SUBST
NAME OF VARIABLE: POTENCY
NUMBER OF SAMPLES: 4

5. PROGRAM: LIZ1WAY.ANC
DATA FILE: BACON.DAT
REF. C, p. 321, Table 15.7
NAME OF POPULATION UNIT: RATION
NAME OF COVARIATE: INITWT
NAME OF VARIABLE: WTGAIN
NUMBER OF SAMPLES: 3

6. PROGRAM: LIZANOVA.RAN
DATA FILE: QUET.DAT
REF. A, p. 300, Table 18.4
Name of treatments: DIET
Name of blocks: AGQTGRP
Name of variable: SBP LEVEL
Number of treatments: 4
Number of blocks: 8

7. PROGRAM: LIZANCOV.RAN
DATA FILE: LIMAS.DAT
REF. C, p. 312, Table 15.2
NAME OF TREATMENTS: VARIETY
NAME OF BLOCKS: FIELD
NAME OF COVARIATE: DRYWT

NAME OF VARIABLE: VITC
NUMBER OF TREATMENTS: 11
NUMBER OF BLOCKS: 5

8. PROGRAM: LIZ2WAY.TWO
   DATA FILE: FEV.DAT
   REF. A, p. 323, Table 19.6
   Name of Row factor: PLANT
   Name of Column factor: TOXSUB
   Name of variable: FEV
   Observations per cell: 12
   Number of rows: 3
   Number of columns: 3

9. PROGRAM: LIZANCOV.2WY
   DATA FILE: WEIGHT.PIG
   REF. C, p. 321, Table 15.7
   NAME OF COVARIATE: INITWT
   NAME OF VARIABLE: WTGAIN
   NAME OF ROW FACTOR: RATION
   NAME OF COLUMN FACTOR: SEX
   OBSERVATIONS PER CELL: 5
   NUMBER OF ROWS: 3
   NUMBER OF COLUMNS: 2

10. PROGRAM: LIZBLOCK.2WY
    DATA FILE: OATS.DAT
    REF. C, p. 237, Table 12.3
    NAME OF VARIABLE: YIELD OF OATS
    NAME OF ROW FACTOR: VARIETY
    NAME OF COLUMN FACTOR: FUNG
    NAME OF BLOCKS: FIELD
    OBSERVATIONS PER CELL: 4
    NUMBER OF ROWS: 4
    NUMBER OF COLUMNS: 4

11. PROGRAM: LIZANCOV.BLK
    DATA FILE: WEIGHT.PIG
    REF. C, p. 321, Table 15.7
    NAME OF COVARIATE: INITWT
    NAME OF VARIABLE: WTGAIN
    NAME OF ROW FACTOR: RATION
    NAME OF COLUMN FACTOR: SEX
    NAME OF BLOCKS: PEN
    OBSERVATIONS PER CELL: 5
    NUMBER OF ROWS: 3
    NUMBER OF COLUMNS: 2

12. PROGRAM: LIZANOVA.NEQ
DATA FILE: MEDCARE.DAT
REF. A, p. 352, Table 20.2
NAME OF ROW FACTOR: AFFCOM
NAME OF COLUMN FACTOR: WORRY
NAME OF VARIABLE: SATISFACTION WITH MEDICAL CARE
NUMBER OF ROWS: 3
NUMBER OF COLUMNS: 2

13. PROGRAM: LIZANAL.NAM OR LIZANAL.MUL
DATA FILE: QUETSBP.DAT
REF. A, p. 60, Problem 2
NUMBER OF VARIABLES: 4
NUMBER OF CROSS TERMS: 3

VARIABLE NAMES

VARIABLE	NAME
1	QUET
2	AGE
3	SMK
4	SBP

CROSS TERM NAMES

TERM	NAME
1	QTSMK
2	AGSMK
3	AQS

CROSSTERM DEFINITIONS

TERM	DEFINITION
1	QUET,SMK
2	AGE,SMK
3	AGE,QTSMK

Answer NO to LOG OPTION query.

NOTE: There is one additional query, namely
IS THIS A RUN FROM A CUSTOM FILE?

Answer
YES<ENTER>

If you answer NO, then additional information must be provided. See the instructions in Chapter VIII.

14. PROGRAM: LIZANAL.PLY
DATA FILE: TENDOM.RDG
REF. C, p. 339, Table 16.3
DEGREE OF POLYNOMIAL: (the best statistical choice is 2. For purposes of experimentation you might want to use 5 or 6 here).

NOTE: There is one additional query, namely
IS THIS A RUN FROM A CUSTOM FILE?

Answer
YES<ENTER>

If you answer NO, then additional information must be provided. See the instructions in Chapter VIII.

15. PROGRAM: LIZEXPO.ANL
DATA FILE: PETPROD.DAT
REF. b, p. 311, Table 10.9
DEGREE OF POLYNOMIAL: (The best statistical choice is 1. You may use a higher degree for experimentation.)
NOTE: There is one additional query, namely
IS THIS A RUN FROM A CUSTOM FILE?
Answer
YES<ENTER>
If you answer NO, then additional information must be provided. See the instructions in Chapter VIII.

16. PROGRAM: LIZRAD.PWR
DATA FILE: TENDOM.RDG
REF. C, p. 339, Table 16.3
NOTE: There is one additional query, namely
IS THIS A RUN FROM A CUSTOM FILE?
Answer
YES<ENTER>
If you answer NO, then additional information must be provided. See the instructions in Chapter VIII.

17. PROGRAM: LIZEXPO.ONE
DATA FILE: PETPROD.DAT
REF. B, p. 311, Table 10.9
NOTE: There is one additional query, namely
IS THIS A RUN FROM A CUSTOM FILE?
Answer
YES<ENTER>
If you answer NO, then additional information must be provided. See the instructions in Chapter VIII.

18. PROGRAM: LIZLOG.ANL
DATA FILE: TENDOM.RDG
REF. C, p. 339, Table 16.3
NOTE: There is one additional query, namely
IS THIS A RUN FROM A CUSTOM FILE?
Answer
YES<ENTER>
If you answer NO, then additional information must be provided. See the instructions in Chapter VIII.

19. PROGRAM: LIZPOLY.ONE
DATA FILE: TENDOM.RDG
REF. C, p. 339, Table 16.3
DEGREE OF POLYNOMIAL: 2
NOTE: There is one additional query, namely
IS THIS A RUN FROM A CUSTOM FILE?

Answer

YES<ENTER>

If you answer NO, then additional information must be provided. See the instructions in Chapter VIII.

20. PROGRAM: LIZPOLY.STP
DATA FILE: TENDOM.RDG
REF. C, p. 339, Table 16.3
DEGREE OF POLYNOMIAL: 3
NOTE: There is one additional query, namely
IS THIS A RUN FROM A CUSTOM FILE?

Answer

YES<ENTER>

If you answer NO, then additional information must be provided. See the instructions in Chapter VIII.

21. PROGRAM: LIZCORR.XTM
DATA FILE: SBPQUET.DAT
REF. A, p. 60, Problem 2
NUMBER OF VARIABLES: 4
NUMBER OF CROSSTERMS: 3

DEFINITION OF CROSSTERMS

TERM	DEFINITION
1	1,3
2	2,3
3	2,4

NOTE: There is one additional query, namely
IS THIS A RUN FROM A CUSTOM FILE?

Answer

YES<ENTER>

If you answer NO, then additional information must be provided. See the instructions in Chapter VIII.

22. PROGRAM: LIZCORR.MUL
DATA FILE: TOBTEST.DAT
REF. C, p. 282, Table 14.1
NUMBER OF VARIABLES: 4
NOTE: There is one additional query, namely
IS THIS A RUN FROM A CUSTOM FILE?

Answer

YES<ENTER>

If you answer NO, then additional information must be provided. See the instructions in Chapter VIII.

23. PROGRAM: LIZSAMP.FRQ
DATA FILE: TEXTILE.DAT
REF. B, p. 43

FREQUENCY DISTRIBUTION OPTION

User set maximum: 68.5, minimum: 38.5
Number of class intervals: 31

This data file produces an excellent example of a HISTOGRAM and FREQUENCY POLYGON.

24. PROGRAMS: LIZ1WAY.AGG, LIZ2WAY.AGG, LIZRAND.AGG, LIZBLOCK.AGG, LIZ1ANC.AGG, LIZ2ANCV.AGG, LIZANRAN.AGG, LIZBLOCK.AGG.
DATA FILE: BACON.AGG
REF. C, P. 321, TABLE 15.7.

This is an AGGREGATE FILE containing 5 columns. The contents of the columns are as follows:

Column 1: independent variable
Name of variable: RATION
Low value: 1
High value: 3

Column 2: independent variable
Name of variable: SEX
Low value: 1
High value 2

Column 3: independent (block) variable
Name of variable: PEN
Low value: 1
High value: 5

Column 4: covariate
Name of variable: INITWT

Column 5: dependent variable
Name of variable: WTGAIN

The above information, together with the following information, is used to answer the queries of the indicated programs.

LIZ1WAY.AGG

Number of samples: 3
Column number for independent variable: 1 (see column 1 above)
Column number for dependent variable: 5 (see column 5 above)

LIZ2WAY.AGG

Number of rows: 3
Number of columns: 2
Column number for row variable: 1
Column number for col. variable: 2
Column number for dependent variable: 5

LIZRAND.AGG

Number of treatments: 3
Number of blocks: 5
Column number for row (treatment) variable: 1
Column number for block variable: 3
Column number for dependent variable: 5

LIZBLOCK.AGG

Number of rows: 3
Number of columns: 2
Number of blocks: 5
Column number for row variable: 1
Column number for col. variable: 2
Column number for block variable: 3
Column number for covariate: 4
Column number for dependent variable: 5

LIZ1ANC.AGG

Same as for LIZ1WAY.AGG and
Column number for covariate: 4

LIZ2ANCV.AGG

Same as for LIZ2WAY.AGG and
Column number for covariate: 4

LIZANRAN.AGG

Same as for LIZRAND.AGG and
Column number for covariate: 4

LIZBLANC.AGG

Same as for LIZBLOCK.AGG and
Column number for covariate: 4

CHAPTER X
THE LIZPACK COMPARATIVE STATISTICS PROGRAMS

The LIZPACK Comparative Statistics Package includes programs for multiple comparison procedures, one-way ANOVA, two- and three-way ANOVA with equal and non-equal cell frequencies, and ANOVA programs for one- and two-way randomized block experiments. In addition, there are four analysis of covariance procedures (one covariate) and a regression model package.

LIZPACK supports both load-and-run programs which use CUSTOM files as input, and programs which accept the conventional AGGREGATE files as input. These programs are listed according to input file type in the APPENDIX to Chapter II.

The load-and-run programs will generally process a larger number of samples or cells than their AGGREGATE file counterparts. This does not become a factor unless the problem is quite large; for example, in the case where this limitation is most severe are the two- and three-way ANOVA (non-equal cell frequencies) programs. The limitation on the number of cells in these cases does not become a factor until the number of cells to be processed exceeds 800.

For large problems, the processing time required by the load-and-run programs may be significantly less than for the AGGREGATE file mode programs.

The LIZPACK comparative statistics programs do not accept multiple diskette data files as input files.

The output for the AGGREGATE file mode programs is the same as for their load-and-run CUSTOM file mode counterparts. (See Chapter VI.)

When operating the AGGREGATE file mode programs you must specify the file column numbers for 1 to 3 independent variables, depending on the type of analysis involved. For each of these you must specify a high value and a low value of the independent variable. These correspond to the levels of the independent variable. For example, if you specify a low value of 3 and a high value of 7 for an independent variable, then LIZPACK will assume levels of 3, 4, 5, 6, and 7 for the independent variable. There must be no gaps in the levels: e.g., if the data file contained observations for levels 3, 5, and 7 only, then the analysis would have to be performed after creating a suitable CUSTOM file or recoding the independent variable. LIZPACK will ignore any cases with observations outside the ranges specified by the low and high values of the independent variable.

For a one-way ANOVA or ANCOVA the file column numbers for a single independent variable must be specified along with its low and high values. The levels of the independent variable identify the groups whose means are to be compared.

For a one-way randomized block experiment or a two-way ANOVA, you must specify the file column numbers of two independent variables, along with their high and low values. The levels of the independent variables identify the cases which belong to a particular cell.

For a two-way randomized block experiment or a three-way ANOVA you must specify the file column numbers of three independent variables along with their high and low values. The levels of the independent variables identify the cases which belong to a particular cell.

For any ANOVA program you must specify a file column number for the dependent variable.

For any ANCOVA program you must specify a file column number for the covariate.

MISSING DATA: Data files for the load-and-run CUSTOM file mode programs may not contain missing data symbols.

For the AGGREGATE file mode programs, the following conventions apply:

(1) On all programs, it is assumed that missing data symbols do not appear in the columns containing the independent variable or else they lie outside the range of the independent variables.

(2) In the case of a one-way ANOVA or ANCOVA it is expected that missing data symbols may occur in the columns containing the dependent variable and/or covariate. When a missing data symbol occurs, the case in which it occurs will be ignored.

(3) In the case of a one or two randomized block experiment ANOVA or ANCOVA, a two-way ANOVA or ANCOVA (equal cell frequencies) or a three-way ANOVA (equal cell frequencies), it is assumed that missing data symbols will not occur.

(4) In the case of two- and three-way ANOVAs (non-equal cell frequencies), it is expected that missing data symbols may occur in the columns containing the dependent variable. When a missing data symbol is encountered, the case in which it occurs is ignored.

IMPORTANT NOTES:

(1) Do not attempt to perform a two- or three-way ANOVA (or two-way ANCOVA) with one observation per cell. Treat such a problem as a randomized block experiment.

(2) The two- and three-way ANOVA progams for unequal cell frequencies do not handle the case of empty cells. Use the regression model when empty cells are present.

RELATED PROGRAMS: The discriminant analysis and canonical correlation package (Package C) contains programs which perform a one-way MANOVA and one-way MANCOVA.

REFERENCES

1. W. G. Cochran and G. M. Cox, Experimental Designs, Wiley, New York, 1957.

2. B. F. Huitema, The Analysis of Covariance and its Alternatives, Wiley, N.Y., 1980.

3. D. G. Kleinbaum and L. L. Kupper, Applied Regression Analysis and other Multivariate Methods, Duxbury Press, North Scituate, Mass., 1978.

4. N. H. Nie, et.al., SPSS: Statistical Package for the Social Sciences, McGraw-Hill, New York, 1975.

5. R. G. D. Steel and J. H. Torrie, Principles and Procedures of Statistics, McGraw-Hill, New York, 1960.

CHAPTER XI
THE LIZPACK FACTOR ANALYSIS PACKAGE

The LIZPACK Factor Analysis Package is based on the methods developed in the classic book Modern Factor Analysis by H. H. Harman, and by Jennrich and Sampson in their Psychometrica papers of 1966 and 1970.

A complete program for Factor Analysis consists of various combinations of the following steps:

(1) The preliminary preparation of the correlation matrix by program LIZCORR.MAT;

(2) A principal component analysis or a preliminary estimation of the principal factors by program LIZCOMP.PCL;

(3) The iterative estimation of the principal factors by program LIZFACTR.ANL;

(4) An orthogonal rotation by means of the programs LIZQUART.MAX, LIZVARI.MAX, or LIZORTHO.MAX;

(5) An oblique rotation by means of program LIZOBLI.MIN; and

(6) Computation of the factor score coefficient matrix and the factor scores by means of LIZSCORE.FAC.

The programs in the Factor Analysis Package generate data files which are used in succeeding programs. The data diskette used for these programs should not contain a data file with the same name. See the OUTPUT section of each program for the names of the system-generated data files.

All system-generated data files are created on the default disk drive. If you have more than one disk drive, you may place the raw data on a drive other than the default drive when using LIZCORR.MAT, and place a formatted diskette on the default drive. The initial correlation matrix will then be written on the diskette on the default drive which may then be used when running the succeeding programs.

Descriptions of the individual programs follow.

1. COMPUTATION OF THE CORRELATION MATRIX
Program LIZCORR.MAT, Diskette 1

The option which must be chosen first is MAIN MENU OPTION 3: CORRELATION MATRIX. The program computes the correlation matrix and writes it to a diskette data file.

The input data file is a regression analysis data file. The data file may be a CUSTOM regression analysis data file or an AGGREGATE file (see Chapter VIII).

You must provide

(1) the number of variables used in the analysis;

If you answer NO to the query IS THIS RUN FROM A CUSTOM FILE? then you must provide

(2) the total number of columns in the AGGREGATE file;

(3) the column number in the AGGREGATE file for each variable.

OUTPUT: A diskette data file containing the correlation matrix. The name of this file is

CORRMAT.RIX

The file will be created on the default drive.

The other options on the MAIN MENU may now be chosen.

If you choose OPTION 4. PRINT CORR. COEFFICIENTS, a table will be printed containing

(1) each pair of variable numbers and their correlation coefficients,

(2) the number of degrees of freedom for each correlation coefficient, and

(3) the value of Student's T for the T-test of significance of the correlation coefficient.

If you choose OPTION 5. CORRELATION MATRIX - SMC'S, then the program will

(1) compute the inverse of the correlation matrix,

(2) place the squared multiple correlation coefficient (SMC's) on the main diagonal of the correlation matrix in the file CORRMAT.RIX.

This option is exercised only if you wish to use the second type of initial estimate for the iterative estimation of the principals factors by program LIZFACTR.ANL which is described below.

NOTE: If you exercise OPTION 5, you may need to use the EDITOR to replace the 1's on the main diagonal of the correlation matrix before using program LIXSCORE.FAC (see the INPUT FILES section of the program).

MISSING DATA: If a missing data symbol occurs in a column which contains a variable, the corresponding case will be ignored.

2. PRINCIPAL COMPONENT ANALYSIS OR PERLIMINARY ESTIMATE OF THE PRINCIPAL FACTORS
Program LIZCOMP.PCL, Diskette 4

The type of output obtained from this program depends on the type of matrix which is input to the program.

If the input matrix for the program is a correlation matrix with 1's on the diagonal, then the program performs a principal component analysis.

If the input matrix is a correlation matrix with SMC's or other estimates of the communalities on the diagonal, then the program output is a preliminary estimate of the principal factors.

The MAIN MENU options are as follows:

OPTION 1. CORR. MATRIX - SYSTEM INPUT

The program inputs the correlation matrix CORRMAT.RIX which was produced by program LIZCORR.MAT as described above. You must provide the number of variables to be used in the analysis.

OPTION 2. CORR. MATRIX - USER INPUT

This allows user input of a correlation matrix. You will be asked to enter those elements of the correlation matrix above the diagonal. The data file CORRMAT.RIX is created. The correlation matrix will contain 1's on the diagonal.

OPTION 3. COMMUNALITIES - USER INPUT

This option allows the user to input estimates of the communalities; this option should only be exercised after exercising option 1 or option 2 above.

OPTION 4. PRINCIPAL COMPONENT ANALYSIS

This option must be used after inputting a correlation matrix by means of one of the options described above.

The output from this option is a principal components analysis or a preliminary estimate of the principal factors (which will still be called "principal components"). After choosing this option, you will be asked to give each variable a name. Use up to 5 alphanumeric symbols in naming the variables.

OUTPUT:

PRINTER:

(1) The contribution to variance (eigenvalues of the correlation matrix) of each component arranged in decreasing order of magnitude. The symbol "PC" stands for "principal component".

(2) The percent contribution to variance by component.

(3) The eigenvalue of the correlation matrix.

(4) The coefficient of the principal components. If the input matrix does not have all 1's on the main diagonal, the matrix may not be positive definite. If zero or negative eigenvalues occur, then the corresponding principal component will have all zero components.

DISKETTE:

The data file COMPON.PCL is produced, which contains the coefficients of the principal components. This file may be used as input for any of the programs described below.

3. ITERATIVE ESTIMATION OF PRINCIPAL FACTORS
Program LIZFACTR.ANL

The input for this program is the output file COMPON.PCL from LIZCOMP.PCL. The factoring methods used is to retain all factors whose corresponding eigenvalues are GREATER than 1, unless the user specifies that a smaller or larger number of factors are to be retained. The preliminary estimation for the principal factors is obtained from the output of LIZCOMP.PCL. The types of preliminary estimates which are obtained from LIZCOMP.PCL are:

TYPE 1: The preliminary estimates for the principal factors will be the principal components if a correlation matrix with 1's on the diagonal was used as input to LIZCOMP.PCL (see Harman).

TYPE 2: If the input matrix for LIZCOMP.PCL contained SMC's on the diagonal, the preliminary estimates lead to the method called PA2 FACTORING by SPSS (see Harman).

TYPE 3: Other types of preliminary estimates may be obtained by user input of estimates for the communalities in program LIZCOMP.PCL.

The method of iterative estimation is the same regardless of the type of the preliminary estimate.

You must first decide whether or not you wish to enter the number of principal factors to be retained; if you answer NO to the query which allows you to specify the number of factors, then the program will retain all factors whose eigenvalue is GREATER THAN 1. Considerable time may be saved if you retain only the minimum number of meaningful factors.

At each stage of the iteration, the same number of factors are retained. The estimate of the i^{th} communality is obtained by taking the sum of squares of the elements in the i^{th} row of the principal factor coefficient matrix, where $i = 1, \ldots, M$ and where M is the number of factors retained.

Iteration will continue until one of the following conditions are fulfilled.

(1) Convergence (to the specified degree of accuracy) occurs, or
(2) 25 iterations have been performed, or
(3) Some communality exceeds 1.

When any of the conditions occur, iteration will stop and the last estimate of the principal factor coefficient matrix (Factor Pattern Matrix) will be used for output.

OUTPUT:

PRINTER:

(1) The contribution to variance (eigenvalue) by factor,
(2) Percent contribution by factor,
(3) Contribution to variance, by variable,
(4) Eigenvectors (columns of the transform matrix), and
(5) Factor pattern matrix.

DISKETTE:

The file "COMPON.PCA" is created. This file contains the factor pattern matrix which may be used as input for any of the programs described below.

4. ORTHOGONAL ROTATION OF PRINCIPAL FACTORS OR PRINCIPAL COMPONENTS

Program LIZQUART.MAX, Diskette 4
Program LIZVARI.MAX, Diskette 4
Program LIZORTHO.MAX, Diskette 4

There are three methods of orthogonal rotation available. The input file is one of the following:

(1) The file COMPON.PCA, created by LIZFACTOR.ANL, will be used if the MAIN MENU option ROTATION OF AXES is chosen.

(2) The file COMPON.PCL, created by program LIZCOMP.PCL, will be used if the MAIN MENU option ROTATION - PRINCIPAL COMPS. is chosen. The program will perform a rotation of principal components. Unless you specify otherwise, only those components with eigenvalaues GREATER THAN 1 will be rotated.

(3) You may elect to enter a factor pattern matrix for rotation by choosing the MAIN MENU option FACTOR MATRIX - USER INPUT. If you choose this option you must provide

- (a) the number of variables;
- (b) The number of factors;
- (c) the coefficients of each factor;
- (d) a name (5 characters or less) for each variable.

The following are the three methods of orthogonal rotation.

(1) QUARTIMAX, program LIZQUART.MAX. A description of the Quartimax criterion is found in the book by Harman (see the reference at the end of this chapter), pp. 283-290.

(2) VARIMAX, program LIZVARI.MAX. This is probably the most popular method of orthogonal rotation. A description of this method may be found in Harman, pp. 290-299.

