LDA

Software System for Longitudinal Data Analysis

Emet D. Schneiderman & Charles J. Kowalski

August, 1993
Longitudinal Data Analysis

Release Notes

Version 3.2

Emet D. Schneiderman
Department of Oral and Maxillofacial Surgery
Baylor College of Dentistry
3302 Gaston Avenue, Dallas, Texas 75246, USA

Charles J. Kowalski
Department of Biologic and Materials Sciences
University of Michigan School of Dentistry
Ann Arbor, MI 48109-1078, USA

August 1993

Supported by DE08730 from the National Institute for Dental Research
# Table of Contents

I. Introduction ................................................. 1
II. System Requirements ...................................... 1
III. Installation .................................................. 2
IV. Program Descriptions ..................................... 2
V. Data Files .................................................... 7
VI. Running the Programs .................................... 8
VII. Examples ................................................... 9
VIII. Output ..................................................... 11
IX. Acknowledgement and Citation .......................... 11
X. User Assistance ............................................. 11
Reference Update:

References have now been published for the following programs. Please see the program description section beginning on page 2 of the Release Notes Version 3.2; the numbers below correspond to the listed programs.


LDA -- LONGITUDINAL DATA ANALYSIS: A PC PACKAGE FOR GROWTH CURVE ANALYSIS AND OTHER LONGITUDINAL STATISTICS WRITTEN IN GAUSS

Emet D. Schneiderman, Charles J. Kowalski, Stephen M. Willis, Thomas R. Ten Have, Amy M. Purey and Ingrid Y. Guo

VERSION 3.2 RELEASE NOTES, AUGUST 4, 1993

I. INTRODUCTION

Welcome to the Longitudinal Data Analysis System in GAUSS. This large set of statistical procedures includes (1) single and (2) multi-sample analyses of growth curves, (3) tracking analyses, (4) computation of orthogonal and orthonormal polynomials, (5) imputation of missing values, (6) prediction of future or unknown values, and (7) diagnosis on the basis of longitudinal observations. Several of the recent programs are ideally suited to the analysis of irregular data sets. These programs have been written to be user-friendly and require only the most basic of computer skills. If you can use a word processor to build a data file, you can run these programs. The majority of these programs are driven by a full screen color menu and the user is prompted for all program input. For these programs all input is filtered -- inappropriate responses result in the user being prompted to re-enter the response. Each program processes the data, creates an output file containing the longitudinal statistics generated, and produces presentation quality graphics on the monitor. These high quality graphics can also be printed, plotted or saved for subsequent manipulation. A few of the programs use an interactive rather than menu-driven format.

II. SYSTEM REQUIREMENTS

This greatly expanded and improved package has been written and compiled using GAUSS386i, Version 3.1. It therefore requires an 80386 or 80486 based personal computer (PC) running the MS-DOS operating system (version 5.0 or higher is recommended, although versions as low as 3.3 will suffice). 80386 computers MUST also be equipped with 80387 numerical coprocessors. At least 4 mb of memory are required, and must be available to GAUSS386i, i.e., not in use by memory resident programs (ISRs) such as Windows or SideKick. At least 5 mb of hard disk space are needed to store the decompressed programs and ancillary files. EGA/VGA graphics
are required to display the color graphics; VGA or SVGA is recommended to display optimally the graphic results. Runtime modules are supplied with the programs so that no additional software (i.e. compiler or interpreter) is required to run these programs. One can create and edit ASCII data sets for use by these programs by using the full-screen editor (EDIT) supplied with MS-DOS version 5.0 or with most popular word processing programs. The programs require no additional installation or modification and are each run with a single command.

An older version of this package (version 2.0) is available to those using IBM-PC/XT/AT, or other 8086/80286 based PCs. It includes the first seven programs provided in the current package (see list below). The appropriate numerical coprocessor, either a 8087 or 80287, is also required to run this version. We regret that we can no longer support this earlier package in addition to the more powerful current version.

III. INSTALLATION

Installation of the programs takes about 5 minutes. To install the programs, place Disk #1 in your drive; if this is the A drive, type A: <return> (if it is the B drive, type B:<return>). At the A>(or B>) prompt, type install <return> and then follow the instructions given by the installation program. You may abort the installation program at any time by pressing ESC. The installation program will place all of the package in a directory of your choice. The default is C:\LDA. You may rename this directory at this time or any other time. This installation process will not alter either of your system files, AUTOEXEC.BAT or CONFIG.SYS.

In order to produce color graphics and drive a printer with your particular hardware, you must modify a file we supply named PQGRUN.CFG. However, before modifying this file, please make a backup copy. Using the MS-DOS editor or a word processor, follow the instructions contained in the beginning of PQGRUN.CFG to select the particular video monitor, printer or plotter that you are using. Then, carefully make the changes, and save the file as an ASCII file.