(3) ORTHOMAX, program LIZORTHO.MAX. This is a class of orthogonal rotations (see Harman, p 299, and Jennrich, 1970). If you choose this method of rotation, you will have the option of entering a parameter GAMMA which controls the rotation criterion. You should not use

- (a) a zero value of GAMMA; use the QUARTIMAX method instead;
- (b) the value of 1 for GAMMA; use the VARIMAX method instead;
- (c) negative values of GAMMA.

If you choose not to enter a value of GAMMA, this program will default to the widely used EQUAMAX method of orthogonal rotation.

OUTPUT:

PRINTER:

- (1) a list of variables,
- (2) the rotated factor pattern matrix,
- (3) the transform matrix,
- (4) contribution to variance by factor,
- (5) percent contribution to variance by factor, and
- (6) contribution to variance by variable.

DISKETTE: The data file COMPON.ROT is created. This file contains the coefficients of principal factors (rotated factor pattern matrix). This file may be used as input for the program LIZSCORE.FAC for the purpose of computing the factor score coefficients and factor scores.

5. OBLIQUE ROTATION OF PRINCIPAL FACTORS OR COMPONENTS
Program LIZOBLI.MIN, Diskette 4

The method of oblique rotation used by LIZPACK is the DIRECT OBLIMIN method. This method obtains the factor pattern method directly without computing an intermediate reference structure; it is the method which is most frequently used by present-day researchers.

The DIRECT OBLIMIN method is discussed in Harman, pp. 320-327. The mathematical foundations for the method were first developed by Jennrich and Sampson (1966). Jennrich and Sampson discuss only the simplest case of the DIRECT OBLIMIN method, but the mathematical formulas for the general case are readily derived from their paper.

DIRECT OBLIMIN is a family of oblique rotations which is controlled by a parameter which is called "delta" by Harman and GAMMA by Jennrich and Sampson. We will refer to the control parameter as GAMMA; in any event our GAMMA and Harman's "delta" are one and the same. You will be given the option of entering a value of GAMMA; if you wish to enter GAMMA, we strongly recommend that you enter ZERO or A NEGATIVE VALUE ONLY. Positive values of GAMMA may produce meaningless results. If you wish to enter positive values of GAMMA despite this warning, then you may do so. In general, the most oblique factors are obtained when GAMMA is zero. As GAMMA approaches minus infinity, the factors tend toward being orthogonal (see Harman, p. 322).

The general operating instructions for LIZOBLI.MIN are the same as for the orthogonal rotation programs of 4. There is a difference in the output; LIZOBLI.MIN outputs both a factor pattern matrix and a factor structure matrix; for orthogonal rotation these matrices are identical.

OUTPUT:

PRINTER:

The output includes

(1) the joint and direct contribution to variance by factor,
(2) contribution to variance by variable,
(3) total variance,
(4) factor pattern matrix,
(5) transform matrix,
(6) factor structure matrix,
(7) correlation among factors.

DISKETTE: The files "FACTOR.PAT" (Factor Pattern Matrix) and "STRUCT.MAT" (Factor Structure Matrix) are created. (See the instructions for program LIZSCORE.FAC below.)

6. FACTOR SCORE COEFFICIENTS AND FACTOR SCORES
Program LIZSCORE.FAC, Diskette 4

This program computes the matrix of Factor Score Coefficients and the Factor Scores for individual cases.

Before computing Factor Scores, the matrix of Factor Score Coefficients must be computed. The input data file and method of computing the factor scores will depend on the main menu option which is chosen. These main menu options are discussed below.

OPTION 1: SCORE COEFFS - PRIN. FACTORS. If this option is chosen, then it is assumed that the Factor pattern matrix was computed using one of the othogonal rotation methods (see # 4 above) after estimating the factors using LIZFACTR.ANL. In this case, the Factor Score Coefficients must be estimated (see Harman, Chapter 16, Section 4). The input file is the file COMPON.ROT created by one of the orthogonal rotation programs.

OPTION 2: SCORE COEFFS - PRIN. FACTORS. If this option is chosen, then it is assumed that the factor pattern matrix was computed by rotating principal components obtained by program LIZCOMP.PCL using one of the methods of ORTHOGONAL rotation. In this case, the Factor Score Coefficients are computed directly (see Harman, Chapter 16, Section 4). The input data file is COMPON.ROT which is created by one of the orthogonal rotation programs.

OPTION 3: SCORE COEFFS. - PRIN. COMP. If this option is chosen, then the Factor Score Coefficients are computed directly from the UNROTATED principal components. The input data file is COMPON.PCL which is created by program LIZCOMP.PCL.

OPTION 4: SCORE COEFFS. - FROM OBLIMIN. If this option is chosen, then it is assumed that the Factor Pattern matrix was computed by an oblique rotation using program LIZOBLI.MIN. You must answer the query "IS THIS RUN FROM ROTATION OF PRINCIPAL COMPONENTS (Y/N)?"

(a) If you answer "Y" then it is assumed that the factor pattern matrix was obtained by rotating the principal components obtained from program LIZCOMP.PCL and the factor score coefficients are computed directly

(b) If you answer "N" then it is assumed that the factor pattern matrix was obtained by rotating the estimated principal factors obtained from program LIZFACTR.ANL. In this case, the Factor Score Coefficients must be estimated.

OUTPUT:

PRINTER: The output consists of

(1) the list of variables
(2) the matrix of Factor Score Coefficients.

After exercising any one of options 1 - 4, the Factor Scores for individual cases may be computed by choosing

OPTION 5: FACTOR SCORES. The input data file for this option is the raw data file which was originally used as input for LIZCORR.MAT. FACTOR SCORES ARE STANDARDIZED, i.e., for each variable in each case, the Z-score is computed prior to computing the Factor Score (see Harman, Chapter 16).

OUTPUT:

PRINTER: For each case the output consists of

(1) case number,
(2) principal factor number, and
(3) factor score.

DISKETTE: The program creates a one-column AGGREGATE file "FACTOR.SCR" which consists of the factor scores for each case in the original file. This file may be merged with the original raw data file using program MERGFILE.AFM.

MISSING DATA: If any column of the raw data file which is designated as containing one of the variables contains a missing data symbol (999999), then the factor score for that case will be output as a missing data symbol. This convention insures that the file FACTOR.SCR will contain the same number of cases as the original data file.

PRACTICE DATA FILE: A practice raw data file is provided for the programs described in this chapter. It is a CUSTOM regression analysis data file containing the famous 5 socioeconomic variables of Harman. The name of this file is SOCIOEC.VAR. (It may be regarded as a 5-column AGGREGATE file.) This file may be used as the raw data input file for programs LIZCORR.MAT and LIZSCORE.FAC.

When using this file, and system files generated from it, always answer the query

NUMBER OF VARIABLES

with

5 <ENTER>.

The suggested names of the variables are

VARIABLE NUMBER	NAME
1	TPOP
2	MSY
3	TEMP
4	MPS
5	MVH

(See Harman, p. 14, Table 2.1.)

The correlation matrix generated by this data has two eigenvalues greater than or at least equal to 1. All programs past LIZCOMP.PCL will automatically use two factors.

REFERENCES

1. H. H. Harman, Modern Factor Analysis, 3rd Ed., The University of Chicago Press, Chicago, 1976.

2. R. I. Jennrich, "Orthogonal Rotation Algorithms," Psychometrika, Vol. 34 (1970), pp. 229-235.

3. R, I. Jennrich and P. F. Sampson, "Rotation for Simple Loadings," Psychometrika, Vol. 31 (1966), pp. 313-323.

4. N. H. Nie, et.al. SPSS -- A Statistical Package for the Social Sciences, McGraw-Hill, New York, 1975.

CHAPTER XII
DISCRIMINANT ANALYSIS, CANONICAL CORRELATION,
AND OTHER MULTIVARIABLE METHODS

The programs described in this chapter are designed to perform several of the procedures related to Multivariate Data Analysis. The mathematical foundations for these programs are found in the book Multivariate Data Analysis by Cooley and Lohnes (see the references at the end of this chapter).

1. MANOVA AND DATA FILE PREPARATION
Program LIZANOVA.MUL, Diskette 5

This program performs the following functions:
(1) It prepares a correlation matrix from a raw data file,
(2) it performs a one-way multiple analysis of variance (MANOVA),
(3) it prepares several diskette data files which are used by other programs which are discussed below.

The procedures followed by this program are described in Cooley and Lohnes, Chapter 8.

The input data file is an AGGREGATE file. Since a one-way MANOVA compares the centroids and dispersion matrix for a set of dependent variables for two or more groups (just as a one-way ANOVA compares the means for a dependent variable for two or more groups), the data file must contain:
(1) the observations for a set of variables. If there are K variables involved in the analysis, where K > 1, then the observations of these variables occupy K columns in the AGGREGATE file (see Chapter III for instructions for the preparation of AGGREGATE files). These K variables are used to prepare a K X K correlation matrix, and are also used for the other computations involved in the MANOVA.
(2) The AGGREGATE file must contain one column which contains an integral independent variable that identifies the group to which a given case belongs. The values assumed by this variable must range from a LOW VALUE to a HIGH VALUE. For example, if the LOW VALUE is 3 and the HIGH VALUE is 7, the values the independent variable takes on must be

3, 4, 5, 6, 7

All cases which have a value of 3 for the independent variable will be assigned to group 1; cases with a value of 4 for the independent variable will be assigned to group 2, etc.

There must be no gaps between the LOW VALUE and the HIGH VALUE for the independent variable, e.g., values of 3, 4, 6, 7 are not acceptable.

If any case has a value for the independent variable which is less than the LOW VALUE or greater than the HIGH VALUE, then that case will be ignored.

The input data diskette may be placed on any drive (if the drive is different from the default drive, be sure to include the drive number in the data file name). The program outputs several data files for use with other programs. These files are created on the default drive. The diskette on the default drive should not contain a file with the same name as a file to be created (see the OUTPUT section below).

The first option chosen from the MAIN MENU is OPTION 2: MANOVA, MATRIX PREPARATION

You must provide the following information:
(1) The number of variables (this refers to the number of DEPENDENT variables only).
(2) The number of groups.
(3) The name of the data file.
(4) The total number of columns in the AGGREGATE file.
(5) Data file column number for the independent variable.
(6) The LOW VALUE for the independent variable.
(7) The HIGH VALUE for the independent variable.
(8) The file column number for each (dependent) variable.
(9) When asked to do so, you must enter a name for each dependent variable (not more than 5 characters).

OUTPUT:

PRINTER

The output from this program consists of
(1) the correlation matrix,
(2) the "total" matrix (dispersion matrix or covariance matrix),
(3) the "among" matrix,
(4) the "within" matrix,
(5) the inverse of the "within" matrix,
(6) Wilk's Lambda,
(7) F-ratio for MANOVA test for the equality of Centroids,
(8) Univariate F-ratios; among MSS, within MSS, F-ratios, and Eta square,
(9) Group Centroids, and
(10) Grand Centroid.

DISKETTE: The following files are created:

CORRMAT.RIX	CENTROID.MAT
TOTAL.MAT	MEANVAR.DAT
AMONG.MAT	VARIABLE.MAT
WITHIN.MAT	GROUP.OBS
INVERSE.MAT	

These files are used by many of the programs described below.

After control returns to the main menu, you may exercise OPTION 3: BOX STATISTIC. This program will then compute and print the Box Statistic, F-ratio for test of the equality of the dispersion matrices of the groups, and the degrees of freedom for the F-ratio (see Cooley and Lohnes, p. 229).

MISSING DATA: If the missing data symbol, 999999, appears in any column which contains a dependent variable, then the corresponding case is ignored. No check is made for missing data symbols in the independent variable.

2. DISCRIMINANT ANALYSIS AND DISCRIMINANT SCORE
Program LIZDISC.ANL, Diskette 5

This program performs a discriminant analysis and, at the user's option, computes the discriminant scores for individual cases. The diskette containing the data files created by LIZANOVA.MUL must be on the default drive at all times when this program is running.

The first main menu option which must be chosen is OPTION 1: DISCRIMINANT ANALYSIS. You must provide the number of variables used in the analysis. The procedures followed by the program are described in Cooley and Lohnes, Chapter 9.

OUTPUT:

PRINTER: The output consists of:

(1) a list of variables used in the analysis,

(2) the eigenvalue of the matrix $W^{-1}A$, where W is the "within matrix", and A is the "among matrix" (these matrices are computed by LIZANOVA.MUL),

(3) Wilk's Lambda,

(4) Tests with successive roots (eigenvalues of $W^{-1}A$) removed). This is a table consisting of:

- (a) the root removed,
- (b) the canonical R,
- (c) the Chi-Square,
- (d) degrees of freedom for Chi-Square,
- (e) Lambda, and
- (f) percent trace.

(5) communalities,

(6) percent contribution (to variance) by discriminant function,

(7) coefficients of discriminant functions (for standardized observations), and

(8) columns of the structure matrix.

When control returns to the main menu, you may elect to compute discriminant scores for individual cases by choosing OPTION 2: DISCRIMINANT SCORES. This option uses the raw data file which was originally used as input to LIZANOVA.MUL. It also uses the diskette data files created by LIZANOVA.MUL.

You must provide:

(1) the name of the raw data file, and

(2) the same information concerning the raw data file as for LIZANOVA.MUL.

OUTPUT:

PRINTER:

The output is a table which consists of

(1) the case number, and

(2) the discriminant scores by principal factor.

DISKETTE:

The file FACTOR.SCR is created. This is a 1-column file containing the discriminant scores for each case. This file may be merged with the raw data file using the files administration program MERGFILE.AFM.

MISSING DATA (raw data file): If any column which is designated as containing one of the dependent variables contains a missing data symbol (999999), then the discriminant score for the corresponding case is output as a missing data symbol.

3. GEISSER CLASSIFICATION
Program LIZCLASS.GEI, Diskette 5

This program classifies individual cases into groups according to a Bayesian method developed by Geisser. The appropriate formulas and the mathematical development for this method are to be found in Cooley and Lohnes, Chapter 10.

This program requires the diskette of data files which was prepared by LIZANOVA.MUL. In addition, it requires the original raw data file which was input to LIZANOVA.MUL. You must provide the same information concerning the data file as in LIZANOVA.MUL and, in addition, you must provide:

(1) the number of groups,
(2) the number of variables, and
(3) the number of subjects (cases) in the raw data file.

This program allows the user to enter prior probabilities if desired. If you answer the query "ARE PRIOR PROBABILITIES TO BE USED?" by NO <ENTER>, then the probabilities for each of the groups will be based on the number of groups. If you answer the above query with YES <ENTER>, then you will have the following options:

(1) PRIOR PROBABILITY BASED ON GROUP SIZE: prior probabilities are automatically computed by the program based on the size of the group relative to the overall number of cases.

(2) USER ENTERED PRIOR PROBABILITIES: If this option is chosen, then you must enter the prior probabilities for each group as directed by the program.

OUTPUT:

PRINTER: The output consists of a table which contains for each subject (case):

(a) the case number,
(b) the actual group number,
(c) the assigned group number, and
(d) the probability that the subject belongs to the group to which the subject is assigned.

In addition, the program prints a table of actual and assigned groups which presents for each group the number of its cases actually assigned to that group, and the number of cases in the group which are assigned to each of the other group (this table is in the form of a matrix). See Cooley and Lohnes, p. 273.

Also printed is the total number of hits, the total number of misses, and the percent effectiveness of the classification.

MISSING DATA: Cases which contain a missing data symbol in any one of the variables will be ignored.

4. MULTIPLE ANALYSIS OF COVARIANCE (MANCOVA)
Program LIZANCOV.ML1, Diskette 5

This program performs a one-way multple analysis of covariance. The mathematical foundation for the procedures used are found in Cooley and Lohnes, Chapter 11.

Before running this program, a multiple analysis of variance must be performed with LIZANOVA.MUL. The following conventions must be followed in submitting the raw data file to LIZANOVA.MUL: Suppose the number of co-variates is P_1 and the number of variables is P_2 with P_1 and P_2 both at least equal to 1.

(1) The total number of variables submitted to LIZANOVA.MUL is $P_1 + P_2$.

(2) The first P_1 variables from the raw data file are the covariates.

(3) The second P_2 variables from the raw data file are the variables.

The above arrangement of the co-variate and variables must be taken into account when entering column numbers for the variable if an AGGREGATE file is used, or in preparing the data file if a CUSTOM file is used.

The program LIZANCOV.ML1 uses the diskette of data files prepared by LIZANOVA.MUL as the input file.

You must provide:
(1) the number of groups,
(2) the number of co-variates, and
(3) the number of variates (dependent variables).

OUTPUT:
PRINTER: The output consists of:
(1) the adjusted total matrix,
(2) the adjusted within matrix,
(3) the adjusted among matrix,
(4) Wilk's Lambda,
(5) Rao's F-approximation with the associated degrees of freedom,
(6) a table of univariate F-ratios consisting of:
variable name,
unadjusted F-ratio,
adjusted F-ratio,
degrees of freedom, and
(7) a table containing the adjusted centroids.

5. CANONICAL CORRELATION
Program LIZCANON.COR, Diskette 4

The procedures followed by this program are described in Cooley and Lohnes, Chapter 6.

The input data file for this program is either a CUSTOM regression analysis file or an AGGREGATE file. The variables for a canonical correlation analysis are divided into a LEFT SET and a RIGHT SET. Suppose the left set contains P_1 variables and the right set contains P_2 variables. The following conventions apply:

(1) The total number of variables is $N = P_1 + P_2$.

(2) P_1 MUST BE AT LEAST EQUAL TO P_2. It is of course possible that $P_1 > P_2$.

(3) The first P_1 variables taken from the data file are the variables for the left set. The remaining variables comprise the right set.

OPTION: You may choose to enter a correlation matrix by choosing the main menu option CORR. MATRIX - USER INPUT.

You must provide:

(1) The number of variables for the left set,
(2) The number of variables for the right set,
(3) The name of the data file,
(4) If the run is not from a CUSTOM file you must provide the column number from the AGGREGATE file for each variable. Remember that the variables for the left set must be entered first. If the run is from a CUSTOM file, the P_1 variables for the left set must appear in the first P_1 columns of the file.
(5) After examining the initial output, you will be asked to specify the number of factors to be retained.

OUTPUT:

PRINTER:

(1) the canonical factor coefficients for the left and right sets.
(2) the correlation matrices for the left and right sets.
(3) Wilk's Lambda.
(4) Chi square with degrees of freedom, all factors.
(5) a table of tests with successive roots removed, which gives the
root removed,
Lambda,
Chi square,
degrees of freedom for Chi square, and
the canonical R.
(6) Redundacy for the right and left sets.
(7) A table of Pearson correlation coefficients for all variables giving:
the variable numbers,
the correlation coefficient, and
degrees of freedom and Student's T for test of significance of the correlation coefficient.

DISKETTE:

The diskette files CORRMAT.RIX and MEANVAR.DAT are created. These are for use by the program only.

MISSING DATA: If a missing data symbol occurs in any column containing a variable, then the corresponding case will be ignored.

REFERENCES

1. W. W. Cooley and Lohnes, P. R., Multivariate Data Analysis. Wiley, New York, 1971.

2. B. F. Huitena, The Analysis of Covariance and its Alternatives. Wiley, New York, 1980.

3. D. F. Morrison, Multivariate Statistical Methods, McGraw-Hill, New York, 1967.

4. N. H. Nie, et.al. SPSS: Statistical Package For the Social Sciences, McGraw-Hill, New York, 1975.

CHAPTER XIII
DESCRIPTIVE STATISTICS

Descriptive statistics may be obtained from the CUSTOM file program LIZDESC.DES which is described in Chapter VI, Section 2 of this manual.

In this chapter we will be concerned with the AGGREGATE file program LIZDESC.AGG, Diskette 6. This program computes and prints descriptive statistics for variables from an AGGREGATE file. The description of this program follows.

DESCRIPTIVE STATISTICS OPTION: The program begins with the usual queries including the name of the data file, which may be a multiple diskette file.

You must provide:

(1) the total number of columns (in the AGGREGATE file),
(2) the number of variables for which statistics are to be computed, — 10 maximum —
and
(3) the data (file) column number for each variable.

If the data file is on more than one diskette, after processing each diskette the program will display the query

DOES THIS FILE CONTAIN MORE THAN ONE DISKETTE?

If so, insert the next diskette, which must have the same file name as originally entered into the program, and type

YES<ENTER>

When all files are read, type

NO<ENTER>

when the above query is displayed. The program will print the same statistics as LIZDESC.DES for each of the specified variables (see Chapter VI).

The formulas used by this program are found in Chapter 14 of N. Nie, *SPSS*, McGraw-Hill. The formula for the coefficient of variability is not found in the above reference, but may be found on page 20 of Steel and Torrie, *Principles and Procedures of Statistics*, McGraw-Hill, 1960. This program uses provisional means algorithms to compute moments.

MISSING DATA: Whenever a missing data symbol is encountered, the case is ignored for the variable containing the symbol; it will not affect a variable which has a valid observation in the same case.

Z-SCORES TO DATA FILE OPTION: This option must be exercised (if at all) immediately after control returns to the main menu from the DESCRIPTIVE STATISTICS option.

Each valid observation of each variable specified in the DESCRIPTIVE STATISTICS option is replaced by its Z-score. NOTE THAT, UNLIKE MOST LIZPACK PROGRAMS AND OPTIONS, THIS PROGRAM ALTERS THE INPUT FILE.

CHAPTER XIV
FREQUENCY DISTRIBUTION, CROSSTABULATION, CONTINGENCY ANALYSIS,
AND THE ANALYSIS OF CROSS-CLASSIFIED DATA

One way frequency distributions are performed by program LIZSAMP.FRQ which is described in Chapter VI, Section 1.

Two- and higher-way frequency distributions, cross-tabulations, contingency analysis, and the analysis of cross classified data are performed by program LIZCROSS.TAB (Diskette 6). The description of this program follows.

POSITIONING THE PRINTER: LIZCROSS.TAB will assume that the printer has been positioned so that the print head is over the first line of the page.

This program will print a two-way table or a higher dimensional table as a series of two-dimensional tables. The tables are printed with horizontal and vertical partitions. The horizontal partitions are automatically printed using dashes (--), ASCII descimal code 45. The user is given the option of choosing a character for the vertical partitions. The character normally used is an I, ASCII decimal code 73. If there is a symbol on your printer which you believe would look better, you may answer the query

    DO YOU WISH TO ENTER CODE FOR A CHARACTER TO BE USED
        FOR VERTICAL PARTITIONS?

with

    YES<ENTER>

You will then be prompted to enter the decimal code for the desired character. If you answer

    NO<ENTER>

the program will use the symbol I.

You must then respond to the standard LIZPACK queries, including the name of the data file.

MULTIPLE DISKETTE DATA FILES: If you answer the query

    IS THIS A MULTIPLE DISKETTE FILE?

with

    YES<ENTER>,

then LIZCROSS.TAB will assume that the data file is on more than one diskette. The following conventions apply:

(1) After the program reads the first diskette (this will not occur until after additional queries have been answered), you will be prompted to

    INSERT NEXT DISKETTE

Insert the next diskette in the appropriate drive (all diskettes must be read from the same drive, and the portions of the data file on each diskette must have the same file name [filespec]). Then type

    YES<ENTER>.

This process is repeated until all diskettes for the data file are read. After the last diskette is read, the prompt

INSERT NEXT DISKETTE

will again appear. Now insert the first diskette of the data file in the drive, and type

NO<ENTER>.

The program is now informed that all diskettes for the data file have been read, and it will proceed to process and print a two-way table.

(2) LIZCROSS.TAB must read the entire data file for each two-way table it prints. There is insufficient memory available in most microcomputers to permit storage of large multi-dimensional tables.

If the data file is on one diskette, simply answer

NO<ENTER>

to the multiple diskette data file query, and no further action with regard to the insertion of disks is necessary.

The following information must be provided:

(1) The number of columns/case in the data file, and

(2) the total number of variables to be cross-tabulated. This must be at least two.

LIZCROSS.TAB divides the variables into the row variable, the column variable, and the control variables.

You must provide:

(1) the data column number (from the data file) for the row variable,

(2) the name of the row variable, and

(3) any observation of the row variable. This "seeds" the routine which finds the maximum and minimum of the row variable. If you plan to enter the maximum-minimum of the row variable, you may enter 0 for this observation. Note that the program will eventually ask you to enter an observation for each variable. These observations may come from different cases as long as they are valid observations (not a missing data symbol) for the variable in question.

(4) You must provide the same information for the column variable as for the row variable.

(5) You must provide the same information for each remaining variable (control variable) as was provided for the row variable.

(6) You must now respond to the query DO YOU WISH TO ENTER MAX-MIN? If you answer

YES<ENTER>

then you will be asked to enter the minimum and maximum value for each variable (row, column, and control, as applicable). If you answer

NO<ENTER>

the program will read the data file and find the extremes for each variable. Note that in this case, the conventions for a multiple diskette data file must be followed.

(7) The program will now display the maximum, minimum, and range (range = maximum - minimum) for each variable. You must enter the number of class intervals, or levels, or categories, you desire for each variable; this

must be at least two. This information will be used by LIZCROSS.TAB to compute the length of each class interval. The formula for computation is

LENGTH OF CLASS INTERVAL = RANGE/(NUMBER OF CLASS INTERVALS - 1)

The program will set up the class intervals so that the center of the first class interval is at the minimum observation for the variable, and the center of the last class interval is at the maximum observation of the variable. An observation will fall in a given inteval if the observation is at least equal to the left endpoint and is less than the right endpoint.

If you enter the maximum and minimum, then any observation which does not fall within a class inteval, together with the corresponding case, will be ignored.

EXAMPLE: Suppose the observations for a variable are 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10. If you specify 10 class intervals for the variable, the class intervals will be

[0.5, 1.5]
[1.5, 2.5]
. . .
[9.5, 10.5]

EXAMPLE: Suppose the minimum value of a variable is 18.5 and the maximum is 36.5. If you specify 19 class intervals for the variable, the class intervals will be

[18, 19]
[19, 20]
. . .
[36, 37]

MISSING DATA: If a missing data symbol occurs in any data file column which contains a variable (row, column, or control), then the entire case will be ignored.

PRINTOUT OF TABLES: The size of each two-way table is determined by
(1) The number (R) of class intervals or levels or categories of the row variable, and
(2) the number (C) of class intervals or levels or categories of the column variable.

The dimension of each two-way table is R x C.

If there are no control variables, then only one two way table will be printed. If there are K control variables (K >= 1) and the i^{th} control variable has V_i class intervals, the total number of R x C tables printed will be

$$V_1 \times V_2 \times \ldots \times V_k.$$

The product R x C must generally be not more than 2200, (slightly larger two-way tables may be permissible).

The number of control variables will affect the amount of memory available to some extent, but there is no inherent limitation on the number of control variables or the number of class intervals for each control variable. However, there are practical considerations. The data file must be read once for each two-way table printed, and a large number of control variables and class intervals within each control variable may result in mountainous printer output.

The printing of a single two-way table may require several pages. In this case the table will be printed as follows:

The table is printed on successive pages until all rows and up to 11 columns have been printed. If there are more than 11 columns, the table is continued on succeeding pages until all rows and up to 11 more columns are printed. This process continues until the entire table is printed. Each page of the table will be numbered.

If some row, some column, or some table contains all zeroes, then that row, column, or table will not be printed.

FORMAT OF TABLES: The main body of the tables are printed by rows and columns. For each class interval (level) of the row variable and column variable there is a cell. Each cell contains the following information, ordered as listed:

(1) cell frequency,
(2) percent of row total for the cell,
(3) percent of column total for the cell, and
(4) percent of table total for the cell.

In the right margin of each row, the frequency for the corresponding class interval of the row variables is printed. Similar information for each class interval of the column variable is printed at the bottom of the table. Percentages of the table total for row and column levels are also printed. The total frequency for the table appears at the bottom right of the table.

The tables are headed by:

CROSSTABULATION FROM LIZPACK DATA FILE (name of data file)

The names of the row variable and column variable are printed. Then the name of each control variable and the number of its current class interval (level) are printed.

MEASURES OF ASSOCIATION: The data which is used as input to LIZCROSS.TAB may be nominal, ordinal, or even continuous; LIZCROSS.TAB will print the same type of measures of association for each two-way table. It is the responsibility of the user to determine which measures are applicable to the particular problem being considered. The computational formulas for the measures of association, together with their interpretation, are found in H. T. Reynolds, The Analysis of Cross-Classifications, The Free Press, New York, 1977. The measures which are computed and printed are listed below with the first page number in Reynolds where the discussion of the measure begins.

NOMINAL MEASURES

Chi Square	p. 7
Phi Squared	p. 46
Contingency Coefficient	p. 46
Tschuprow's T	p. 47
Lambda (Row Variable Dependent)	p. 48
Lambda (Col. Variable Dependent)	p. 48
Symmetric Lambda	p. 51
Cramer's V	p. 47
Goodman-Kruskal's Tau A	p. 52
Goodman-Kruskal's Tau B	p. 52

ORDINAL MEASURES

Kendall's Tau A	p. 72
Kendall's Tau B	p. 78
Kendall's Tau C	p. 78
Wilson's E	p. 76
Gamma	p. 74
Somer's D (Row var. dep.)	p. 75
Somer's D (Col. var. dep.)	p. 75
Somer's Symmetric D	p. 75
Partial Gamma	p. 103

For quick identification of type, LIZCROSS.TAB prints nominal and ordinal measures in separate columns. If there is enough space, these measures will be printed at the bottom of the page containing the table. Otherwise, they will be printed on the next page. All measures are printed for each two-way table.

NOTE: The design of the tables for the cross-tabulation program in SPSS (referenced below) is generally considered to be the best model available. The design of the tables for LIZCROSS.TAB was greatly influenced by the SPSS model.

REFERENCES

1. N. Nie, et. al. SPSS: Statistical Package for the Social Sciences, McGraw-Hill, New York, 1975.

2. H. T. Reynolds, The Analysis of Cross-Classifications, The Free Press, New York, 1977.

CHAPTER XV
THE LIZPACK NONPARAMETRIC STATISTICS PROGRAMS

Unlike most other LIZPACK statistical programs which read the data file one case at a time, the non-parametric statistical programs must have the entire data set resident in memory. This limits the number of cases which can be processed. However, these programs will still process rather large problems, and nonparametric statistical projects are usually comparatively small.

A reference for this chapter is the book Nonparametric Statistical Methods by Hollander and Wolfe (see complete reference at the end of this chapter).

1. Program LIZSPEAR.KEN, Diskette 6

This program computes the Spearman and Kendall rank order coefficients between each possible pair of a set of variables. For the mathematical basis for the Kendall Coefficients, see Hollander and Wolfe, pp. 185-193, and for the Spearman Coefficients, see N. Nie, et al., pp. 288-292.

You must provide
(1) the name of the data file,
(2) number of variables to be correlated,
(3) whether or not the run is from a CUSTOM file. A CUSTOM file is an AGGREGATE file for which the total number of columns equals the number of variables to be correlated.
(4) If the run is not from a CUSTOM file, you must specify the number of columns in the input AGGREGATE file and the file column number for each variable.
(5) The name of each variable.

MISSING DATA: If any column which contains a variable contains a missing data symbol, then the entire case will be ignored.

OUTPUT: The output consists of two tables:
Table 1 consists of
(1) variable names,
(2) Kendall's K,
(3) Tau, and
(4) standard deviation for each pair of variables.
The program also prints the number of valid observations per variable and the number of variables.

Table 2 consists of
(1) variable names,
(2) Spearman's R,
(3) T statistic for test of significance (this applies to "large" samples), and
(4) degrees of freedom for T.

2. Program LIZWILCO.XON, Diskette 6

This program performs the Wilcoxon signed rank test for one sample of paired observations (see Hollander and Wolfe, pp. 27-33). It also produces Hodges-Lehmann's estimation of the median, pp. 33-35. The Main Menu OPTION 1, WILCOXEN SIGNED RANK TEST, must be performed before selecting OPTION 2, ESTIMATE OF THE MEDIAN. The program is limited to 2000 paired observations. OPTION 2 must use a scratch pad disk file SCRATCH.PAD; you may wish to insert a blank diskette immediately after using OPTION 1 and before exercising OPTION 2. A run using 2000 paired observations will produce a scratch file which will completely fill one diskette.

You must provide
(1) the name of the data file,
(2) data column number for variable 1 (the "X" variable, or first variable in each ordered pair,
(3) data column number for variable 2 (the "Y" variable, or second variable for each ordered pair), and
(4) the total number of columns in the data file.

OUTPUT: The output consists of a table which contains, for each ordered pair,
(1) the case number,
(2) the value of the "X" variable,
(3) the value of the "Y" variable,
(4) the difference X - Y
(5) the absolute value of the difference, and
(6) the rank of the pair.

Next, the program prints
(1) total number of valid observations (see MISSING DATA below),
(2) number of ties on X and Y (the number of tied pairs),
(3) number of ties on absolute differences,
(4) Wilcoxen T-plus statistic,
(5) large sample T-plus,
(6) Wilcoxen T-minus statistic, and
(7) large sample T-minus.
If OPTION 2 is exercised, the program will also print
(8) the estimate of the median of the differences from Walsh averages.

MISSING DATA: If a missing data symbol occurs in either variable of an ordered pair, the pair will be ignored.

3. Program LIZKRUSK.WAL, Diskette 6

This program performs the Kruskal-Wallis one-way ANOVA, (Hollander and Wolfe, p. 115-120) and Jonckheere's test for ordered alternatives (Hollander and Wolfe, pp. 120-124). Both OPTION 1, KRUSKAL-WALLIS TEST and OPTION 2, JONCKHEERE'S TEST are performed in the same manner as the one-way ANOVA of LIZ1WAY.AGG of the CORE PACKAGE. The queries and responses are essentially the same.

You must provide:
(1) the number of samples,
(2) the name of the AGGREGATE file,
(3) the total number of columns in the AGGREGATE file,
(4) file column number for the independent variable. The values of the independent variable serve to identify the sample to which the case belongs. The values of the independent variable must be integers.
(5) the high values and low value of the independent variable. There must be no gaps between the high value and the low value; every value in between must appear in some valid case. Cases with values of the independent variable lying outside the range specified by the high and low value are ignored. Missing value symbols are not permitted (or needed) as values for the independent variable.
(6) the file column number for the dependent variable.

MISSING DATA: If a missing data symbol occurs in the column which contains the dependent variable, the case will be ignored.

OUTPUT:
OPTION 1: For each sample, a table is printed which consists of
(1) the sample number,
(2) the rank (of the sample),
(3) average rank (of samples), and
(4) number of observations (in the sample).
Also, the program prints
(1) the total observations,
(2) the number of tied groups (samples),
(3) uncorrected H-statistic, and
(4) corrected H.
OPTION 2: The output consists of
(1) total number of observations,
(2) Jonckheere's J-statistic,
(3) large sample J, and
(4) number of ties on Mann-Whitney counts.

4. Program LIZ2WAY.FRD, Diskette 6

This program performs Friedman's two-way ANOVA, which is analogous to the one-way randomized block experiment, and Page's test for ordered alternatives. The operation of this program is entirely similar to that of LIZRAND.AGG.

The mathematical foundation of Friedman's test is found on pp. 139-147 of Hollander and Wolfe, and for Page's test on pp. 147-151 of the same book.

You must provide
(1) the number of blocks,
(2) the number of treatments,
(3) name of blocks,
(4) name of treatments,
(5) name of AGGREGATE file,
(6) total number of columns in the AGGREGATE file,
(7) column number for independent variable (row variable) and the high and low values of the independent variable. This serves to identify the treatment to which a case belongs.
(8) The column number for independent variable (column variable) and the high and low values of the independent variable. This serves to identify the block to which a case belongs.

In 7 and 8 above, the values of the independent variable must be integers. There must be no gaps between the high and low values; each integer between the high and low values must appear in some case.
(9) You must provide the file column number for the dependent variable.

OUTPUT: The program prints a table which contains for each treatment
(1) the treatment number,
(2) the rank of the treatment,
(3) the average rank of the treatment, and
(4) the number of observations for the treatment.

The program also prints
(1) the total number of observations,
(2) uncorrected S-statistic,
(3) large sample S-statistic,
(4) degrees of freedom for chi-square test (for large sample S-statistic),
(5) total number of ties,
(6) Page's L-statistic,
(7) large sample L-statistic,
(8) number of treatments, and
(9) number of blocks.

MISSING DATA Missing data symbols are not expected nor checked for in this program.

REFERENCES

1. M. Hollander and D. A. Wolfe, Nonparametric Statistical Methods, John Wiley and Sons, New York, 1973

2. N. H. Nie, et al, SPSS: Statistical Package for the Social Sciences, McGraw-Hill, New York, 1975.

CHAPTER XVI
TIME SERIES ANALYSIS AND FORECASTING

The time series analysis and forecasting package contains a number of programs which implement procedures for the analysis and forecasting of time series. The present version does not contain an implementation of the Box-Jenkins model. If the demand is sufficient, a Box-Jenkins model will be implemented in a future supplement to LIZPACK.

DATA FILES: Because of the nature of time series and the type of computations that are performed by the programs in this package, use of the general AGGREGATE file was not deemed to be desirable. All programs in this package use a two-column AGGREGATE file. Column 1 contains the observations of the independent variable, which is usually time, and which is usually measured at equally spaced intervals. It is imperative that the observations of the independent variable be arranged in order of increasing magnitude. Column 2 contains the observations of the dependent variable of the time series. The observations of the dependent and independent variable are arranged in two-column cases.

For most purposes, the particular values of the equally spaced observations of its independent variable are not important. You may use values such as 1, 2, 3, 4, etc. If you are content with such an arrangement, then you may use the file maker included with the program LIZPLOT.SRS, which automatically inserts the values of the independent variable.

MISSING DATA: The use of missing data symbols in time series analysis is usually not advantageous and is not permitted.

GRAPHICS: Graphics are used extensively. After a graph is drawn, the program will pause to allow you to inspect the graph. Press Key C to continue.

DIAGNOSTIC PROGRAMS: There are a number of forecasting methods which are currently in use. The best method to be used will depend on the time series itself. There are three diagnostic programs, LIZ1TIME.SR1, LIZ2TIME.SR2, and LIZANOVA.SRS which are intended to assist you in identifying the characteristics of a particular time series. No matter what method is used, it is always advisable to study the results very carefully before accepting a forecast.

PERIODS PER UNIT: When analyzing a series with seasonal fluctuations, you will be asked to provide the PERIODS/UNIT. If, for example,

(a) periods are months and units are years, the answer is 12.
(b) periods are days and units are weeks, the answer is 7.

The description of the individual programs in the package are given below.

1. MEAN, VARIANCE, AUTOCORRELATIONS, AND DIFFERENCING
Program LIZ1TIME.SR1, Diskette 7

The program performs the following functions:

(1) It graphically displays the time series.

(2) It computes the mean and variance of the series (the variance is the sum of squares of the terms divided by the number of terms).

(3) It computes the autocorrelations and draws the correlogram (see Kendall, p. 39, p. 70,).

(4) It computes the partial autocorrelations and draws the partial correlogram (see Kendall, p. 78).

(5) It computes differenced series.

You must provide the name of the data file containing the time series. The graph of time series will be drawn, the mean and variance will be printed, and then an OPTIONS menu will appear.

OPTION 1: AUTOCORRELATIONS - CORRELOGRAM. This performs functions 3 and 4 as listed above. You must provide:

(1) name of the source file. This file contains the time series and may contain the original series or a differenced series.

(2) name of autocorrelations file. Partial autocorrelations are estimated (see Box and Jenkins, p. 497).

OUTPUT: This option prints two tables which contain the autocorrelations and partial autocorrelations for each case up to one-third the number of cases, or 50 cases, whichever is smaller.

OPTION 2: DIFFERENCING SERIES. Computes the first difference of a time series. You must provide a name for the data file which is to contain the differenced series and the name of the source file. Note that higher order differences may be obtained by repeatedly differencing a series.

OUTPUT:

PRINTER: The printer output consists of a table containing
(a) the values of the independent variable, and
(b) the values of the differenced series.

DISKETTE: A diskette time series data file is generated; it contains the differenced series.

OPTION 3: DIFFERENCED SERIES -- SEASONAL. This computes a series of differences X(I)-X(I-P) where X(J) is the J^{th} observation of the time series and P is the number of periods/unit. Otherwise, this option is the same as OPTION 2. The series it produces will contain P-1 less observations than the original.

OPTION 4: SHOWPACK option

OPTION 5: KILL FILE. Repeated differencing of series may produce unneeded intermediate files; you may KILL such files using this option.

OPTION 6: END RUN. Control is returned to the Main Menu.

2. PERIODOGRAM, WHITE NOISE TEST
Program LIZ2TIME.SR2, Diskette 7

This diagnostic program performs the following functions:

(1) It computes and draws the graph of the periodogram (see Fuller, p. 275-287). This is done from Main Menu Option 1. This option must be exercised prior to selecting either OPTIONS L or C.

(2) OPTION L: LOG PERIODOGRAM. In this option, the natural logarithm of the periodogram ordinate is plotted versus the independent variable; because of the size of the spectral ordinate, this graph is frequently more revealing than that of (1).

(3) OPTION C: CUMULATIVE PERIODOGRAM. This option computes and graphs the cumulative periodogram and performs a white noise test (see Fuller, p. 285). If the number of observations in the time series is less than 63, then you will be asked to enter the Kolmogorov-Smirnov statistic for an appropriate integer. This statistic may be found in Birnbaum, J. Amer. Stat. Assoc., Vol. 47, pp. 425-441. If you enter 0 for this statistic, only the cumulative periodogram is plotted.

OUTPUT:

OPTION 1:

PRINTER: The output consists of a table consisting of
(1) frequency (angle),
(2) periodogram ordinate,
(3) cumulative ordinate,
(4) "COEFF. A", coefficient of the cosine term, and
(5) "COEFF. B", coefficient of the sine term.
The number of ordinates is also printed.

DISKETTE: The diskette file PERIOD.GM is created for system use.

OPTION C: The program prints a table consisting of the
(1) ordinate number,
(2) ordinate,
(3) interval for white noise test, and
(4) F-ratio for significance of the ordinate.
The degrees of freedom for the F-ratios are also printed.

3. ANOVA FOR TREND AND SEASONAL EFFECTS
Program LIZANOVA.SRS, Diskette 7

This is a one-way randomized block ANOVA especially designed for time series. You must provide
(1) the name of the units,
(2) the name of the periods,
(3) the name of the (dependent) variable,
(4) the name of the data file,
(5) the number of units, and
(6) the number of periods/unit. (NOTE: The total number of observations in the data file must equal the product of the number of units and the number of periods per unit).

OUTPUT: The program prints an ANOVA table consisting of
(1) source,
(2) degrees of freedom (DF),
(3) sum of squares (SS),
(4) mean sum of squares (MSS), and
(5) F-ratio.

The program also prints a table of means for each period and unit and the Grand Mean,
Proportion of variability due to trend effect (percent),
Proportion of variability due to seasonal effect, and
Proportion of variability due to the residual.
(See Levenbach and Cleary.)

4. MOVING AVERAGES
Program LIZMOVE.AVE, Diskette 7

The Moving Averages program is designed to remove the "trend component" from a time series by performing one or more smoothings and producing a residual time series which contains the "seasonal component" and "random component." The seasonal component may be extracted using LIZMOVE.AVE or via harmonics using LIZTRIG.SRS which is described later. The program can treat additive, multiplicative, or mixed models.

Main menu option 3 is the moving averages option. After selecting this option, the series is displayed graphically. Press Key C and the first OPTIONS MENU appears.

LIMITS: This program will perform a moving average of length up to 61 for a polynomial of degree up to 5.

OPTIONS MENU 1

The options here are mostly concerned with options in which the user enters the weights to be used in the moving average. If you select OPTIONs 1, 2, or 3 below, you will be asked if you wish to enter weights. If you answer NO, then a simple moving average will be performed. If you answer YES, then you will be asked to enter weights as follows.

(1) You must enter the length of the moving average, i.e., the number of weights to be entered or used, by answering the query "# OBSERVATIONS/INTERVAL?" This must be done even if you do not choose to enter weights. The length is usually odd.

(2) Enter weights in order from first to last. This allows you to use smoothings such as those obtained from Spenser's formulas, or other methods described in the literature.

Many sources give weights as integers (see Kendall) and then give the sum of the weights which can be divided into the individual integral weights so that the weights sum to unity. You may enter the integral weights; the summation and division will be performed automatically. The entered weights must not sum to 0.

The OPTIONs 1, 2, and 3 on this menu do not produce forecasted or extrapolated values, so the length of the series will be shortened at both ends.

OPTION 1: DISJOINT INTERVALS

This is not really a moving average. If the specified length of the moving average is N, this option computes the average of the first N observations, the second N observations, the third N observations, etc., using the weights entered. The variance of the indicated observations are also computed.

You must provide the number of observations per interval and the weights as described above.

OUTPUT FOR OPTIONS 1, 2, AND 3:
A table is printed containing
(1) the period (observation of the independent variable),
(2) the average, and
(3) the variance.
A disk data file named AVERAGES.MOV is created. This is a time series data file which contains the smoothed series.
The smoothed series is plotted on the graphics screen.

OPTION 2: CENTERED INTERVALS

This option performs a centered moving average as described by Kendall, p. 38.

You must provide the number of observations per interval and the weights as described above. If you answer the query "IS MODEL MULTIPLICATIVE?" with YES, a multiplicative model is assumed; if you answer NO, an additive model is assumed. OUTPUT is described above.

OPTION 3: ADJACENT INTERVALS

This option performs an ordinary moving average (see Kendall, pp. 29-38).

You must provide the same information as for OPTION 2.

OUTPUT is described above.

OPTION 4: CLEAR SCREEN

This clears the graphics screen. LIZMOVE.AVE will plot the original and smoothed series on the same graph unless this option is exercised before selecting OPTION 1, 2, or 3.

OPTION 5: SHOWPACK

This creates a graphic disk file containing the current graphics screen.

OPTION 8: SEASONAL ADJUSTMENTS, ADD., and
OPTION 9: SEASONAL ADJUSTMENTS, MULT.

This performs an adjustment for seasonal variation. This process used is described in Kendall, pp. 36-63 (we slightly modify the procedure given for multiplicative models).

You must provide
(1) the number of observations used in the original moving average.
(2) you must answer YES or NO to the query RUN FROM CENTERED INTERVALS? (YES means the original moving average was done by OPTION 2.)
OPTION 8 must not be selected until OPTION 2 or 3 has been exercised.

OUTPUT:
The program prints a table containing
(1) period (observation of the independent variable),
(2) unadjusted value of the dependent variable, and
(3) adjusted value of the dependent variable.
The program creates the time series data disk file SEASONAL.DAT which contains the seasonally adjusted time series.
The adjusted time series is plotted on the graphics screen. The screen will be cleared before this plot is made.

OPTION 6: MOVING AVES., COMPUTED WTS.

Control is transferred to a third OPTIONS MENU which is described as OPTIONS MENU 3 below.

OPTION 7: CONTINUE LISTING

Control is transferred to a second OPTIONS MENU which is described as OPTIONS MENU 2 below.

OPTIONS MENU 2

OPTION 1: RESIDUALS, CENTERED INTERVALS, or
OPTION 2: RESIDUALS, ADJACENT INTERVALS

This option must be selected after exercising Options 2 or 3 on the options menu described above. Choose the option corresponding to the type of moving average which was performed.

You must answer the query "IS MODEL ADDITIVE?" with YES or NO.

The program computes the residuals using the original time series and the latest smoothed series; the latest smoothed series is contained in the file AVERAGES.MOV.

OUTPUT:
The program prints a table containing the
(1) period (value of the independent variable), and
(2) residual.
The program creates the time series disk data file RESIDUAL.DAT which may be RENAMEd for input to LIZMOVE.AVE or LIZTRIG.SRS.

OPTION 3: END RUN

Control is transferred to the main menu.

OPTION 4: ANOVA FOR SEASONAL EFFECTS

This option performs a one-way ANOVA for seasonal effects in the time series. You must provide the number of periods per unit.

OUTPUT: The program prints a one-way ANOVA table containing the
(1) source,
(2) degrees of freedom (DF),
(3) sum of squares (SS),
(4) mean sum of squares, and
(5) F-ratios
A second table is printed which contains the mean for each period.

OPTION 5: CONTINUE LISTING

Control is returned to the first OPTIONS MENU.

OPTIONS MENU 3

This menu is concerned with moving average processes in which the weights are computed by the system.

OPTION 1: COMPUTE WEIGHTS

This option must be selected before exercising any other option on the menu.

You must provide

(1) the number of observations per interval. This number is usually odd. If you enter an even number, then you must follow with a moving average of length 2 in order to realign the data. The last procedure will produce a smoothed series with an extra observation at the beginning and the end of the series. These extra observations should be deleted from the file using the EDITOR or MINI-EDITOR.

A moving average of odd length produces a smoothed time series of the same length as the original series.

(2) the degree of the polynomial (see Kendall, Chapter 3).

The program computes the weights which can be used by the other options on this menu.

OPTION 2: MOVING AVERAGES

This option performs a moving average using the weights computed in OPTION 1.

You must specify whether or not the model is multiplicative.

The OUTPUT is the same as for OPTIONS 1, 2, or 3 of OPTIONS MENU 1.

OPTION 3: REPEAT MOVING AVERAGES

This does another moving average on the time series contained in the data file AVERAGES.MOV. The file AVERAGES.MOV may have been created by OPTION 2 above, or by the present OPTION 3.

You must provide a new name for the current contents of the file AVERAGES.MOV. The newly smoothed series then will be written to the file AVERAGES.MOV.

The OUTPUT is the same as for OPTIONS 1, 2, or 3 of OPTIONS MENU 1.

OPTION 4: RESIDUALS

The program computes the residuals. The operation and OUTPUT in this option is the same as OPTION 1 or 2 on OPTIONS MENU 2.

OPTION 5: SEASONAL ADJUSTMENTS

The operation and OUTPUT for this option is the same as for OPTION 8 or 9 of OPTION MENU 1.

OPTION 6: CONTINUE LISTING

Control is returned to OPTION MENU 1.

OPTION 7: KILL DATA FILE

This option allows you to KILL unwanted data files produced by repeated smoothings.

5. TRIGONOMETRIC REGRESSION: REMOVAL OF SEASONAL COMPONENT FROM DETRENDED SERIES VIA HARMONICS

Program LIZTRIG.SRS, DISKETTE 7

This program uses trigonometric regression to remove the seasonal component from a series which

(1) has been detrended using LIZMOVE.AVE or by some other process, or
(2) is without trend.

The method used is described in Kendall, pp. 112-113.

The normal input data file for this program is the file RESIDUAL.DAT which was created by LIZMOVE.AVE. Be sure to RENAME this file before using it for input; program LIZTRIG.SRS creates yet another file named RESIDUAL.DAT which contains the random component of the series.

You must provide the following information:

(1) the name of the data file,
(2) the number of periods per unit, and
(3) answer YES or NO to the query "IS MODEL MULTIPLICATIVE?"

The program first draws the graph of the input time series on the graphics screen, then computes the Fourier representation of the seasonal component as a Fourier polynomial (see Kendall, p. 113). The representation is

$$S(T)=A(0)+\sum_{I=1}^{M} A(I)*COS(V(I)*T)+\sum_{I=1}^{M} B(I)*SIN(V(I)*T)$$

where the A(I) and B(I) are constants computed by the computer and
(1) V(I) = 2*PI*I/P, P the periods/unit,
(2) M = [P/2], the largest integer not more than P/2.
(3) T is the independent variable.

OUTPUT: The program prints a table consisting of the
(1) frequency (I/P, I, P as above),
(2) Coeff. A (coefficient of cosine term), and
(3) Coeff. B (coefficient of sine term).

The program prints a second table consisting of the
(1) period (value of independent variable),
(2) observed Y,
(3) predicted Y (this is S(T(I)) where S(T) is defined as above, and T(I) is the I^{th} value of the independent variable), and
(4) the residual.

A graph of the seasonal component is sent to the graphics screen.

The program creates the disk time series data file RESIDUAL.DAT which contains the residual time series, or random component. This file may be submitted to program LIZ2TIME.SR2 for a white noise test. The seasonal component is written to the time series disk data file FOURIER.DAT.

6. SINGLE EXPONENTIAL SMOOTHING, Program LIZ1FORE.CST
DOUBLE EXPONENTIAL SMOOTHING, Program LIZ2FORE.CST
TRIPLE EXPONENTIAL SMOOTHING, Program LIZ3FORE.CST
All programs are on Diskette 7

Exponential smoothing is a popular method of forecasting because the method is easy to use. As with all forecasting methods, you should be cautious in interpreting the results. Exponential smoothing is not intended for use in forecasting time series with seasonal fluctuations. The textbook by Bowerman and O'Connell (1981, Chapters 3-6) contains an excellent exposition on the use of exponential smoothing. The authors include a fully worked out example for each method.

The method of operation of all three programs is essentially the same. You must provide the
(1) name of the data file, and
(2) number of future forecast periods (this will always be one for LIZ1FORE.CST because for this program, all future forecasts are the same). Do not attempt to forecast too far ahead by these methods. Exponential smoothing is a short-term forecasting method.

The graph of the input series is sent to the graphics screen.

You will be asked

DO YOU WISH TO ENTER ALPHA?

Alpha is the smoothing constant. If you answer YES, you must enter a positive number not more than 1. If you answer NO, the system will compute an "optimal" alpha between .01 and .3. This is the standard recommended range for alpha. The best alpha for a given series may not lie in this range. With a little experience, you may do better by experimenting and choosing your own values of alpha.

If you answered YES to the previous query, you will be asked

DO YOU WISH TO ENTER THE NUMBER OF
OBSERVATIONS TO BE USED TO COMPUTE ALPHA?

There is no "best" method for choosing this number. It seems in general that the number should not be less than 6 nor more than half the length of the series. If you answer NO, then the program will use half the length of the series. For a long series, this number is generally far too large. If you have no better information available, we suggest you enter the number 6 for single or double smoothing, and not more than 12 for triple smoothing. You may wish to make several runs using different numbers each time (but generally not less than 6).

You will now be asked

DO YOU WISH TO ENTER THE NUMBER OF OBSERVATIONS
TO BE USED TO COMPUTE INITIAL ESTIMATES?

This refers to the number of observations used to compute the initial estimates of the smoothed statistics. If you answer NO, the program will use all the observations. This is a standard choice. You may wish to try other choices.

OUTPUT: The program prints a table consisting of the

(1) period (value of the independent variable),
(2) observed Y (value of dependent variable),
(3) smoothed Y, and
(4) residual.

Future forecasts are printed and labeled.

The following statistics are also printed: the smoothing constant, the mean of the series, the residual SS, the total SS (based on the mean of the series), and the number of observations (length of the series). The smoothed values are plotted as small rectangles on the same graph with the original series. The forecasted values are also graphed.

The program creates a time series disk data file RESIDUAL.DAT which contains the residual time series. This file may be used for diagnostic purposes.

7. FORECASTING SEASONAL SERIES WITH POLYNOMIALS AND DUMMY VARIABLES

Program LIZ4FORE.CST, Diskette 7

This program uses the technique of dummy variables combined with polynomials to fit certain time series which show seasonal fluctuations. An excellent discussion of this method is contained in Bowerman and O'Connell (1979, Chapter 8). We use a different, but equivalent, method of coding the dummy variables than that used in Bowerman and O'Connell.

After selecting the main menu option FORECASTING-SIMULATION, you must provide

(1) name of source file,

(2) name of destination file (this file is a regression analysis file which may be used either by this program or by one of the regression analysis modeling programs of the CORE PACKAGE),

(3) degree of the polynomial (this is usually 1 or 2, but at your option, may be more),

(4) number of periods/unit,

(5) number of forecast periods,

(6) whether or not the model is additive, and

If you answer the query "CREATE DATA FILE ONLY?" with NO, this program performs the least squares curve fit. Otherwise, answer YES and it creates the destination file only. This data file can then be used as input for one of the modeling programs of the CORE PACKAGE. The description of the destination file follows.

If the degree of the polynomial is to be M and the number of periods per unit is P, then the destination file will be a CUSTOM regression analysis file which will contain

(1) M + P variables for an additive model, and

(2) M + 2P - 1 variables for a multiplicative model.

To describe the variables in a given case, let T be the value of the independent variable for the case.

ADDITIVE MODEL

(1) The first M variables are

$$T, T^2, T^3, \ldots, T^M$$

(2) The next P - 1 variables are D(1), D(2), . . . , D(P-1) where

(a) D(I) = 0 if the case is not observed in period I

(b) D(I) = 1 if the case is observed in period I.

(c) D(I) = -1 if the case is observed in period P.

(3) The last variable is the dependent variable.

MULTIPLICATIVE MODEL

(1) The first M + P - 1 are the same as for the additive model.

(2) The next P-1 variables are

T*D(1), T*D(2), . . . , T*(D(P-1).

(3) The last variable is the dependent variable.

OUTPUT: The program prints an ANOVA table for regression. The table is entirely similar to those printed by the regression programs of the CORE PACKAGE.

The partial regression coefficients are printed. The variables corresponding to polynomial terms and dummy variables are identified. For a multiplicative model, the "dummy variables" actually break down as follows: if P is the number of periods per unit, then

(1) the first P-1 variables indicated as dummy variables are the coefficients of the dummy variables D(I), and

(2) the next P-1 variables indicated as dummy variables are actually the variables T*D(I), where T is the independent variable.

The constant term is printed.

A table of means and standard deviations is printed.

The standard error of estimate is printed.

A forecasting simulation table is printed containing the

(1) period,
(2) observed Y,
(3) forecasted Y, and
(4) residual.

The graph of the original series and the forecasted values are sent to the graphics screen.

Two disk data files are created:

(1) the destination file, which is described above, and
(2) the file RESIDUAL.DAT, containing the residual time series.

8. FORECASTING SEASONAL SERIES WITH POLYNOMIALS AND FOURIER POLYNOMIALS
Program LIZTRIG.PLY, Diskette 7

This program produces a data file which may be used as input for either of the multiple regression analysis modeling programs LIZANAL.MUL or LIZANAL.NAM of the CORE PACKAGE. The purpose is to fit a curve of the form

$$\text{(I)} \quad Y=P(T)+F(T)+T*G(T)$$

where P(T) is a polynomial and F(T) and G(T) are Fourier polynomials. For our purposes, a Fourier polynomial of degree M is a function of the form

$$\sum_{I=1}^{M} (A(I)*\cos(V(I)*T)+B(I)*\sin(V(I)*T))$$

where

(1) the A(I) and B(I) are constants, and
(2) V(I) = 2*PI*I/P where P is the number of periods.

The data file produced by this program is a CUSTOM regression analysis data file which may be described as follows:

Let M be the degree of polynomial P(T) of (I), and let P and Q be the degrees of the polynomials F(T) and G(T), respectively. Then the data file contains M+2P+2Q+1 variables.

The arrangement of the variables for a particular case is now described. Let T be the value of the independent variable for the case.

(1) The first M variables are powers of T: T, T^2, ..., T^M
(2) The next P variables are cos(V(1)*T), cos(V(2)*T), ..., cos(V(P)*T).
(3) The next P variables are sin(V(1)*T), sin(V(2)*T), ..., sin(V(P)*T).
(4) The next Q variables are T*cos(V(1))*t), ..., T*cos(V(Q)*T).
(5) The next Q variables are T*sin(V(1)*T), ..., T*sin(V(Q)*T).
(6) The last variable is the dependent variable.

You must provide the
(1) name of the source file,
(2) name of the destination file,
(3) degree of polynomial (this is the degree of P(T) in (I)). If you enter 0, the term P(T) will not appear.
(4) number of periods per unit,
(5) degree of Fourier polynomial (this is the degree of F(T) in (I)). If you enter 0, the term F(T) will not appear.
(6) degree of the mixed polynomial (this is the degree of the Fourier polynomial G(T) in (I)). If you enter 0, the polynomial G(T) will not appear.

The file will be produced and the number of variables in the destination file will be displayed on the screen. Record this number and press any alphanumeric key. LIZPACK DIRECTORY will reappear.

9. AUTOREGRESSION AND SIMULATION
Program LIZCORR.LAG, Diskette 7

The operation and output of this program are entirely similar to that of LIZANAL.MUL of the CORE PACKAGE. You should read the instructions found in Chapter VI, Section 15 before using LIZCORR.LAG. In the present program, variables are referred to as LAGS. This program produces a forecast equation which yields the observation X(T) at time T, of the dependent variable in terms of lagged values of the dependent variable (see Kendall, pp. 142-150 and Granger and Newbold, pp. 176-179). For modeling purposes, LIZCORR.LAG should be used exactly like LIZANAL.MUL (see Chapter VIII, "Performing Regression Analysis with LIZPACK").

When performing a regression analysis run with LIZCORR.LAG, the partial regression coefficient for each lagged variable is printed with its lag number. Remember that if X(T) is the present observation, LAG 1 refers to X(T-1), LAG 2 refers to X(T-2), etc. LAG 0 is the dependent variable, X(T).

The OPTIONS menus for LIZCORR.LAG are quite similar to those of LIZANAL.MUL. There is both a FORWARD MODE and a BACKWARD MODE to allow for forward and backward stepwise autoregression and other manipulations of the LAGS. After selecting either MODE from the main menu, you must provide the ORDER OF THE PROCESS, i.e., the number of lags to be used as independent variables in the autoregression. There is one difference on the BACKWARD MODF menu; the option PARTIAL CORRELATION COEFFICIENT does not appear. In its place is OPTION 1: SIMULATION.

The SIMULATION option must be selected after exercising OPTION 8: REGRESSION RUN.

When OPTION 1 is exercised, a graph of the time series appears on the graphics screen and the "forecasted" values, the values of the dependent variable computed from the autoregression equation, are plotted on the same graph as small rectangles.

10. PRELIMINARY ESTIMATES FOR THE PARAMETERS OF GROWTH CURVES
Program LIZNLIN.EST, Diskette 7

This program may be used at your option to produce preliminary estimates for the parameters of certain growth curves, and these estimates may be used as input to the non-linear least squares regression program LIZNON.LIN which is described below.

The growth models with which LIZNLIN.EST and LIZNON.LIN are concerned are

(1) the modified exponential curve

$$y = \alpha(1-\beta\exp(-kt))$$

(2) the logistic curve

$$y = \alpha/(1+\beta\exp(-kt)$$

(3) the Gompertz curve

$$y = \alpha\exp(-\beta\exp(-kt)).$$

See Draper and Smith, pp. 505-513 for a discussion of these curves.

The non-linear regression techniques which are currently in use may be sensitive to the initial estimates. Thus LIZNLIN.EST attempts to provide a reasonable set of preliminary estimates to be used by LIZNON.LIN. The methods used to obtain these estimates are described in Mills, pp. 554-566. Before the advent of the use of computers in data analysis, these methods were used to fit a time series with a growth curve. Sometimes the preliminary estimates themselves provide a fairly good fit for the time series.

Select the menu option corresponding to the curve you wish to fit. The only information required from you is the name of the data file which contains the time series you wish to fit.

OUTPUT: The output is the inital estimates of the parameters ALPHA, BETA, and K.

11. NON-LINEAR LEAST SQUARES ESTIMATES OF THE PARAMETERS FOR GROWTH CURVES
Program LIZNON.LIN, Diskette 7

Read the description of the program LIZNLIN.EST above.

The purpose of the present program is to estimate the parameters ALPHA, BETA, and K (as described in 11) by the method of non-linear least squares (see Draper and Smith, pp. 462-464.

Select the menu option corresponding to the type of curve you wish to fit.

You must provide the following information.

(1) You must answer the query

COMPUTE RESIDUALS BY SUBTRACTION?

with YES or NO. If you answer NO, then the residuals will be computed by division.

(2) You must answer the query

DO YOU WISH TO ENTER INITIAL ESTIMATES?

with YES or NO. If you enter NO, then the program will compute (very rough) initial estimates by the methods described in Draper and Smith, pp. 510-511. The estimates obtained in this manner may be adequate; on the other hand, they may lead to divergence or overflow. If you answer YES, you must

(a) enter estimates obtained from LIZNLIN.EST or

(b) enter your own estimates.

(3) At the prompt, enter the name of the data file.

The program provides the scattergram and the graph of the fitted curve.

ABOUT CONVERGENCE: After each iteration, the screen will show

(1) the largest absolute difference between two successive parameter estimations,and

(2) the residual sum of squares for the present estimation of the parameters.

Let SS(K) denote the residual sum of squares from the K^{th} iteration. Convergence will occur when

ABS(SS(K)-SS(K-1))/SS(K-1) < .00000001

for 5 successive iterations. You may reduce the convergence criterion or force a printout by proceeding as follows:

(1) press any alphanumeric key except Key P, and hold it down until new numbers appear on the screen (i.e., until an interation has been completed). This will produce a new convergence criterion: Convergence will occur when the absolute difference of (1) is not more than .0000001.

The above process may be repeated as often as desired; each repetition will reduce the convergence criterion by a factor of 10.

(2) Press Key P and hold it down until new numbers appear on the screen. Iteration will stop, and the program will print the results of the last iteration just as if convergence had occurred.

You may have to try more than one type of curve in order to fit your series.

OUTPUT: The program prints
(1) the estimated model parameters, and
(2) a table containing the
  (a) independent variable,
  (b) observed Y,
  (c) computed Y (computed value of the dependent variable),
  (d) the residual, and
  (e) the residual SS.

The program creates the disk data file RESIDUAL.DAT which contains the residual series.

12. PLOTS OF TIME SERIES AND SCATTERGRAMS FROM AGGREGATE FILES
Program LIZPLOT.SRS, Diskette 7

This program allows you to examine the graphs of a time series or a scattergram in which the independent (horizontal) and dependent (vertical) variables are taken from an AGGREGATE file. The independent variable may be taken from one file and the dependent variable may be taken from another. This program does not check for missing data symbols.

The following are descriptions of the main menu options. In all options, the number of observations of the independent variable and the dependent variable must be the same.

OPTION 3: This option provides a graph of a time series with the independent and dependent vaiables taken from the same time series data file. You must provide the name of the data file.

OPTION 4: A time series graph is drawn where the independent variable is taken from one file and the dependent variable is taken from another. You must provide
(1) the name of the 1st file, which contains the independent variable.
(2) the name of the 2nd file, which contains the dependent variable.

Both files must be time series data files.

OPTION 5: This produces scattergrams with both variables taken from the same AGGREGATE file. You must provide
(1) the name of the AGGREGATE file,
(2) the total number of columns in the AGGREGATE file,
(3) the column number for the independent variable, and
(4) the column number for the dependent variable.

OPTION 6: This produces a scattergram with the variables taken from different AGGREGATE files. You must provide:
(1) the name of the 1st data file,
(2) the total number of columns (in the 1st file),
(3) the column number for the independent variable,
(4) the name of the 2nd data file,
(5) the total number of columns in the 2nd file, and
(6) the column number for the dependent variable.

13. OPERATIONS ON TIME SERIES
Program TRANSPAK.SRS, Diskette 7

This program allows you to add, subtract, multiply, or divide the observations of the dependent variable in two time series. The following are examples of how this program may be used:

(1) to compute residuals from a file containing the original series and another file containing the "smoothed" version of the series.

(2) to change forms of a model, e.g., from additive to multiplicative, after a smoothing of some sort has been performed.

You must provide the

(1) name of the first source file,

(2) name of the second file (depending on the option you choose, the observations of the dependent variable in this file will be added to, subtracted from, multiplied into, or divided into the corresponding observations in the first file. The first and second source files must contain the same number of cases), and

(3) the name of the destination file (this file will contain the transformed series).

An OPTIONS menu will appear. Select the operation you wish to perform. The destination file will be created. A new OPTIONS menu will appear which has options to allow you to

(1) perform a new operation with all new files, a new second source file, or a new destination file;

(2) reload the LIZPACK DIRECTORY; or

(3) temporarily END LIZPACK.

14. TRANSFORMATIONS OF TIME SERIES
Program LIZENCO.SRS, Diskette 7

This program will allow you to perform some of the common transformations on the dependent variable of a time series. The transformations are preprogrammed and are selected from a menu.

You may also use the program TRANSPAK.REG of the CORE PACKAGE (see Chapter VI, Section 22); TRANSPAK.REG will perform more transformations than LIZENCO.SRS. More sophisticated transformations may be performed with the main transformation programs TRANSPAK.LPK and LIZENCO.LPK (see Chapter VII).

You must provide

(1) the name of the source file, and

(2) the name of the destination file (this file will contain the transformed series).

When the OPTIONS menu appears, select the proper option. The transformation will be performed.

Since the program initially copies the contents of the source file into the destination file and then performs all transformations on the dependent variable in the destination file, you may perform repeated transformations with this program.

OUTPUT: Destination file (diskette).

PRACTICE DATA FILES

(I) File AIRLINE.DAT (Kendall, p. 12, Table 1.3).
This series consists of 8 years of monthly data (96 cases).
(Kendall, p. 64) Suggested exercises using this file for program LIZMOVE.AVE are:

1. In LIZMOVE.AVE, choose the MOVING AVERAGES option from the main menu, and then OPTION 6 from the first OPTIONS menu.
2. Select OPTION 1: COMPUTE WEIGHTS. Specify
   (a) 15 as the number of observations per interval, and
   (b) 1 as the degree of polynomial.
3. After the weights have been computed, select OPTION 2: MOVING AVERAGES; specify a multiplicative model.
4. When the OPTIONS menu reappears, select OPTION 3: REPEAT MOVING AVERAGES. Give the moving averages file the name TEMP.DAT.
5. When the OPTIONS menu reappears, select OPTION 7, and KILL the file TEMP.DAT. Then select OPTION 4: RESIDUALS.
6. CONTINUE LISTING to obtain OPTIONS MENU 2. EXIT to the MAIN MENU. Press E to temporarily END LIZPACK.
7. RENAME file RESIDUAL.DAT to AIRPASS.MIL. RENAME file AVERAGES.MOV to AIRLINE.TRN.
8. Type RUN<ENTER> to reenter LIZMOVE.AVE. Select the MOVING AVERAGES option from the MAIN MENU. When you are asked to enter the name of the data file, enter AIRPASS.MIL. Then select OPTION 6 from the first OPTIONS MENU.
9. Select OPTION 1: COMPUTE WEIGHTS. Specify
   (a) 7 as the number of observations per interval, and
   (b) 1 as the degree of polynomial.
10. Select OPTION 2 and perform a moving average. Specify a multiplicative model.
11. Select OPTION 1: COMPUTE WEIGHTS. Specify
    (a) 5 as the number of observations per interval, and
    (b) 1 as the degree of polynomial.
12. Select OPTION 3: REPEAT MOVING AVERAGES. Again, give the name TEMP.DAT to the moving averages file.
13. Select OPTION 7 and KILL the file TEMP.DAT. Then select OPTION 4: RESIDUALS.

At this point, you have created 4 data files:

(1) AIRLINES.TRN, which contains the trend component of the series.
(2) AIRPASS.MIL, which contains the seasonal and random components of the series.
(3) AVERAGES.MOV, which contains the seasonal component as computed by LIZMOVE.AVE.
(4) RESIDUAL.DAT, which contains the random component of the series.

Now RENAME file RESIDUAL.DAT to AIRLINE.RES and use this file as input for LIZ2TIME.SR2. Do a periodogram, then a cumulative periodogram, and examine the results of the white noise test.

Use the data file AIRPASS.MIL as input to program LIZTRIG.SRS to remove the seasonal component. This will produce a new file RESIDUAL.DAT which contains the random component; use the last file as input for program LIZ2TIME.SR2. Do a periodogram, then a cumulative periodogram, and again examine the results of the white noise test.

(II) File IMPORTS.DAT (Kendall, p. 144, Table 11.1)

This file consists of 11 years of quarterly data (44 cases).

Use ths file as input for program LIZCORR.LAG. Answer the query
ORDER OF PROCESS?
with
4<ENTER>

See Kendall, pp. 143-145 for a discussion of this series.

(III) Files USWHITE.POP and USBLACK.POP

These files contain census figures for the white and black populations of the United States observed at ten-year intervals. The files are incomplete since there are observations only to up to the year 1970, and the observations for that year are estimated.

These files may be used as input for programs LIZNLIN.EST and LIZNON.LIN. Fit a LOGISTIC CURVE to these series by obtaining preliminary estimates of the parameters from LIZNLIN.EST and then entering these into LIZNON.LIN.

Several other good examples of time series are found in Box and Jenkins beginning on p. 525.

REFERENCES

1. B. L. Bowerman and R. T. O'Connell, Forecasting and Time Series, Duxbury Press, North Scituate, Mass., 1979.

2. G. E. P. Box and G. M. Jenkins, Time Series Analysis, Forecasting, and Control, Holden-Day, San Francisco, 1970.

3. N. R. Draper and H. Smith, Applied Regression Analysis, 2nd Edition, Wiley, New York, 1981.

4. M. Firth, Forecasting Methods in Business and Management, Edward Arnold, London, 1977.

5. W. A. Fuller, Introduction to Statistical Time Series, Wiley, New York, 1976.

6. C. W. J. Granger, Forecasting in Business and Economics, Academic Press, New York, 1980.

7. C. W. J. Granger and P. Newbold, Forecasting Economic Time Series, Academic Press, New York, 1977.

8. M. G. Kendall, Time-Series, Hafner Press, New York, 1973.

9. H. Levenbach and J. Cleary, The Beginning Forecaster, Lifetime Learning Publishers, Belmont, CA, 1981

10. ------, The Professional Forecaster, Lifetime Learning Publishers, Belmont, CA., 1981.

11. F. Mills, Introduction to Statistics, Henry Holt, New York, 1956.

APPENDIX

Sample Output from Selected LIZPACK Programs

APPENDIX

1. Sample listing of an AGGREGATE FILE by the Editor

2. One-way frequency distribution from LIZSAMP.FRQ

3. Descriptive statistics as computed by LIZDESC.DES. The output from LIZDESC.AGG is similar except that the number of missing cases for each variable is also printed.

4. Test for polynomial trend in means -- equal spacing (LIZMEAN.MCP)

5. One-way ANOVA from LIZ1WAY.ONE or LIZ1WAY.AGG

6. One-way ANOVA for randomized block experiment from LIZANCOV.RAN or LIZRAND.AGG

7. ANOVA for Randomized Block Experiment from LIZANOVA.RAN or LIZRAND.AGG

8. Two-way ANOVA with equal cell frequencies; from LIZ2WAY.TWO or LIZ2WAY.AGG

9. Two-way randomized block experiment from LIZBLOCK.2WY or LIZ2BLOCK.AGG

10. Two-way ANCOVA by LIZ2WAY.ANC or LIZ2ANCV.AGG (2 pages)

11. Three-way ANOVA by LIZANOVA.3WY (equal cell frequencies). Data from Steel and Torrie. This is a 3 x 3 x 2 factorial experiment.

12. Printout of correlation matrix and a regression run by LIZCORR.MUL

13. Regression and Residual Analysis from LIZANAL.RES. Data from Steel and Torrie. Note the probability of several outliers. (2 pages)

14. Regression and Residual Analysis from LIZANAL.RES (on the following page). The data used here were introduced by Longley in 1967. Many of the statistical packages of that time could not obtain the estimates of the partial regression coefficients to even one decimal place. The Longley Data are a formidable challenge to any Regression program; however, many modern packages will perform excellently with these data. LIZPACK actually outperforms some mainframe packages on these data. The model is overparameterized, and a sensitivity test indictes that no matter how accurately the estimates are computed, they are not reliable estimates of the true population parameters (see Weisberg, Applied Regression Analysis). This run is from a scaled version of the Longley data.

15. Regression run from LIZEXPO.ONE

16. Regression run from LIZLOG.ANL

17. Regression run from LIZRAD.PWR

18. Polynomial regression by LIZPOLY.STP. Degree of polynomials: 3, 2, 1. (2 pages)

19. Computation of residuals from LIZPOLY.RES

20. Principal Component Analsysis from LIZCOMP.PCL

21. Oblique rotation of two principal factors by LIZOBLI.MIN

22. One-way MANOVA by LIZANOVA.MUL; two variables, three groups.

23. An example of a small discriminant analysis problem (2 x 2) from LIZDISC.ANL. Only one eigenvalue is retained in the analysis.

24. Crosstabulation; multipage printout. (4 pages)

25. Autoregression and Simulation from LIZCORR.LAG

LISTING OF LIZPACK DATA FILE CEMENT/AGG
NUMBER OF COLUMNS: 6
TOTAL NUMBER OF RECORDS: 78

COLUMN 1	LABEL: CASE NUMBER
COLUMN 2	LABEL: TCA
COLUMN 3	LABEL: TCS
COLUMN 4	LABEL: TCAF
COLUMN 5	LABEL: DCS
COLUMN 6	LABEL: HEAT

RECORD #,RECORD		RECORD #,RECORD		RECORD #,RECORD		RECORD #,RECORD		RECORD #,RECORD		RECORD #,RECORD	
1	1	2	7	3	26	4	6	5	60	6	78.5
7	2	8	1	9	29	10	15	11	52	12	74.3
13	3	14	11	15	56	16	8	17	20	18	104.3
19	4	20	11	21	31	22	8	23	47	24	87.6
25	5	26	7	27	52	28	6	29	33	30	95.9
31	6	32	11	33	55	34	9	35	22	36	109.2
37	7	38	3	39	71	40	17	41	6	42	102.7
43	8	44	1	45	31	46	22	47	44	48	72.5
49	9	50	2	51	54	52	18	53	22	54	93.1
55	10	56	21	57	47	58	4	59	26	60	115.9
61	11	62	1	63	40	64	23	65	34	66	83.8
67	12	68	11	69	66	70	9	71	12	72	113.3
73	13	74	10	75	68	76	8	77	12	78	109.4

1. Sample listing of an AGGREGATE FILE by the Editor

(Reduced size copy)

NAME OF STUDY: WEEKLY EARNINGS OF TEXTILE WORKERS
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 3 AUGUST 1983
RUN FROM LIZPACK DATA FILE: TEXTILE/DAT

TABLE FOR FREQUENCY DISTRIBUTION

CLASS	ENDPOINT A	ENDPOINT B	FREQUENCY
1	38	39	1
2	39	40	1
3	40	41	2
4	41	42	2
5	42	43	3
6	43	44	3
7	44	45	8
8	45	46	14
9	46	47	19
10	47	48	15
11	48	49	20
12	49	50	27
13	50	51	28
14	51	52	20
15	52	53	14
16	53	54	8
17	54	55	6
18	55	56	7
19	56	57	3
20	57	58	4
21	58	59	3
22	59	60	3
23	60	61	2
24	61	62	2
25	62	63	1
26	63	64	1
27	64	65	2
28	65	66	0
29	66	67	0
30	67	68	1

CASES	MEAN	VARIANCE	STD DEVIATION
220	50.1561364	21.5359286	4.64068192

2. One-way frequency distribution from LIZSAMP.FRQ

(Reduced size copy)

NAME OF STUDY: POTENCY OF CARDIAC ARREST SUBSTANCES
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 3 AUGUST 1983
RUN FROM LIZPACK DATA FILE: DOSAGE/GMS

SAMPLE 1 , 10 CASES

MINIMUM	MAXIMUM	RANGE	MEAN	VARIANCE
19	29	10	25.9	9.43333334
STANDARD DEV.	COEFF. OF VAR.	STD. ERROR	SKEWNESS	KURTOSIS
3.0713732	11.8585838	.971253486	-.950944911	-.121293439

SAMPLE 2 , 10 CASES

MINIMUM	MAXIMUM	RANGE	MEAN	VARIANCE
17	28	11	22.2	12.1777778
STANDARD DEV.	COEFF. OF VAR.	STD. ERROR	SKEWNESS	KURTOSIS
3.48966729	15.719222	1.10352969	.0756770121	-1.48821434

SAMPLE 3 , 10 CASES

MINIMUM	MAXIMUM	RANGE	MEAN	VARIANCE
16	25	9	20	8.66666667
STANDARD DEV.	COEFF. OF VAR.	STD. ERROR	SKEWNESS	KURTOSIS
2.94392029	14.7196015	.930949337	.141099138	-1.45828403

SAMPLE 4 , 10 CASES

MINIMUM	MAXIMUM	RANGE	MEAN	VARIANCE
16	25	9	19.6	8.71111112
STANDARD DEV.	COEFF. OF VAR.	STD. ERROR	SKEWNESS	KURTOSIS
2.95145915	15.0584651	.933333333	.609556677	-1.03483913

3. Descriptive statistics as computed by LIZDESC.DES. The output from LIZDESC.AGG is similar except that the number of missing cases for each variable is also printed.

(Reduced size copy)

NAME OF STUDY: P
NAME OF INVESTIGATOR: I.Q. SMART
DATE OF RUN: 4 AUGUST 1983
RUN FROM LIZPACK DATA FILE: DOSAGE/GMS

ANOVA TABLE FOR POTENCY DATA

SOURCE	DF	SS	MSS	F RATIO
BETWEEN (SUBST)	3	249.875008	83.2916693	8.54516998
CONTR 1	1	222.605	222.605	22.8377889
CONTR 2	1	27.225	27.225	2.7931035
CONTR 3	1	.0450000092	.0450000092	4.61670094E-03
REM	0	7.74860382E-06		
WITHIN (ERROR)	36	350.899994	9.74722206	
TOTAL	39	600.775002		

4. Test for polynomial trend in means -- equal spacing (LIZMEAN.MCP)

(Reduced size copy)

NAME OF STUDY: POTENCY OF CARDIAC ARREST SUBSTANCES
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 3 AUGUST 1983
RUN FROM LIZPACK DATA FILE: DOSAGE/GMS

ANOVA TABLE FOR POTENCY DATA

SOURCE	DF	SS	MSS	F RATIO
BETWEEN (SUBST.)	3	249.875008	83.2916693	8.54516998
WITHIN (ERROR)	36	350.899994	9.74722206	
TOTAL	39	600.775002		

UNIT	MEAN
SUBST. 1	25.9
SUBST. 2	22.2
SUBST. 3	20
SUBST. 4	19.6
GRAND MEAN	21.925

5. One-way ANOVA from LIZ1WAY.ONE or LIZ1WAY.AGG

(Reduced size copy)

NAME OF STUDY: VITAMIN C CONTENT OF LIMA BEAN VARIETIES
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 4 AUGUST 1983
RUN FROM LIZPACK DATA FILE: LIMAS/DAT

ANOVA TABLE FOR DRYWT DATA

SOURCE	DF	SS	MSS	F RATIO
VARIETY	10	2166.71303	216.671303	22.7086491
FIELD	4	367.85376	91.96344	9.6384037
ERROR	40	381.654236	9.5413559	
TOTAL	54	2916.22102		

FACTOR	MEAN	FACTOR	MEAN
VARIETY 1	35.42	FIELD 1	31.7090909
VARIETY 2	47.76	FIELD 2	31.1
VARIETY 3	36.6	FIELD 3	33.9818182
VARIETY 4	39.38	FIELD 4	34.7636364
VARIETY 5	24.48	FIELD 5	38.3818182
VARIETY 6	30.72		
VARIETY 7	34.9		
VARIETY 8	34.34		
VARIETY 9	32.84		
VARIETY 10	23.88		
VARIETY 11	33.54		

GRAND MEAN 33.9872728 STD DEVIATION 7.34874772

ANOVA TABLE FOR VITC DATA

SOURCE	DF	SS	MSS	F RATIO
VARIETY	10	51018.1782	5101.81782	34.3134844
FIELD	4	4968.94031	1242.23508	8.35494629
ERROR	40	5947.30371	148.682593	
TOTAL	54	61934.4223		

6. One-way ANOVA for randomized block experiment from LIZANCOV.RAN or LIZRAND.AGG

(page 1 of 2 pages)

(Reduced size copy)

FACTOR	MEAN	FACTOR	MEAN
VARIETY 1	88.1	FIELD 1	90.0636364
VARIETY 2	35.98	FIELD 2	103.172727
VARIETY 3	69.92	FIELD 3	90.5363637
VARIETY 4	67.64	FIELD 4	87.4727273
VARIETY 5	125.76	FIELD 5	73.3454546
VARIETY 6	107.74		
VARIETY 7	97.12		
VARIETY 8	70.94		
VARIETY 9	84.74		
VARIETY 10	154.44		
VARIETY 11	75.72		

GRAND MEAN 88.9181818 STD DEVIATION 33.8664103

ANALYSIS OF COVARIANCE TABLE FOR DRYWT,VITC DATA

SOURCE	DF	SS(X,X)	SS(X,Y)	SS(Y,Y)	DF	ADJ. SS(Y,Y)	ADJ. MSS(Y,Y)	F RATIO
TOTAL	54	2916.22102	-12226.1373	61934.4223				
FIELD	4	367.85376	-1246.65637	4968.94031				
VARIETY	10	2166.71303	-9784.13318	51018.1782				
ERROR	40	381.654236	-1195.34778	5947.30371	39	2203.45345	56.4988064	
VARIETY + ERROR	50	2548.36726	-10979.481	56965.482	49	9661.07499		
VARIETY ADJUSTED					10	7457.62154	745.762154	13.199609

6. One-way ANOVA for randomized block experiment from LIZANCOV.RAN or LIZRAND.AGG

(Reduced size copy)

NAME OF STUDY: RELATIONSHIP OF DIET AND AGE TO SBP LEVEL
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 4 AUGUST 1983
RUN FROM LIZPACK DATA FILE: QUET/DAT

ANOVA TABLE FOR SBP LEVEL DATA

SOURCE	DF	SS	MSS	F RATIO
DIET	3	33.5609369	11.186979	11.9182122
AGQTGRP	7	462.859689	66.1228127	70.444909
ERROR	21	19.7115603	.938645726	
TOTAL	31	516.132186		

FACTOR	MEAN	FACTOR	MEAN
DIET 1	12.575	AGQTGRP 1	9.975
DIET 2	10.3625	AGQTGRP 2	5.075
DIET 3	10.4625	AGQTGRP 3	14.6
DIET 4	9.9375	AGQTGRP 4	6.9
		AGQTGRP 5	12.8
		AGQTGRP 6	8.8
		AGQTGRP 7	17.475
		AGQTGRP 8	11.05

GRAND MEAN 10.834375 STD DEVIATION 4.08037074

7. ANOVA for Randomized Block Experiment from LIZANOVA.RAN or LIZRAND.AGG

(Reduced size copy)

NAME OF STUDY: EFFECT OF TOXIC SUBSTANCES ON FEV LEVEL BY PLANT
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 4 AUGUST 1983
RUN FROM LIZPACK DATA FILE: FEV/DAT

TWO-WAY ANOVA TABLE FOR FEV DATA

SOURCE	DF	SS	MSS	F(FIXED)	F(RANDOM)	F(CR,RF)	F(CF,RR)
PLANT	2	3.29889584	1.64944792	6.14449939	.269185576	.269185576	6.14449939
TOXSUB	2	66.889369	33.4446845	124.587652	5.45808484	124.587652	5.45808484
PLANT X TOXSUB	4	24.5101976	6.12754941	22.8262579	22.8262579	22.8262579	22.8262579
ERROR	99	26.5758581	.268443011				
TOTAL	107	121.274321					

FACTOR	MEAN	FACTOR	MEAN
PLANT 1	3.70361111	TOXSUB 1	4.75416667
PLANT 2	3.90305556	TOXSUB 2	4.12277778
PLANT 3	4.13138887	TOXSUB 3	2.86111111

GRAND MEAN 3.71268519 STD DEVIATION 1.06461487

CELL	MEAN	CELL	MEAN	CELL	MEAN	CELL	MEAN	CELL	MEAN	CELL	MEAN
1 1	4.87833334	1 2	3.46416667	1 3	2.74833334						
2 1	5.23416667	2 2	3.7775	2 3	2.6975						
3 1	4.13	3 2	5.12666667	3 3	3.1375						

8. Two-way ANOVA with equal cell frequencies; from LIZ2WAY.TWO or LIZ2WAY.AGG

(Reduced size copy)

NAME OF STUDY: EFFECT OF USE OF FUNGICIDES ON YIELD OF OATS
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 4 AUGUST 1983
RUN FROM LIZPACK DATA FILE: OATS/DAT

TWO-WAY ANOVA TABLE FOR YIELD OF OATS DATA

SOURCE	DF	SS	MSS	F(FIXED)	F(RANDOM)	F(CR,RF)	F(CF,RR)
FIELD	3	2842.87305					
VARIETY	3	2848.02185	949.340617	31.6564834	14.5687433	14.5687433	31.6564834
FUNG	3	170.536926	56.8456421	1.89556108	.872362937	1.89556108	.872362937
VARIETY X FUNG	9	586.465515	65.162835	2.17290419	2.17290419	2.17290419	2.17290419
ERROR	45	1349.49695	29.9888211				
TOTAL	63	7797.39429					

FACTOR	MEAN	FACTOR	MEAN
VARIETY 1	42.45625	FUNG 1	50.6875
VARIETY 2	53.40625	FUNG 2	55.2
VARIETY 3	54.30625	FUNG 3	53.125
VARIETY 4	61.06875	FUNG 4	52.225

GRAND MEAN 52.809375 STD DEVIATION 11.1251141

CELL	MEAN	CELL	MEAN	CELL	MEAN	CELL	MEAN	CELL	MEAN	CELL	MEAN
1 1	36.05	1 2	50.625	1 3	45.85	1 4	37.3				
2 1	50.85	2 2	55.375	2 3	53.1	2 4	54.3				
3 1	53.925	3 2	51.375	3 3	55.875	3 4	56.05				
4 1	61.925	4 2	63.425	4 3	57.675	4 4	61.25				

9. Two-way randomized block experiment from LIZBLOCK.2WY or LIZ2BLOCK.AGG

(Reduced size copy)

NAME OF STUDY: EFFECT OF RATIONS ON WEIGHT GAIN IN BACON PIGS
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 4 AUGUST 1983
RUN FROM LIZPACK DATA FILE: WEIGHT/PIG

TWO-WAY ANOVA TABLE FOR INITWT DATA

SOURCE	DF	SS	MSS	F(FIXED)	F(RANDOM)	F(CR,RF)	F(CF,RR)
RATION	2	5.3999939	2.69999695	.061784827	.240355724	.240355724	.061784827
SEX	1	32.0333252	32.0333252	.733028035	2.85163029	.733028035	2.85163029
RATION X SEX	2	22.4666748	11.2333374	.257055775	.257055775	.257055775	.257055775
ERROR	24	1048.8	43.7000001				
TOTAL	29	1108.7					

FACTOR	MEAN	FACTOR	MEAN
RATION 1	39.8	SEX 1	41.1333334
RATION 2	40.7	SEX 2	39.0666667
RATION 3	39.8		
GRAND MEAN 40.1		STD DEVIATION 6.18312497	

CELL MEAN	CELL MEAN	CELL MEAN	CELL MEAN	CELL MEAN	CELL MEAN
1 1 41	1 2 38.6				
2 1 40.6	2 2 40.8				
3 1 41.8	3 2 37.8				

TWO-WAY ANOVA TABLE FOR WTGAIN DATA

SOURCE	DF	SS	MSS	F(FIXED)	F(RANDOM)	F(CR,RF)	F(CF,RR)
RATION	2	2.26864624	1.13432312	2.06770688	4.76479555	4.76479555	2.06770688
SEX	1	.43440342	.43440342	.791854563	1.82473886	.791854563	1.82473886
RATION X SEX	2	.476126671	.238063335	.433955006	.433955006	.433955006	.433955006
ERROR	24	13.1661577	.548589905				
TOTAL	29	16.3453341					

10. Two-way ANCOVA by LIZ2WAY.ANC or LIZ2ANCV.AGG
(page 1 of 2 pages)

(Reduced size copy)

FACTOR	MEAN	FACTOR	MEAN
RATION 1	9.649	SEX 1	9.184
RATION 2	9.288	SEX 2	9.42466667
RATION 3	8.97600001		

GRAND MEAN 9.30433334 STD DEVIATION .750754427

CELL	MEAN	CELL	MEAN
1 1	9.606	1 2	9.692
2 1	8.99	2 2	9.586
3 1	8.956	3 2	8.996

ANALYSIS OF COVARIANCE TABLE FOR INITWT,WTGAIN DATA

SOURCE	DF	SS(X,X)	SS(X,Y)	SS(Y,Y)	DF	ADJ. SS(Y,Y)	ADJ. MSS(Y,Y)	F RATIO
TOTAL	29	1108.7	78.587	16.3453341				
PEN	4	605.866669	39.9053345	4.85178566				
SEX	1	32.0333252	-3.73032761	.43440342				
RATION	2	5.3999939	-.146999359	2.26864624				
RATION X SEX	2	22.4666748	3.11232758	.476126671				
ERROR	20	442.933334	39.3666649	8.31437207	19	4.81557398	.253451262	
RATION + ERROR	22	448.333328	39.2196655	10.5830183	21	7.15212868		
RATION ADJUSTED					2	2.33655471	1.16827735	4.60947539
SEX + ERROR	21	474.96666	35.6363373	8.74877549	20	6.07501195		
SEX ADJUSTED					1	1.25943798	1.25943798	4.96915252
RATION X SEX + ERROR	22	465.400009	42.4789925	8.79049874	21	4.9132646		
RATION X SEX ADJUSTED					2	.0976906289	.0488453144	.192720739

ERROR REGRESSION COEFFICIENT .0888771782 STANDARD ERROR .0239209292
RATION REGRESSION COEFFICIENT .0874788089 STANDARD ERROR .0513574907
SEX REGRESSION COEFFICIENT .0750291343 STANDARD ERROR .0533235689
RATION X SEX REGRESSION COEFFICIENT .0912741548 STANDARD ERROR .0105012808

Two-way ANCOVA by LIZ2WAY.ANC or LIZ2ANCV.AGG
(page 2 of 2 pages)

(Reduced size copy)

NAME OF STUDY: EFFECT OF FUNGICIDES ON LEGUME SEED GERMINATION
NAME OF INVESTIGATOR: WILLIAM J. GRAY
DATE OF RUN: 16 DECEMBER 1983
RUN FROM LIZPACK DATA FILE: LEGUME/DAT

THREE WAY ANOVA TABLE FOR SEEDLINGS EMERGED DATA

SOURCE	DF	SS	MSS	F. ALL FIXED	F. ALL RANDOM
SPECIES	2	9900.11121	4950.0556	4.42921297	NO EXACT TEST
SOIL	2	16436.1112	8218.0556	7.35335547	NO EXACT TEST
FUNG	1	1932.01856	1932.01856	1.72873243	NO EXACT TEST
SPECIES X SOIL	4	658.444336	164.611084	.147290779	.615581611
SPECIES X FUNG	2	194.036987	97.0184937	.0868102511	.362811538
SOIL X FUNG	2	1851.14807	925.574036	.828185549	3.46128792
SPECIES X SOIL X FUNG	4	1069.62964	267.40741	.239270921	.239270921
ERROR	36	40233.3334	1117.59259		
TOTAL	53	72274.8334			

MIXED EFFECTS MODELS

SOURCE	F (C RANDOM)	F (B RANDOM)	F (A RANDOM)	F (C FIXED)	F (B FIXED)	F (A FIXED)
SPECIES	51.0217735	30.0712169	4.42921297	30.0712169	51.0217735	NO EXACT TEST
SOIL	8.67887439	7.35335547	49.924072	49.924072	NO EXACT TEST	8.67887439
FUNG	1.72873243	2.08737333	19.9139204	NO EXACT TEST	19.9139204	2.08737333
SPECIES X SOIL	.615581611	.147290779	.147290779	.147290779	.615581611	.615581611
SPECIES X FUNG	.0868102511	.362811538	.0868102511	.362811538	.0868102511	.362811538
SOIL X FUNG	.828185549	.828185549	3.46128792	3.46128792	3.46128792	.828185549
SPECIES X SOIL X FUNG	.239270921	.239270921	.239270921	.239270921	.239270921	.239270921

FACTOR	MEAN	FACTOR	MEAN	FACTOR	MEAN
SPECIES 1	76.6666667	SOIL 1	77.7222223	FUNG 1	63.962963
SPECIES 2	82.1111111	SOIL 2	86.3333334	FUNG 2	75.925926
SPECIES 3	51.0555556	SOIL 3	45.7777778		

GRAND MEAN 69.9444445 STANDARD DEVIATION 36.9279854

CELL	MEAN	CELL	MEAN	CELL	MEAN	CELL	MEAN	CELL	MEAN
1 1 1	88.6666667	1 1 2	92	1 2 1	95.3333334	1 2 2	90.3333334	1 3 1	22
1 3 2	71.6666667	2 1 1	94	2 1 2	91.6666667	2 2 1	96.3333334	2 2 2	97.3333334
2 3 1	55.6666667	2 3 2	67.6666667	3 1 1	50.6666667	3 1 2	59.3333333	3 2 1	65.6666667
3 2 2	73	3 3 1	17.3333333	3 3 2	40.3333333				

MEANS FOR INTERACTIONS

CELL	MEAN	CELL	MEAN	CELL	MEAN	CELL	MEAN	CELL	MEAN
AB 1 1	90.3333334	AB 1 2	92.3333334	AB 1 3	46.8333333	AB 2 1	87.3333334	AB 2 2	96.8333334
AB 2 3	61.6666667	AB 3 1	55	AB 3 2	69.3333334	AB 3 3	28.8333334		
AC 1 1	68.6666667	AC 1 2	84.6666667	AC 2 1	78.6666667	AC 2 2	85.5555556	AC 3 1	44.5555556
AC 3 2	57.5555556								
BC 1 1	74.4444445	BC 1 2	81	BC 2 1	85.7777778	BC 2 2	86.8888889	BC 3 1	31.6666667
BC 3 2	59.8888889								

11. Three-way ANOVA by LIZANOVA.3WY (equal cell frequencies). Data from Steel and Torrie. This is a 3 x 3 x 2 factorial experiment.

(Reduced size copy)

```
NAME OF STUDY: EFFECT OF CERTAIN CHEMICALS ON LEAF BURN RATE OF TOBACCO
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN:5 AUGUST 1983
RUN FROM LIZPACK DATA FILE:  TOBTEST/DAT
```

PEARSON CORRELATION COEFFICIENTS

VARIABLES	COEFFICIENT, R	DF	T (R=0)
1 2	.209400418	28	1.13316516
1 3	.0918776927	28	.488236144
1 4	-.717729229	28	-5.45417842
2 3	.40699788	28	2.35774221
2 4	-.499638451	28	-3.05210606
3 4	.18034204	28	.970187611

```
NAME OF STUDY:  EFFECT OF CERTAIN CHEMICALS ON LEAF BURN RATE OF TOBACCO
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 5 AUGUST 1983
RUN FROM LIZPACK DATA FILE:  TOBTEST/DAT
```

ANOVA TABLE FOR REGRESSION

SOURCE	DF	SS	MSS	F RATIO
REGRESSION	3	5.50667035	1.83555678	40.3470351
ERROR	26	1.18284965	.0454942174	
TOTAL	29	6.68952001		

PARTIAL REGRESSION COEFFICIENTS

VAR. #	COEFFICIENT, B	STD ERROR	DF	T (FOR B=0)
1	-.531216034	.0695195378	26	-7.64124808
2	-.439626401	.0729675412	26	-6.02495842
3	.209228133	.0406234952	26	5.15042174

Y-INTERCEPT 1.80893442

OTHER STATISTICS

VARIABLE	MEAN	STANDARD DEVIATION
1	3.27866667	.582667672
2	.807666666	.605176996
3	4.65433333	1.06742904
4	.686	.480284399

STANDARD ERROR OF ESTIMATE: .213293735

12. Printout of correlation matrix and a regression run by LIZCORR.MUL

(Reduced size copy)

NAME OF STUDY: BURN RATE OF TOBACCO VS CHEMICAL COMPOSITION OF LEAF
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 30 JANUARY 1983
NAME OF DATA FILE: TOBTEST/DAT

ANOVA TABLE FOR REGRESSION-DEPENDENT VARIABLE: BRNRT

SOURCE	DF	SS	MSS	F RATIO
REGRESSION	3	5.50667039	1.8355568	40.3470366
ERROR	26	1.18284961	.0454942159	
TOTAL	29	6.68952		

PARTIAL REGRESSION COEFFICIENTS

VAR. #	COEFFICIENT, B	STD ERROR	MEAN	T (FOR B=0)
NPCT	-.531216042	.0695195372	3.27866667	-7.64124826
KPCT	-.439626398	.07296754	.807666667	-6.02495847
CLPCT	.20922813	.0406234953	4.65433333	5.15042165

Y-INTERCEPT 1.80893446

DF FOR T-TEST ON REGRESSION COEFFS.: 26
STANDARD ERROR OF ESTIMATE: .213293732
MEAN OF DEPENDENT VARIABLE: .686
STD. ERROR OF EST. AS PCT OF MEAN OF DEP. VAR.: 31.0923807
PCT OF VARIANCE EXPLAINED BY REGRESSION: 82.3178701
NUMBER OF OBSERVATIONS: 30

13. Regression and Residual Analysis from LIZANAL.RES. Data from Steel and Torrie. Note the probability of several outliers.

(page 1 of 2 pages)

(Reduced size copy)

CASE	Y, OBSERVED	Y, COMPUTED	RESIDUAL	NORMAL DEV.	STUD. RES.	V(I,I)	COOK'S DIST.	T FOR OUTLIER
1	.34	.737590751	-.397590751	-1.86405267	-1.96177752	.0971473996	.103527051	-2.08410159
2	.11	-9.44416411E-03	.119444164	.559998474	.602436837	.135926576	.0142730497	.594904605
3	.38	.79703588	-.41703588	-1.95521864	-2.02193343	.0649023743	.070937739	-2.15972261
4	.68	.629924243	.0500757569	.234773692	.248173921	.105075135	1.80786367E-03	.2436433
5	.18	.118718126	.0612818742	.28731212	.311508614	.149316966	4.25815547E-03	.306030948
6	1.53	1.22453035	.305469651	1.43215485	1.51027672	.100778055	.0639075456	1.55052137
7	.77	1.16964912	-.399649115	-1.87370305	-2.01384463	.134335508	.157337828	-2.14948201
8	.89	1.19538671	-.305386712	-1.431766	-1.5581791	.155675608	.111914406	-1.60468003
9	1.05	.973160964	.0768390363	.360249857	.37642905	.0841141028	3.25337194E-03	.370129023
10	1.15	1.1538179	-3.81790381E-03	-.0178997469	-.019601125	.166065801	1.91271571E-05	-.0192206264
11	1.49	1.37194127	.118058729	.553503038	.592192768	.126397598	.0126850216	.584649093
12	.18	.215436193	-.0354361927	-.166137994	-.193809833	.265170965	3.38868567E-03	-.190183606
13	.34	.449176552	-.109176552	-.511860106	-.540325577	.102588741	8.34371089E-03	-.532832831
14	.92	.681476968	.238523033	1.11828431	1.18880389	.115120724	.0459652823	1.19875298
15	1.35	1.05405932	.295940676	1.38747948	1.44736629	.0810407942	.0461853639	1.48014034
16	1.33	1.10031027	.229689734	1.07687053	1.1419969	.110804755	.0406285319	1.14900787
17	.23	.184217384	.0457826159	.214645857	.24960068	.260473923	5.48582937E-03	.245047368
18	0	.171783247	-.171783247	-.805383474	-.882600338	.167321697	.039133005	-.87872463
19	.08	-.0796603351	.159660335	.748546776	.823050959	.172849689	.0353897605	.81779169
20	.11	.305290385	-.195290385	-.915593643	-.98312253	.132658276	.0369572303	-.982465086
21	1.17	1.07573309	.0942669148	.441958205	.459951546	.0767097801	4.39416821E-03	.452865789
22	1.01	1.01742693	-7.42692873E-03	-.0348201922	-.0358998192	.05924225	2.02898236E-05	-.0352035415
23	1.4	1.16450912	.235490884	1.10406847	1.16629281	.103858026	.03941109	1.17478761
24	.51	.458757369	.0512426312	.240244431	.25917964	.140778979	2.75153276E-03	.254475493
25	.36	.124330675	.235669325	1.10490507	1.2288071	.191495386	.0894094137	1.24153519
26	.89	.849755937	.0402440634	.188679072	.198451388	.0960608811	1.04629868E-03	.194745145
27	.91	.994006341	-.0840063405	-.393852833	-.41684848	.107287733	5.22077554E-03	-.410126336
28	.26	.391077575	-.131077575	-.614540211	-.651737492	.110890603	.0132441766	-.644366369
29	.73	.830548365	-.100548365	-.471407971	-.491785949	.0811563681	5.34039321E-03	-.48449446
30	.23	.229453528	5.46471682E-04	2.56206161E-03	3.07270401E-03	.304755306	1.0346538E-06	3.01303472E-03

```
DF FOR T-TEST FOR CASE BEING OUTLIER:  25
T(2,1)= 1.01019394
T(1,1)=-3.75839022E-08
T(1,2)= .269450558
STATISTIC FOR DURBIN-WATSON TEST:    D= 1.91854403        4-D= 2.08145597
SUM OF SQUARES OF RESIDUALS 1.18284962
```

13. Regression and Residual Analysis from LIZANAL.RES. Data from Steel and Torrie. Note the probability of several outliers.

(Reduced size copy)

14. Regression and Residual Analysis from LIZANAL.RES (on the following page). The data used here were introduced by Longley in 1967. Many of the statistical packages of that time could not obtain the estimates of the partial regression coefficients to even one decimal place. The Longley Data are a formidable challenge to any Regression program; however, many modern packages will perform excellently with these data. LIZPACK actually outperforms some mainframe packages on these data. The model is overparameterized, and a sensitivity test indicates that no matter how accurately the estimates are computed, they are not reliable estimates of the true population parameters (see Weisberg, Applied Regression Analysis). This run is from a scaled version of the Longley data.

NAME OF STUDY: REGRESSION AND RESIDUAL ANALYSIS-LONGLEY DATA
NAME OF INVESTIGATOR: WILLIAM J. GRAY
DATE OF RUN: 15 DECEMBER 1983
NAME OF DATA FILE: LONGLEY/DAT

ANOVA TABLE FOR REGRESSION-DEPENDENT VARIABLE: EMP

SOURCE	DF	SS	MSS	F RATIO
REGRESSION	6	184.172403	30.6954005	330.28546
ERROR	9	.836423755	.0929359727	
TOTAL	15	185.008827		

PARTIAL REGRESSION COEFFICIENTS

VAR. #	COEFFICIENT. B	STD ERROR	MEAN	T (FOR B=0)
GNPDF	.0150618419	.0849148946	101.68125	.177375736
GNP	-.0358191755	.0334909903	387.698438	-1.06951676
UNEMP	-.0202022979	4.88399446E-03	319.33125	-4.13642931
ARFC	-.0103322692	2.14274116E-03	260.66875	-4.82198664
POPN	-.0511042122	.226073084	117.424	-.226051732
YEAR	1.82915161	.455478364	1954.5	4.01589133

Y-INTERCEPT -3482.25891

DF FOR T-TEST ON REGRESSION COEFFS.: 9
STANDARD ERROR OF ESTIMATE: .304854019
MEAN OF DEPENDENT VARIABLE: 65.317
STD. ERROR OF EST. AS PCT OF MEAN OF DEP. VAR.: .466729977
PCT OF VARIANCE EXPLAINED BY REGRESSION: 99.5479007
NUMBER OF OBSERVATIONS: 16

CASE	Y, OBSERVED	Y, COMPUTED	RESIDUAL	NORMAL DEV.	STUD. RES.	V(I,I)	COOK'S DIST.	T FOR OUTLIER
1	60.323	60.0556603	.267339751	.876943503	1.1560135	.424536986	.142839959	1.18111257
2	61.122	61.2160139	-.0940138996	-.308389898	-.467567864	.564978244	.240561314	-.446280853
3	60.171	60.124713	.0462870598	.151833523	.190100301	.362074843	2.93019276E-03	.179589202
4	61.187	61.5971146	-.410114557	-1.34528178	-1.69790052	.372227865	.244193245	-1.94170499
5	63.221	62.9112854	.309714601	1.01594397	1.63842956	.615510966	.613916559	1.8440268
6	63.639	63.8883114	-.249311402	-.817805858	-1.02999024	.369573614	.0889453263	-1.03393163
7	64.989	65.1530495	-.164049476	-.538124697	-.754659169	.491531397	.0786485625	-.735132978
8	63.761	63.7741805	-.0131804049	-.0432351358	-.0614304127	.504656134	5.49234227E-04	-.0579291947
9	66.019	66.004695	.0143050551	.0469242794	.0636860072	.457116932	4.87879744E-04	.0600573387
10	67.857	67.4016056	.455394417	1.49381143	1.82581961	.332615238	.23521465	2.1694513
11	68.169	68.1862688	-.0172688067	-.0566461507	-.0708011367	.359981547	4.02607307E-04	-.0667705493
12	66.513	66.5520554	-.0390553773	-.128111735	-.178195115	.483124157	4.24000008E-03	-.168301125
13	68.655	68.8105493	-.155549735	-.510243347	-.645055643	.374303373	.0355603083	-.622729969
14	69.564	69.6496716	-.0856715739	-.281024912	-.319920929	.228378471	4.32751009E-03	-.303354172
15	69.331	68.989068	.341931969	1.12162526	1.41634536	.372872492	.17038881	1.51478992
16	70.551	70.7577582	-.206758142	-.678220161	-1.21540674	.688614697	.466684544	-1.25336415

DF FOR T-TEST FOR CASE BEING OUTLIER: 8
T(2,1)= 54.47289
T(1,1)=-3.20114175E-05
T(1,2)=-.594469071
STATISTIC FOR DURBIN-WATSON TEST: D= 2.55948685 4-D= 1.44051315
SUM OF SQUARES OF RESIDUALS .836424843

(Reduced size copy)

NAME OF STUDY: PRODUCTION OF PETROLEUM PRODUCTS IN THE U.S.
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 5 AUGUST 1983
RUN FROM LIZPACK DATA FILE: PETPROD/DAT

ANOVA TABLE FOR REGRESSION

SOURCE	DF	SS	MSS	F RATIO
REGRESSION	1	.903573336	.903573336	520.300615
ERROR	16	.0277861932	1.73663707E-03	
TOTAL	17	.93135953		

REGRESSION COEFFICIENTS

A= 1085.52393	B= 1.04413124

OTHER STATISTICS

VARIABLE	MEAN	STANDARD DEVIATION
1	9.5	5.50284018
2	1678.81111	403.649458

THE SQUARED CORRELATION COEFFICIENT IS: .970165986

15. Regression run from LIZEXPO.ONE

(Reduced size copy)

NAME OF STUDY: YIELD OF ALASKAN FIELD PEAS VS TENDOMETER READING
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 5 AUGUST 1983
RUN FROM LIZPACK DATA FILE: TENDOM/RDG

ANOVA TABLE FOR REGRESSION

SOURCE	DF	SS	MSS	F RATIO
REGRESSION	1	9731.65064	9731.65064	203.27496
ERROR	23	1101.10938	47.8743208	
TOTAL	24	10832.76		

REGRESSION COEFFICIENTS

A=-373.903725	B= 91.4577014

OTHER STATISTICS

VARIABLE	MEAN	STANDARD DEVIATION
1	107.956	24.4823819
2	52.14	21.702294

THE SQUARED CORRELATION COEFFICIENT IS: .898353755

16. Regression run from LIZLOG.ANL

(Reduced size copy)

NAME OF STUDY: YIELD OF ALASKAN FIELD PEAS VS TENDOMETER READING
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 5 AUGUST 1983
RUN FROM LIZPACK DATA FILE: TENDOM/RDG

ANOVA TABLE FOR REGRESSION

SOURCE	DF	SS	MSS	F-RATIO
REGRESSION	1	4.43282124	4.43282124	139.429042
ERROR	23	.731231364	.031792668	
TOTAL	24	5.16405261		

REGRESSION COEFFICIENTS

A= 5.33465734E-03	B= 1.95194302

OTHER STATISTICS

VARIABLE	MEAN	STANDARD DEVIATION
1	107.956	24.4823819
2	52.14	21.702294

THE SQUARED CORRELATION COEFFICIENT IS: .858399707

17. Regression run from LIZRAD.PWR

(Reduced size copy)

NAME OF STUDY: YIELD OF ALASKAN FIELD PEAS VS TENDOMETER READING
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 5 AUGUST 1983
RUN FROM LIZPACK DATA FILE: TENDOM/RDG

ANOVA TABLE FOR POLYNOMIAL REGRESSION, DEGREE 3

SOURCE	DF	SS	MSS	F RATIO
REGRESSION	3	10017.1881	3339.06269	85.9768635
ERROR	21	815.57193	38.8367586	
TOTAL	24	10832.76		

PARTIAL REGRESSION COEFFICIENTS

TERM	COEFFICIENT, B	STD ERROR	DF	T (FOR B=0)
1	-6.69128138	5.20361355	21	-1.2858913
2	.0781729648	.0476771569	21	1.63963143
3	-2.58171633E-04	1.42023672E-04	21	-1.81780706

Y-INTERCEPT 192.56475

OTHER STATISTICS

TERM	MEAN	STANDARD DEVIATION
1	107.956	23.9669056
2	12205.934	5418.59862
3	1441811.49	951143.828
4	52.14	21.2453524

STANDARD ERROR OF ESTIMATE: 6.23191452

THE SQUARED CORRELATION COEFFICIENT IS: .924712453

ANOVA TABLE FOR POLYNOMIAL REGRESSION, DEGREE 2

SOURCE	DF	SS	MSS	F RATIO
REGRESSION	2	9888.85512	4944.42756	115.241916
ERROR	22	943.904877	42.9047672	
TOTAL	24	10832.76		

18. Polynomial regression by LIZPOLY.STP.
Degree of polynomials: 3, 2, 1
(page 1 of 2 pages)

(Reduced size copy)

PARTIAL REGRESSION COEFFICIENTS

TERM	COEFFICIENT, B	STD ERROR	DF	T (FOR B=0)
1	2.71328375	.586807256	22	4.62380741
2	-8.37858184E-03	2.59549656E-03	22	-3.22812289

Y-INTERCEPT -138.506843

STANDARD ERROR OF ESTIMATE: 6.55017307

THE SQUARED CORRELATION COEFFICIENT IS: .912865708

ANOVA TABLE FOR POLYNOMIAL REGRESSION, DEGREE 1

SOURCE	DF	SS	MSS	F RATIO
REGRESSION	1	9441.75409	9441.75409	156.117484
ERROR	23	1391.00592	60.4785181	
TOTAL	24	10832.76		

PARTIAL REGRESSION COEFFICIENTS

TERM	COEFFICIENT, B	STD ERROR	DF	T (FOR B=0)
1	.827577657	.0662343052	23	12.4946982

Y-INTERCEPT -37.2019736

STANDARD ERROR OF ESTIMATE: 7.77679357

THE SQUARED CORRELATION COEFFICIENT IS: .871592658

18. Polynomial regression by LIZPOLY.STP.
Degree of polynomials: 3, 2, 1
(page 2 of 2 pages)

(Reduced size copy)

NAME OF STUDY: YIELD OF ALASKAN FIELD PEAS VS TENDOMETER READING
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 4 AUGUST 1983
RUN FROM LIZPACK DATA FILE: TENDOM/RDG

INDEPENDENT VAR.	OBSERVED Y	COMPUTED Y	RESIDUAL
76.2	24	19.5926327	4.40736729
76.8	22	20.4511488	1.54885125
77.3	26.5	21.1619697	5.33803028
79.2	22	23.8248767	-1.82487667
80	25	24.9279999	.0720000863
87.8	37.5	35.1213207	2.37867928
93.2	36	41.5809087	-5.58090872
93.5	39.5	41.9254449	-2.4254449
94.3	32	42.8368337	-10.8368337
96.8	26.5	45.6157888	-19.1157888
97.5	55.5	46.3751249	9.12487507
99.5	49.5	48.4994049	1.00059509
104.2	56	53.2275767	2.77242333
106.3	55.5	55.2204977	.279502273
106.7	58	55.5917217	2.40827829
119	61.5	65.6978199	-4.19781989
119.7	69	66.1967057	2.8032943
119.8	71.5	66.2673047	5.23269534
119.8	73	66.2673047	6.73269534
123.5	76.5	68.7616449	7.73835516
141	78.5	77.4502198	1.04978019
142.3	74	77.8908497	-3.89084965
145.5	71.5	78.8548049	-7.35480487
149	77	79.7126198	-2.71261978
150	85.5	79.9199999	5.58000011

19. Computation of residuals from LIZPOLY.RES

(Reduced size copy)

```
NAME OF STUDY HARMAN'S 5 SOCIOECONOMIC VARIABLES
NAME OF INVESTIGATOR J. Q. SMART
DATE OF RUN 8 NOVEMBER 1983
COMPUTATION OF PRINCIPAL COMPONENTS

VARIABLES
---------------------------------------------------------------------------
TPOP  MSY   TEMP  MPS   MVH
---------------------------------------------------------------------------


CONTRIBUTION TO VARIANCE (EIGENVALUE), BY COMPONENT
---------------------------------------------------------------------------
PC 1  2.87331358   PC 2  1.79666011   PC 3  .214836802   PC 4  .0999340522   PC 5  .0152554622
---------------------------------------------------------------------------


PCT CONTRIBUTION, BY COMPONENT
---------------------------------------------------------------------------
PC 1  57.47 PC 2  35.93 PC 3  4.3   PC 4  2     PC 5  .31
---------------------------------------------------------------------------

CONTRIBUTION TO VARIANCE, BY VARIABLE
---------------------------------------------------------------------------
TPOP  1.00003364   MSY   .999924115   TEMP 1.00009368   MPS   .999992503   MVH   .999956067
---------------------------------------------------------------------------
TOTAL VARIANCE 5.00000001
---------------------------------------------------------------------------


EIGENVECTORS (COLUMNS OF TRANSFORM MATRIX)
---------------------------------------------------------------------------
EIGENVECTOR  1
 .342794949     .452460959     .396752163     .550049238     .466695041
EIGENVECTOR  2
 .601592241    -.406465168     .541621832    -.0778736708   -.416478482
EIGENVECTOR  3
-.0590588591   -.688705281    -.248427146     .664074211     .139593992
EIGENVECTOR  4
-.203953461     .353591137    -.0230165225    .50038643     -.763190642
EIGENVECTOR  5
 .689560321     .175281182    -.697845045    -6.248C3512E-04 -.0824316436

---------------------------------------------------------------------------


COEFFICIENTS OF PRINCIPAL COMPONENTS
---------------------------------------------------------------------------
PRINCIPAL COMPONENT NUMBER  1  EIGENVALUE 2.87331358
TPOP  .58106662    MSY   .766959841   TEMP  .672528689   MPS   .932380282   MVH   .791087822
PRINCIPAL COMPONENT NUMBER  2  EIGENVALUE 1.79666011
TPOP  .806371535   MSY  -.544824083   TEMP  .725987468   MPS  -.104381518   MVH  -.558245884
PRINCIPAL COMPONENT NUMBER  3  EIGENVALUE .214836802
TPOP -.0273740712  MSY  -.31921828    TEMP -.115147202   MPS   .307801658   MVH   .0647025008
PRINCIPAL COMPONENT NUMBER  4  EIGENVALUE .0999340522
TPOP -.0644744772  MSY   .111778459   TEMP -7.27606311E-03 MPS  .158183898 ' MVH  -.241262479
PRINCIPAL COMPONENT NUMBER  5  EIGENVALUE .0152554622
TPOP  .0851696669  MSY   .0216495054  TEMP -.0861929381  MPS  -7.718124E-05  MVH  -.01018138

---------------------------------------------------------------------------
```

20. Principal Component Analsysis from LIZCOMP.PCL

(Reduced size copy)

```
NAME OF STUDY: HARMAN'S 5 SOCIOECONOMIC VARIABLES-OBLIQUE ROTATION
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 8 NOVEMBER 1983

VARIABLES
------------------------------------------------------------------------------------------
TPOP   MSY   TEMP   MPS   MVH
------------------------------------------------------------------------------------------


COEFFICIENTS OF PRINCIPAL FACTORS (FACTOR PATTERN MATRIX)
------------------------------------------------------------------------------------------
PRINCIPAL FACTOR NUMBER  1
TPOP -.0738706261   MSY  .959086852    TEMP .0506229313   MPS  .79959097    MVH  .98681777
PRINCIPAL FACTOR NUMBER  2
TPOP  .995486123    MSY -.0996655076   TEMP .970061313    MPS  .368989076   MVH -.099534753

------------------------------------------------------------------------------------------


COLUMNS OF TRANSFORM MATRIX
------------------------------------------------------------------------------------------
COLUMN  1
 .867090804     .631264704
COLUMN  2
-.498150113     .775567451

------------------------------------------------------------------------------------------


FACTOR STRUCTURE MATRIX, BY COLUMNS
------------------------------------------------------------------------------------------
PRINCIPAL FACTOR NUMBER  1
TPOP  .0864173786   MSY  .943039229    TEMP .206817165    MPS  .859003674   MVH  .970791201
PRINCIPAL FACTOR NUMBER  2
TPOP  .983591859    MSY  .0547616754   TEMP .978212355    MPS  .497735061   MVH  .0593575185

------------------------------------------------------------------------------------------


JOINT AND DIRECT CONTRIBUTION TO VARIANCE BY FACTORS
------------------------------------------------------------------------------------------
PF 1  2.54102217    PF 2  2.08800489
 1  2  .0247318602

------------------------------------------------------------------------------------------


CONTRIBUTION TO VARIANCE, BY VARIABLE
------------------------------------------------------------------------------------------
TPOP  .988083936    MSY  .921415247    TEMP .935216077    MPS  .767133103   MVH  .975350923
------------------------------------------------------------------------------------------
TOTAL VARIANCE 4.65375892
------------------------------------------------------------------------------------------


CORRELATION AMONG FACTORS
------------------------------------------------------------------------------------------
 1  2  .161014806

------------------------------------------------------------------------------------------
```

21. Oblique rotation of two principal factors by LIZOBLI.MIN

(Reduced size copy)

```
NAME OF STUDY REF. COOLEY AND LOHNES-PP.250-255
NAME OF INVESTIGATOR J. Q. SMART
DATE OF RUN      4 NOVEMBER 1983
RUN FROM LIZPACK DATA FILE:  INFOANAL/AGG


--------------------------------------------------------------------------------
UNIVARIATE F-RATIOS-NUM DF  2   DEN DF   193
--------------------------------------------------------------------------------
VARIABLE        AMONG MSS      WITHIN MSS     F-RATIO        ETA SQUARE
--------------------------------------------------------------------------------
 1              3925.31934     1168.64055     3.35887654     .033636234
 2              1294.31372     299.086631     4.32755459     .0429203565
--------------------------------------------------------------------------------
CORRELATION MATRIX
--------------------------------------------------------------------------------
POSITION OF ENTRY              ENTRY
--------------------------------------------------------------------------------
 1  1                          1
 1  2                          .854806529
--------------------------------------------------------------------------------
 2  2                          1
--------------------------------------------------------------------------------
TOTAL MATRIX
--------------------------------------------------------------------------------
POSITION OF ENTRY              ENTRY
--------------------------------------------------------------------------------
 1  1                          233398.266
 1  2                          101419.102
--------------------------------------------------------------------------------
 2  2                          60312.3472
--------------------------------------------------------------------------------
AMONG MATRIX
--------------------------------------------------------------------------------
POSITION OF ENTRY              ENTRY
--------------------------------------------------------------------------------
 1  1                          7850.63867
 1  2                          4505.79492
--------------------------------------------------------------------------------
 2  2                          2588.62744
--------------------------------------------------------------------------------
WITHIN MATRIX
--------------------------------------------------------------------------------
POSITION OF ENTRY              ENTRY
--------------------------------------------------------------------------------
 1  1                          225547.627
 1  2                          96913.3067
--------------------------------------------------------------------------------
 2  2                          57723.7197
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
WILK'S LAMBDA .956817289
--------------------------------------------------------------------------------
F-RATIO FOR MANOVA TEST FOR EQUALITY OF CENTROIDS
F= 2.14241172  WITH NUM. DF  4   AND DEN. DF   384
--------------------------------------------------------------------------------
```

22. One-way MANOVA by LIZANOVA.MUL; two variables, three groups.
(page 1 of 2 pages)

(Reduced size copy)

```
GROUP CENTROIDS
--------------------------------------------------------------------------------
GROUP 1
--------------------------------------------------------------------------------
VR 1  148.333333   VR 2  74.9166667
--------------------------------------------------------------------------------
GROUP 2
--------------------------------------------------------------------------------
VR 1  159.525641   VR 2  81.1538462
--------------------------------------------------------------------------------
GROUP 3
--------------------------------------------------------------------------------
VR 1  164.051724   VR 2  84.0172414
--------------------------------------------------------------------------------
GRAND CENTROID
--------------------------------------------------------------------------------
VR 1  157.438776   VR 2  80.0918368
--------------------------------------------------------------------------------
STANDARD DEVIATIONS
--------------------------------------------------------------------------------
VR 1  34.5964476   VR 2  17.5867589
--------------------------------------------------------------------------------
BOX STATISTIC   4.97763205
F-RATIO FOR TEST OF EQUALITY OF DISPERSION MATRICES   .816933503
NUM. DF   6   DEN. DF   649635
--------------------------------------------------------------------------------
```

22. One-way MANOVA by LIZANOVA.MUL; two variables, three groups.
(page 2 of 2 pages)

(Reduced size copy)

NAME OF STUDY REF. COOLEY AND LOHNES-PP. 250-255
NAME OF INVESTIGATOR J. Q. SMART
DATE OF RUN 4 NOVEMBER 1983

VARIABLES

INF01 INF02

EIGENVALUES

FN 1 .045002359 FN 2 1.23691585E-04

WILK'S LAMDBA .956817289

TESTS WITH SUCCESIVE ROOTS REMOVED

ROOT REMOVED	CANONICAL R	CHI SQUARE	DF	LAMBDA	PERCENT TRACE
0	0	8.49749416	4	.956817289	0
.045002359	.203002024	.0238091814	1	.999876324	99.7258977

COMMUNALITIES

INF01 .780438355 INF02 .996646428

PCT CONTRIBUTION, BY DISCRIMINANT FUNCTION

FN 1 88.85 FN 2 0

COLUMNS OF STRUCTURE MATRIX

COLUMN 1
.883424221 .998321806

COEFFICIENTS OF DISCRIMINANT FUNCTIONS-FOR STANDARDIZED OBSERVATIONS

DISCRIMINANT FUNCTION 1
.111591446 .982932709

CENTROIDS OF GROUPS IN DISCRIMINANT SPACE

GROUP 1

INF01 -.295071383 INF02 -6.40124988E-04

GROUP 2

INF01 .0612565064 INF02 1.31361629E-04

GROUP 3

INF01 .22286682 INF02 4.85539525E-04

23. An example of a small discriminant analysis problem (2 x 2) from LIZDISC.ANL. Only one eigenvalue is retained in the analysis.

(Reduced size copy)

NAME OF STUDY: MULTI-PAGE PRINTOUT FROM LIZCROSS.TAB
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 4 NOVEMBER 1983

CROSSTABULATION FROM LIZPACK DATA FILE XTABDATA/BIG

ROW VARIABLE TEST | COLUMN VARIABLE AGGRP

LEVEL	1	2	3	4	5	6	7	8	9	10	11	TOTAL
	1	6	7	6	2	10	15	7	11	10	27	
1	.9	5.2	6.1	5.2	1.7	8.7	13	6.1	9.6	8.7	23.5	115
	16.7	25	25.9	28.6	13.3	23.8	29.4	29.2	30.6	30.3	33.3	28.8
	.3	1.5	1.8	1.5	.5	2.5	3.8	1.8	2.8	2.5	6.8	
	1	2	1	0	3	3	2	0	0	1	0	
2	7.7	15.4	7.7	0	23.1	23.1	15.4	0	0	7.7	0	13
	16.7	8.3	3.7	0	20	7.1	3.9	0	0	3	0	3.3
	.3	.5	.3	0	.8	.8	.5	0	0	.3	0	
	0	0	0	0	0	0	0	0	1	0	0	
3	0	0	0	0	0	0	0	0	100	0	0	1
	0	0	0	0	0	0	0	0	2.8	0	0	.3
	0	0	0	0	0	0	0	0	.3	0	0	
	0	0	1	1	0	1	0	1	0	0	0	
4	0	0	25	25	0	25	0	25	0	0	0	4
	0	0	3.7	4.8	0	2.4	0	4.2	0	0	0	1
	0	0	.3	.3	0	.3	0	.3	0	0	0	
	1	6	7	6	2	10	15	7	11	10	27	
5	.9	5.2	6.1	5.2	1.7	8.7	13	6.1	9.6	8.7	23.5	115
	16.7	25	25.9	28.6	13.3	23.8	29.4	29.2	30.6	30.3	33.3	28.8
	.3	1.5	1.8	1.5	.5	2.5	3.8	1.8	2.8	2.5	6.8	
	1	2	1	0	3	3	2	0	0	1	0	
6	7.7	15.4	7.7	0	23.1	23.1	15.4	0	0	7.7	0	13
	16.7	8.3	3.7	0	20	7.1	3.9	0	0	3	0	3.3
	.3	.5	.3	0	.8	.8	.5	0	0	.3	0	
	0	0	0	0	0	0	0	0	1	0	0	
7	0	0	0	0	0	0	0	0	100	0	0	1
	0	0	0	0	0	0	0	0	2.8	0	0	.3
	0	0	0	0	0	0	0	0	.3	0	0	
	0	0	1	1	0	1	0	1	0	0	0	
8	0	0	25	25	0	25	0	25	0	0	0	4
	0	0	3.7	4.8	0	2.4	0	4.2	0	0	0	1
	0	0	.3	.3	0	.3	0	.3	0	0	0	
	1	6	7	6	2	10	15	7	11	10	27	
9	.9	5.2	6.1	5.2	1.7	8.7	13	6.1	9.6	8.7	23.5	115
	16.7	25	25.9	28.6	13.3	23.8	29.4	29.2	30.6	30.3	33.3	28.8
	.3	1.5	1.8	1.5	.5	2.5	3.8	1.8	2.8	2.5	6.8	
TOTAL	6	24	27	21	15	42	51	24	36	33	81	399
	1.5	6	6.8	5.3	3.8	10.5	12.8	6	9	8.3	20.3	

24. Crosstabulation; multipage printout. Page 1 of 4 pages.

(Reduced size copy)

CROSSTABULATION FROM LIZPACK DATA FILE XTABDATA/BIG CONTINUED

ROW VARIABLE TEST COLUMN VARIABLE AGGRP

LEVEL	1	2	3	4	5	6	7	8	9	10	11	TOTAL
	1	2	1	0	3	3	2	0	0	1	0	
10	7.7	15.4	7.7	0	23.1	23.1	15.4	0	0	7.7	0	13
	16.7	8.3	3.7	0	20	7.1	3.9	0	0	3	0	3.3
	.3	.5	.3	0	.8	.8	.5	0	0	.3	0	
	0	0	0	0	0	0	0	0	1	0	0	
11	0	0	0	0	0	0	0	0	100	0	0	1
	0	0	0	0	0	0	0	0	2.8	0	0	.3
	0	0	0	0	0	0	0	0	.3	0	0	
	0	0	1	1	0	1	0	1	0	0	0	
12	0	0	25	25	0	25	0	25	0	0	0	4
	0	0	3.7	4.8	0	2.4	0	4.2	0	0	0	1
	0	0	.3	.3	0	.3	0	.3	0	0	0	
TOTAL	6	24	27	21	15	42	51	24	36	33	81	399
	1.5	6	6.8	5.3	3.8	10.5	12.8	6	9	8.3	20.3	

24. Crosstabulation; multipage printout. Page 2 of 4 pages.

(Reduced size copy)

ROW VARIABLE TEST COLUMN VARIABLE AGGRP

LEVEL	12	13	14	TOTAL
	8	2	3	
1	7	1.7	2.6	115
	33.3	33.3	33.3	28.8
	2	.5	.8	
	0	0	0	
2	0	0	0	13
	0	0	0	3.3
	0	0	0	
	0	0	0	
3	0	0	0	1
	0	0	0	.3
	0	0	0	
	0	0	0	
4	0	0	0	4
	0	0	0	1
	0	0	0	
	8	2	3	
5	7	1.7	2.6	115
	33.3	33.3	33.3	28.8
	2	.5	.8	
	0	0	0	
6	0	0	0	13
	0	0	0	3.3
	0	0	0	
	0	0	0	
7	0	0	0	1
	0	0	0	.3
	0	0	0	
	0	0	0	
8	0	0	0	4
	0	0	0	1
	0	0	0	
	8	2	3	
9	7	1.7	2.6	115
	33.3	33.3	33.3	28.8
	2	.5	.8	
	0	0	0	
10	0	0	0	13
	0	0	0	3.3
	0	0	0	
TOTAL	24	6	9	399
	6	1.5	2.3	

24. Crosstabulation; multipage printout. Page 3 of 4 pages.

(Reduced size copy)

ROW VARIABLE TEST COLUMN VARIABLE AGGRP

LEVEL	12	13	14	TOTAL
	0	0	0	
11	0	0	0	1
	0	0	0	.3
	0	0	0	
	0	0	0	
12	0	0	0	4
	0	0	0	1
	0	0	0	
TOTAL	24	6	9	399
	6	1.5	2.3	

NOMINAL MEASURES

CHI SQUARE: 152.388617 DF: 143
PHI SQUARED: .381926357
CONTINGENCY COEFFICIENT: .525711364
TSCHUPROW'S T: .0516799157
LAMBDA (ROW DEP.): 3.52112676E-03
LAMBDA (COL. DEP.): .0471698113
SYMMETRIC LAMBDA: .0265780731
CRAMER'S V: .186334586
GOODMAN-KRUSKAL TAU A: .0189291239
GOODMAN-KRUSKAL TAU B: .0306513319

ORDINAL MEASURES

KENDALL'S TAU A: -.0417878868
KENDALL'S TAU B: -.0509460147
KENDALL'S TAU C: -.045472533
WILSON'S E: -.0429464529
GAMMA: -.0619816185
SOMER'S D (ROW DEP.): -.0465311961
SOMER'S D (COL. DEP.): -.0557797055
SOMER'S SYMMETRIC D: -.0507374361

PARTIAL GAMMA: -.0619816185

24. Crosstabulation; multipage printout. Page 4 of 4 pages.

(Reduced size copy)

NAME OF STUDY: FORCASTING IMPORTS DATA (KENDALL-P.144)
NAME OF INVESTIGATOR: J. Q. SMART
DATE OF RUN: 30 JANUARY 1984
RUN FROM LIZPACK DATA FILE: IMPORTS/DAT

ANOVA TABLE FOR REGRESSION

SOURCE	DF	SS	MSS	F RATIO
REGRESSION	4	2068417.67	517104.416	239.425751
ERROR	35	75591.93	2159.76943	
TOTAL	39	2144009.59		

PARTIAL REGRESSION COEFFICIENTS

LAG. #	COEFFICIENT, B	STD ERROR	DF	T (FOR B=0)
4	.0542283959	.185832477	35	.291813341
3	.0204812471	.211748323	35	.096724483
2	.469826413	.197849125	35	2.37467016
1	.503741226	.172453577	35	2.9210251

Y-INTERCEPT -51.7799216

STANDARD ERROR OF ESTIMATE: 46.4733195

PERIOD	OBSERVED Y	FORECASTED Y	RESIDUAL
5	1457	1449.35696	7.64304543
6	1403	1463.95221	-60.952209
7	1389	1446.6132	-57.6131978
8	1379	1414.90471	-35.9047089
9	1408	1403.21408	4.7859211
10	1426	1409.90924	16.0907602
11	1460	1431.63754	28.3624625
12	1442	1457.27329	-15.2732868
13	1414	1466.12133	-52.1213288
14	1472	1445.23217	26.7678275
15	1520	1462.76913	57.2308731
16	1540	1512.64905	27.3509484
17	1611	1544.94506	66.0549388
18	1612	1594.23556	17.7644363
19	1632	1631.10957	.890431881
20	1659	1644.19296	14.8070445
21	1581	1671.06119	-90.061194
22	1643	1644.91855	-1.91854477
23	1672	1641.1416	30.8583975
24	1686	1684.74597	1.25403547
25	1722	1702.46333	19.5366693
26	1681	1731.1317	-50.131701
27	1726	1729.25142	-3.25142288
28	1642	1734.15342	-92.1534176
29	1777	1714.09383	62.9061661
30	1787	1741.33177	45.6682272
31	1779	1810.51561	-31.515604
32	1850	1809.39372	40.6062789
33	1948	1848.92638	99.073617
34	1903	1932.02913	-29.0291324
35	1945	1956.42411	-11.4241071
36	1937	1962.29643	-25.2964282
37	1992	1982.39193	9.60806561
38	1980	2004.75903	-24.7590246
39	1966	2026.66833	-60.6683254
40	2024	2014.67067	9.32932664
41	2026	2040.04688	-14.0468817
42	2130	2067.36682	62.6331825
43	2078	2121.12427	-43.1242724
44	2197	2146.97789	50.0221148

25. Autoregression and Simulation from LIZCORR.LAG

(Reduced size copy)

SHOWPACK and DATAPACK

SHOWPACK 1.0

Programming and Manual by Michael Dudgeon
Crimson Software

SHOWPACK 1.0 is a program written to help create, label, and edit hi-resolution graphics screen displays. Graphical aids are provided for plotting data from tables, drawing with joysticks, and creating pie charts and bar graphs. Graphs can be plotted using data tables which have been stored on disk in Coco 5-byte floating point format. High-res screens can be saved and loaded from disk, and a slide-show mode exists for stepping through and displaying a series of graphs or charts stored on disk. SHOWPACK provides two sizes of hi-res letters for typing text and labels onto the screen, and the cursor can be moved in small enough steps to have properly labeled axes.

Sample Run

SHOWPACK is on Diskette 4 of LIZPACK, and may be loaded by LIZPACK DIRECTORY just as if it were a LIZPACK program. To directly load and run SHOWPACK, proceed as follows: let N be the disk drive number from which you wish SHOWPACK to load. Type POKE 32767,N<ENTER>, and then type RUN "SHOWPACK:N"<ENTER>. The number N is 0, 1, 2, or 3. You will be asked whether this is a COLD START (Y-N)? If you have just loaded SHOWPACK off the disk, type <Y><ENTER> else just hit <ENTER>. The Y tells the computer to load a machine language program which must be in memory to produce text on the hi-res screens. Since you just loaded SHOWPACK, type Y. The hi-res screen will appear and display whatever is randomly in memory at the time. In order to clear the screen hit <CLEAR><W>. You will be asked to verify the clear command. After long work on a screen one missed key should not destroy the effort. Note that the <CLEAR> key prefaces all major commands. The screen will be cleared to the color white which is the default background color. The default foreground color is black.

Now just type and 8x8 dot capital letters will appear. The arrow keys move the cursor and automatically repeat if held down. <SHIFT><0> unlocks the shift and allows you to type lower-case letters. <CLEAR><R> reverses the characters, making them white on black. <CLEAR><S> gives you small letters (42/line) while <CLEAR><B> gives you big ones (24 lines of 32 characters/line). After you finish experimenting, type <CLEAR><W> to clear the screen. Try using <CLEAR><M> for a display of a MENU of SHOWPACK commands.

Now use the arrow keys to move the cursor to the approximate center of the screen. Type <CLEAR><G> to enter the GRAPHICS MODE. You will be asked to choose (1) PIE CHART, (2) BAR GRAPH, (3) PLOT DATA FROM DISK, (4) DRAW WITH JOYSTICK, (5) DRAW LINES, or (6) ESCAPE. Type <1> and <ENTER>. You will now be be given prompts to help in making a pie chart to fit your specifications. For RADIUS OF GRAPH type 50. You will then be asked about the center. Since you previously left the cursor near the center of the screen, type 0,0. Any response other than 0,0 is taken as the X,Y location of the pie's center in the screen's 256 by 192 grid coordinates. Next you will be asked how many sections (slices) are in the pie. Pretend there are 5 categories you wish to describe so type 5. Then from S1 to S5 you will be asked for the percentage for each section. For these percentages type 20, 25, 15, 30, and 10. After a slight delay the graphics screen will reappear with the pie. Using the arrows you can move the cursor around and type labels for the pie chart. If you find it necessary to use smaller letters, type <CLEAR><S> for the 5x7 letters. If you want to fine adjust the vertical position of the cursor, hit <CLEAR>< > or <CLEAR>< > to move the cursor up or down one row.

After you have labeled and prettied up your graph hit <CLEAR><F>. Flash! The screen is FLIPped! Whatever was black is white and vice-versa. Type <CLEAR><F> again to flip it back. You will find that certain displays look better one way and others look better the other.

Now insert a scratch disk of your choosing. Type <CLEAR><D> to access the DISK COMMANDS. Type 2 for a SAVE of this screen. For the filename enter DEM01. The screen will now be saved to disk. When the graphics screen reappears, type <CLEAR><W> to wipe it clean. Then type <CLEAR><D> to get back to the DISK COMMANDS. Choose menu option (1) for LOADing hi-res screens. For the filename type DEM01. The screen you just created will be loaded and displayed.

Clear the screen again. Hit <CLEAR><G> and select option (2) for the BAR GRAPH. You will be asked for the minimum value of the range you want on your graph. Suppose you want to have a vertical scale from 0-100, so type 0. You will then be asked to enter the maximum vertical value . Type 100. For the vertical axis increment use 10. That means each horizontal line in the graphical grid will be 10 more than the one below it. The lines in this graph will represent 0, 10, 20,..., 100. You will then be asked for the number of bars on the graph. This will be the number of categories you want to have displayed on the graph. Type 4 for this example. You will then be asked for the values for each of the four bar entries. Let's say that you are charting the sales of "superitems" in thousands from January through April. For entry #1 or January you sold 56 thousand. Type 56. For #2 or February you sold only 8 thousand. Type 8. In March or #3 you sold a whopping 97 thousand. Type 97. For 4 or April type 61. Now the computer will make the bar graph. By using the arrows and <CLEAR> arrows for fine positioning, label the horizontal lines from 0 to 100. There is one extra line on top to provide a less cramped looking graph. Label the four categories which have the orange or blue bars Jan, Feb, Mar and Apr. Add any titles you wish. You may want to FLIP the image to make the result look cleaner. If you want, you can also save this screen to disk.

General Information

*The angle symbols, < and >, enclose a key to be used in selecting an option and are not to be typed.

*When you are viewing the graphics screen, an option is selected by pressing the <CLEAR> key followed by the letter or numbers for the option.

*When you are in a menu screen mode and there is a flashing cursor, you must hit <ENTER> to register your response.

*Letters will not appear correctly in PMODE 3. PMODE 3 is for graphs where control of color selection is necessary.

*The <CLEAR><P> option dumps the graphics screen to an Epson printer. The printer driver program starts at line 1000 and needs to be replaced by one catered to any non-Epson printer that you plan to use .

*The <ENTER> key acts as a down arrow and returns the cursor to the left-hand side of the screen.

*After drawing pie charts or making pictures in PMODE 3, you can switch to

PMODE 4 and use letters to label them. Yellow and blue come across as blue or orange while magenta and cyan come across the same. An example would be to create a pie chart as blue on red in PMODE 3, transfer it to PMODE 4 where it would be a blue/orange pie on a white background.

*Cursor arrows have an automatic repeat feature when held down.

*SHOWPACK uses memory locations from 30713-32767 for its machine language routines.

*Use DISK menu option (8) for creating a data table for use in plotting points.

Description of Commands

<CLEAR><A> EXIT SHOWPACK. Loads LIZPACK DIRECTORY. You must have executed LIZPACK.SET prior to using SHOWPACK. If a LIZPACK program fails to operate properly after using SHOWPACK, you will have to turn off the computer, turn it back on, and rerun LIZPACK.SET.

<CLEAR><B> Puts the computer into a mode for writing 8x8 dot alphanumeric characters to the graphics screen. The letters are sized to produce up to 24 character lines with 32 letters per line.

<CLEAR><C> The COLOR command allows the user to specify the PMODE and foreground and background colors. The default PMODE is 4 and the default colors are black for foreground and white for background. Input PMODE and then the names of the foreground and background colors. (BLACK, GREEN, YELLOW, BLUE, RED, WHITE, CYAN, MAGENTA, and ORANGE). The computer only needs the first three letters of each color.

<CLEAR><D> DISK OPTION menu:

(1)-LOAD SCREEN:Enter the filename of the screen you want loaded from disk. If you do not give an extension, /BIN is assumed. GRAPH1 and GRAPH2 are sample graphs which are stored on the SHOWPACK disk.

(2)-SAVE SCREEN-Enter filename of screen you wish to save. Extension is /BIN if not specified

(3)-LOAD & PLOT TABLES: This option assumes that a table containing the X,Y (horizontal, vertical) coordinates of the data points has been created and stored in a disk file. The disk file can be created by LIZPACK, DISK option (8) in SHOWPACK, or a program of your own creation. The computer requests the name for the data file containing the X-Y values you wish to plot. /DAT is the assumed extension if none is given. Your next choice is whether new graphical axes are to be plotted.

For the NO AXIS case: a) the computer scans the data and scales the plot to fill the area bounded by a rectangle four units from the edge of the screen, b) the data is plotted without tracing out new axes, c) you are asked whether the file is (D)irect access or (S)equential. If the file is direct access, you will also be asked to specify the length of the record (10 is standard). The sequential files must be stored on disk as X,Y paired coordinates.

For the PLOT AXIS case choose whether the USER or COMPUTER is to define the axes. The COMPUTER generated axes break the screen into quarters and the plot is scaled to nearly fill the screen. The USER defined axes require that you know your data ranges well enough to answer prompts asking for upper and lower X and Y axis limits and data values for the spacing between axis "tick" marks. User specified axes are advantageous for multiple plots overlayed on the same graph.

In writing a program to create a direct access data table file, use 10 bytes of data for each record entry: 5 bytes for the MKN$ of X and 5 bytes for the MKN$ of Y. A direct data file called DIRDATA1 with a record length of 10 is provided on disk for testing out the plot option.

(4)-SET UP SLIDE SHOW: This option allows you to tell the computer the names of up to 99 screens stored on disk you wish to be able to display with three keystrokes. First enter the number of screens you want to set up. Then in order type in the filenames of the screens. Once the last screen is defined you will be asked for a filename to designate the slide show package. The list of slides is saved to disk under this name and is used in DISK option (5) to load and run a slide show.

(5)-LOAD SLIDE SHOW asks for and loads the slide show file. The list of slides is read from disk and stored in memory. In addition, the first screen is automatically loaded into a buffer space in memory. Once a show is loaded, the <CLEAR> key plus a two digit number will access these screens. If you type <CLEAR><0><2> the, second screen will be loaded onto the screen and screen 3 will be moved to the hidden display buffer. When you type <CLEAR><0><3> screen 3 will instantly appear because it is already in memory. <CLEAR>< > displays the next screen in the sequence while <CLEAR>< > loads and displays the previous screen. This allows sequential display of screens with minimal disk access time and should prove valuable in formal presentations. In addition, the <CLEAR> <right arrow> steps to the next slide without you having to remember a slide's number. <CLEAR> <left arrow> steps backward through the list of slides. SHOW1 on LIZPACK Diskette 8 (Practice Data Files) is a sample slide show with four graphics pages.

(6)-INSTALL TEMPORARY SLIDE SHOW FILENAMES: This option allows you to put a list of display filenames in memory, but does not save the list to disk. Once installed the slides can be called using the <CLEAR> key followed by its two digit number (method described in option 5). Enter the filenames (/ BIN extensions assumed) for the files you wish to add to the display list.

(7)-OVERLAY A SCREEN: This option allows you to overlay a screen from disk on top of the existing screen. For this option to work, screens must have white backgrounds and black foregrounds because the computer overlays the black parts of the picture.

(8)-CREATE/SAVE DATA TABLE: This subprogram is for generating DATA TABLEs to be used in plotting graphs using DISK MODE or GRAPHICS option (3). Once a data table has been saved on disk using option (8), it can be used for plotting graphs or drawing crude pictures. The user is asked for the X,Y values of the data points on the graph where Y is the vertical coordinate and X is the horizontal coordinate of a point.
*** Note the following differences: +Y is up for the plot option while +Y is down for CoCo ECB line commands.***.
PLOT allows you to enter data in a more natural form with + being up and - down. The plot routine allows variables to have both positive and negative values and has an option which scales the axes to compensate for the range of X and Y. Data entry is terminated when a response of <END,0> is given. Table data is then saved to disk in a direct access data file with a record length of ten and an extension of /DAT. Option (3) in either the DISK or GRAPHICS MODE can be used to load the data table file and plot the points. Note that the PLOT routine is such that straight lines are drawn between adjacent points

in the table list. This means that the order of entering data table points is important, and the option serves equally well for connect-the-dot drawing and graph plotting.

<CLEAR><F> FLIPs the screen by doing a machine language bit reversal of the graphics screen i.e. black becomes white, white becomes black.

<CLEAR><G> GRAPHICS OPTIONS:

(1)-PIE CHART: First enter the radius of the circle or pie. Then enter the X and Y coordinates of the center of the pie (separated by a comma). The PMODE 4 (hi-res) screen coordinates are for a 256 by 192 grid with the origin in the upper lefthand corner. PMODE 3 has a grid resolution of 128 by 192 with four colors. If you want the center of the pie to be where the cursor was positioned when you left the graphics screen, then enter 0,0. Next enter the number of sections or slices for the pie. Then as prompted, enter the percentage of the pie to be occupied by each of the sections. For 50% type <5><0><ENTER>. The graph will be printed in the foreground color. (See <CLEAR><C>)

(2)-BAR GRAPH: The computer will clear the screen and plot a bar graph in PMODE 4. First enter the minimum vertical axis value you want on this graph. If your bars range from 500-1000, you would type 500. This number represents the bottom axis line. Next enter the maximum vertical value for the range. On our sample graph from 500-1000 type 1000. Then type in the vertical-axis increment. This means each line will be displaced that much from the one below it. A minimum value of 500, a maximum of 1000, and an increment of 100 will give you lines representing 500, 600, 700,..., 1000. Then enter the number of bars you want on your graph. For each of the bars enter its value or magnitude in the scale range you just set up. The BAR GRAPH command will clear the screen black and draw in white, orange, and blue.

(3)-PLOT DATA FROM DISK: This is identical to option (3) in the DISK COMMANDS. Enter the disk filename for the data you wish to plot. Then respond to the second prompt question with <D> or <S> depending whether the file is direct or sequential. Finally specify the record size noting that 10 is standard since the CoCo floating point format uses five bytes for the x-axis value and five bytes for the y-axis value of a data point.

(4)-DRAW WITH JOYSTICK: This option is probably as much for entertainment as for serious graphics, but who can tell when it might prove useful. The right joystick is used as a pen. The <spacebar> or <down arrow> lowers the pen and puts the cursor in the drawing mode. The <up arrow> lifts the pen for repositioning. The <W> key opts for wiping the screen clear to obtain a black background. Once the drawing session is completed, the <Q> key can be used to quit and return to the main graphics screen mode where text can be added and the other SHOWPACK options are available.

(5)-DRAW LINE OR AXIS: The DRAW LINE option gives the user easy access to the line drawing command in ECB. The user will be asked for the X,Y coordinates for the starting point and end point defining the line, and then SHOWPACK will plot the line. The coordinates are those used by the CoCo, namely 0,0 is in the upper left corner of the TV screen, X is the horizontal axis value ,and Y is the vertical location (measured from the top of the screen). SHOWPACK lines 3070-3080 can be modified if other features are to be added to DRAW LINE.

(6)-ESCAPE from this menu back to the graphics screen.

<CLEAR><H> Change the HUE of the screen. This command flips the color set of the current PMODE. White trim becomes green and vice-versa. It is switching between SCREEN 1,0 and SCREEN 1,1.

<CLEAR><M> Displays a MENU of the major SHOWPACK commands.

<CLEAR><P> PRINT screen. The SHOWPACK program currently has a graphics print routine for an EPSON FX80 at line 1000. The program at line 1000 should be replaced with one for your printer. Any print driver machine language code that you add must reside in memory lower than decimal 30713. Make sure to change the CLEAR statement at the beginning of SHOWPACK to protect your print driver routine.

<CLEAR><R> REVERSEs letters. Letters become black on white if they were white on black or vice-versa.

<CLEAR><S> Engages the SMALL letter mode. Letters are now 5x7 (42 per line).

<CLEAR><W> WIPES screen. This command clears the screen to the background color.

<CLEAR><two digit number> Loads in screen N as defined in the SLIDE SHOW list (See <CLEAR><D>, options (4)&(5) -SLIDE SHOW). Screen N+1 is also loaded into a hidden buffer for quick display. This command is ignored if the slide show list has not been loaded.

<CLEAR>< > Shifts the cursor up one picture element row.

<CLEAR>< > Shifts the cursor down one row.

<CLEAR>< > Loads in the next screen of a slide show.

<CLEAR>< > Loads in the slide show screen whose number is one less than the one currently being displayed.

<SHIFT><0> Locks and unlocks the shift key to put the keyboard in and out of the capital letter lock mode.

List of Command Functions

1. MENU: <clear><M>
2. DISK OPTIONS: <clear><D>
load and save screens, plot data created by LIZPACK, setup a slide show, create a data table for plotting.
3. GRAPHICS OPTIONS: <clear><G>
pie chart, bar graph, plot tables, draw lines, draw with joystick.
4. LOAD or CREATE SLIDE SHOW: access through DISK OPTIONS <clear><D>
5. PLOT or CREATE DATA TABLE: see DISK OPTIONS.
6. BIG LETTERS: <clear><B>
7. SMALL LETTERS: <clear><S>
8. HUE: <clear><H> changes hue.

9. FLIP: <clear><F> inverts the polarity of the screen image. Black becomes white and vice versa.
10. WIPE: <clear><W> clears the screen to the background color.
11. PRINT: <clear><P> sends the image on the screen to a graphics printer.
12. COLOR: <clear><C> allows changes in PMODE and colors.
13. REVERSE: <clear><R> reverses the polarity of the letter signals-white becomes black and black become white.
14. LOAD LIZPACK DIRECTORY: <CLEAR><A>

Demo Files on Disk

The following is a list of sample files that can be used to try out some of the SHOWPACK commands:
1. GRAPH1, GRAPH2, GRAPH3, and GRAPH4 are screens that can be LOADed and examined or can be used in CREATing a SLIDE SHOW.
2. GINNY, DIRDATA1, and DIRDATA2 are direct data tables which can be used with the Disk or Graphics options for PLOTting DATA tables.
3. SEQDATA is a sequential data table file that can be used with the disk PLOT DATA option.
4. SHOW1 is slide show that can be installed with the disk LOAD SLIDE SHOW option.

All Demo files are on the diskette containing LIZPACK practice data files, diskette 8.

CRIMSON SOFTWARE
5 Bellview Drive
Tuscaloosa, AL 35405

DATAPACK

Programmed by Michael Dudgeon
Manual by William J. Gray

DATAPACK provides a means for converting LIZPACK data files to ASCII files and vice versa. The conversion process involves the processing of data one character at a time, and is consequently quite time consuming.

DATAPACK is located on Diskette 4 of LIZPACK and may be loaded as if it were a LIZPACK program.

After loading the program, you may select one of the two options which are described below.

OPTION 1: ASCII TO LIZPACK
OPTION 2: LIZPACK TO ASCII

For either option, you must specify whether the ASCII file is

1. space delimited,
2. FORTRAN format, or
3. custom designed.

If the file is in FORTRAN format, you must specify the FORTRAN format in which the file was written. The format statement may not exceed 80 characters. If the file is custom designed, you must specify the record length, number of slots for each record, and the length of each slot. You must specify whether the data is integer or floating point. If it is floating point, you must specify the number of digits to the right of the decimal point.

You must provide the name of the source and destination files. The destination file is then assembled, and control is returned to the first OPTIONS MENU with the variable tables cleared.

DISK 1 (FREE 4)

	LIZSAMP	. FRQ	5
*	LIZ PACK	. SET	1
*	LIZ PACK	. DIR	2
	LIZ DESC	. DES	4
	LIZANOVA	. NEQ	3
	LIZ 2WAY	. TWO	4
	LIZ ANOVA	. RAN	2
	LIZ 1WAY	. ONE	3
	LIZ MEAN	. MCP	6
	LIZ BLOCK	. 2WY	4
	LIZ 2WAY	. ANC	4
	LIZ ANCOV	. BLK	4
	LIZ 1WAY	. ANC	4
	LIZ ANCOV	. RAN	3
	LIZ ANAL	. RES	7
	DELETE	. VAR	1
	LIZ POLY	. RES	2
	TRANSPAK	. REG	2
*	EDITOR	. BIG	3

DISK 2 (FREE 1)

*	EDITOR	. BIG	3
*	LIZPACK	. SET	1
*	LIZPACK	. DIR	2
	LIZRAD	. PWR	4
	SCATTRGM	. BIN	1
	TRANSPAK	. LPK	2
	LIZANAL	. PLY	6
	LIZLOG	. ANL	4
	FILEASSM	. AFM	2
	FILEMOD	. CFM	1
	LIZPOLY	. ONE	4
	LIZPOLY	. STP	4
	LIZEXPO	. ANL	6
	LIZANAL	. MUL	5
	LIZ CORR	. XTM	4
	FILEPROD	. AFM	2
	LIZENCO	. LPK	2
	LIZEXPO	. ONE	4
	MERGFILE	. AFM	1

DISK 3 (FREE 3)

	LIZANRAN	. AGG	3
	LIZANOVA	. 3WY	4
	LIZ3WAY	. NEQ	4
	LIZMEAN	. AGG	6
	LIZ1WAY	. AGG	3
	LIZ1ANC	. AGG	4
*	LIZ PACK	. DIR	2
*	LIZ PACK	. SET	1
	LIZ BLOCK	. AGG	4
	LIZ BLANC	. AGG	4
	LIZ RAND	. AGG	2
*	EDITOR	. BIG	3
	LIZSAMP	. AGG	5
	LIZ 2WAY	. AGG	4
	LIZ 2ANCV	. AGG	4
	LIZ2NEC	. AGG	3
	LIZ 3WAY	. AGG	4
	LIZ3NEQ	. AGG	5

DISK 4 (FREE 22)

*	LIZ PACK	. SET	1
*	LIZPACK	. DIR	2
	LIZ CORR	. MAT	3
	LIZ COMP	. PCL	4
	LIZ FACTR	. ANL	4
	LIZ QUART	. MAX	5
	LIZ VARI	. MAX	5
	LIZORTHO	. MAX	5
	LIZ OBLI	. MIN	4
	LIZSCORE	. FAC	3
*	EDITOR	. BIG	3
	SHOWPACK	. BAS	4
	SHOWPACK	. BIN	1
	DATA PACK	. BAS	2

DISK 5 (FREE 44)

	LIZ CANON	. COR	5
	LIZ CLASS	. GET	2
	LIZ DISC	. ANL	5
	LIZ ANOVA	. MUL	4
	LIZ ANCOV	. ML1	3
*	LIZ PACK	. SET	1
*	LIZ PACK	. DIR	1
*	EDITOR	. BIG	3

DISK 6 (FREE 28)

	LIZ DESC	. AGG	4
*	LIZ PACK	. SET	1
	LIZ CROSS	. TAB	6
*	LIZ PACK	. DIR	2
	LIZ3CORR	. MCP	3
	LIZ WILCO	. XON	3
	LIZKRUSK	. WAL	3
	LIZ2WAY	. FRD	3
	LIZ SPEAR	. KEN	3
*	EDITOR	. BIG	3
	LIZ2CORR	. MCP	3
	LIZ1CORR	. MCP	3
	LIZ4CORR	. MCP	3

DISK 7 (FREE 6)

LIZ TIME	.SRI	4
LIZ MOVE	.AVE	7
LIZ CORR	.LAG	6
LIZ NLIN	.EST	2
LIZ 2 TIME	.SR2	4
LIZ TRIZ	.SRS	3
LIZ 2 FORE	.EST	5
LIZ 1 FORE	.EST	3
* LIZ PACK	.DIR	2
LIZ 3 FORE	.EST	5
LIZ NON	.LIN	4
LIZ PLOT	.SRS	3
TRANSPAK	.SRS	1
LIZTRIG	.PLY	1
LIZENCO	.SRS	1
LIZ ANOVA	.SRS	2
LIZ 4 FORE	.EST	5
* LIZ PACK	.SET	1
* EDITOR	.BIG	3

DISK 8 (FREE 9)

FEV	.DAT	1
QUET	.DAT	1
DOSAGE	.GMS	1
SOYBEAN	.DAT	1
PIGS	.WT	1
MEDCARE	.DAT	1
OATS	.DAT	1
TOBTEST	.DAT	1
WEIGHT	.PIG	1
BACON	.DAT	1
LIMAS	.DAT	1
QUETSBP	.DAT	1
AIRLINE	.DAT	1
RATION	.DAT	1
PETPROD	.DAT	1
BIGFILE	.AGG	1
BUTTERFAT	.NEW	1
BIGMESS	.GOB	2
TEXTILE	.DAT	1
QUETSMK	.DAT	1
IMPORTS	.DAT	1
USWHITE	.POP	1
GRAPH1	.BIN	3
SHEEP	.DAT	1
VEGPRICE	.DAT	1
USBLACK	.POP	1
SOCIOEC	.VAR	1
CONSUMER	.DAD	1
TENDOM	.RDG	1
FEVAGG	.DAT	1
XTABDATA	.BIG	2
GRAPH2	.BIN	3
MEDCARE	.AGG	1
LIMAAGG	.DAT	1
LONGLEY	.DAT	1
DOSAGE	.AGG	1
DRUGABS	.DAT	1
BACON	.AGG	1
GRAPH3	.BIN	3
GRAPH4	.BIN	3
SHOW1	.S-S	1
DIRDATA1	.DAT	1
DIRDATA2	.DAT	2
SEQDATA	.DAT	1
GINNY	.DAT	1

TOLERANC	.DAT	1
SHOOT	.DAT	1
LEGUME	.DAT	1
LEGUME	.AGG	1