IV. PROGRAM DESCRIPTIONS

For selecting a data set (or sets) for analysis, most of the programs include an automatic feature for searching the disk and displaying all of the available data files. Data files may be in GAUSS or ASCII format (see below). For most of the programs, data files may have up to 250 subjects with 25 time points. A few of the computer-intensive programs are limited to 150 subjects (noted below). Programs using polynomials can compute equations up to degree 21. Generally, however, the number of subjects must exceed the maximum degree of the polynomial used. Time points may be either evenly or unevenly spaced, and fractional time points are allowed. Programs that permit the consecutive processing of data sets will accept up to 10 data files.
We have at any given moment several programs under development. As they are completed and tested, we will add them to the distribution disks. Current users will be notified from time to time about updates. The following programs are currently available (names of older versions shown in parentheses):

1. **RAO.GCG** (GCA7.GCG)  

2. **KTRK.GCG** (TRACKING.GCG)  

3. **FDTRK.GCG**  

4. **FDTRK2.GCG**  

5. **HILLS.GCG** (HILLS4.GCG)  

7. **2STG.GCG** Two-stage polynomial growth curve model. Similar to RAO above, except instead of being arbitrary, the covariance matrix Σ has the special structure where $Σ = WAW^T + σ^2 I$. The program tests whether the two-stage model adequately fits the sample, and if so, determines the AGC and confidence intervals, often yielding more precise estimates and narrower confidence bands than RAO. See Ten Have TR, Kowalski CJ and Schneiderman ED (1991) *PC Program for Analyzing One-Sample Longitudinal Data Sets Which Satisfy the Two-Stage Polynomial Growth Curve Model*. Amer. J. Human Biol. 3:269-279.


19. **ZBANDS.GCG**† Allows the computation of confidence bands for average and individual growth curves generated by our program **ZCOEFF.GCG** (see above). Confidence bands for the individual can be used for diagnosis and prognostication. Suitable for irregular observations with missing data. Limited to 150 subjects. Group indicators allowed. See Schneiderman ED, Kowalski CJ, Willis SM and Guo IY. *Confidence Bands for Average and Individual Growth Curves*. Submitted for publication, Computers in Biology and Medicine.


22. **MC.GCG**† Computes McMahan’s tracking indices that reflect individuals’ maintenance of relative position within the distribution of a measurement as that distribution changes over time. Based on the two-stage model (see above). The jackknife is used to compute the standard errors of the indices. See Guo IY, Ten Have TR, Kowalski CJ, Schneiderman ED and Willis SM. *A PC Program for Computing McMahan’s Tracking Indices from One-Sample Longitudinal Data Sets*. In Press, Int. J. Biomed. Comput.

* Denotes interactive programs.
† Group indicators allowed or required in data set
V. DATA FILES

Data files of two different formats can be processed by our programs. These are GAUSS and ASCII. The GAUSS format is created using the enclosed GAUSS utility, ATOG.EXE. When using this format, the dimensions of the data file are determined by our programs. When using the ASCII format, the dimensions of the data file must be supplied explicitly by the user.

To create an ASCII data file, simply enter the data (a single response variable measured at several times) into a file using an editor or word processor. Each row is a subject and each column is a time point. For some of these programs, data must be complete. Other programs allow missing data (see above) in which a missing datum is indicated by a period (.). Each column of data must be separated from the next by at least one space. With the exception of data sets for programs that allow or require group indicators (indicated above with t), data files should not contain any extraneous information such as ID numbers, dates etc. Be sure your word processor does not enter any hidden control characters (non-ASCII) in the file. Most word processors have a non-document or ASCII text file mode. This is the mode you want. Below are two examples of typical ASCII data files. Type A is suitable for all of the single sample procedures (most of the earlier programs). Type B is suitable for the programs based on Zerbe's work (ZCOEFF, ZDIST, ZDIAG, ZBANDS) and the Randomization tests (ZRTA and ZRTE). This latter format includes a column containing group indicators. These indicators may be any one column to the left or right of the data itself and must be consecutive integers beginning with 1. The response variables must be in contiguous columns. Type B, with no missing data is suitable for the Potthoff-Roy procedure (PR).

A. Complete data file with no missing values and no group indicators:

```
21.32 34.63 56.76 67.22 13.41
20.33 14.54 20.61 21.10 24.44
9.00 10.15 12.20 12.25 12.29
23.34 43.35 63.36 93.01 93.33
10.33 11.18 12.41 13.09 14.01
```
B. Data file with missing values (indicated by ",") and group indicators in the first column:

<table>
<thead>
<tr>
<th></th>
<th>39.4</th>
<th>41.9</th>
<th>42.5</th>
<th>43.8</th>
<th>45.7</th>
<th>46.9</th>
<th>47.6</th>
<th>48.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>34.3</td>
<td>36.8</td>
<td>41.2</td>
<td>42.5</td>
<td>43.8</td>
<td>45.7</td>
<td>46.3</td>
<td>47.6</td>
</tr>
<tr>
<td>2</td>
<td>41.0</td>
<td></td>
<td>42.3</td>
<td>43.1</td>
<td>45.1</td>
<td>45.7</td>
<td>47.6</td>
<td>48.0</td>
</tr>
<tr>
<td>1</td>
<td>40.0</td>
<td>40.0</td>
<td>41.5</td>
<td>42.3</td>
<td>43.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>40.6</td>
<td>43.8</td>
<td>44.4</td>
<td>45.7</td>
<td>45.7</td>
<td>46.3</td>
<td>46.9</td>
<td>47.6</td>
</tr>
<tr>
<td>2</td>
<td>38.1</td>
<td>39.4</td>
<td>40.6</td>
<td>43.2</td>
<td>43.2</td>
<td>46.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>40.0</td>
<td>43.0</td>
<td>44.3</td>
<td>46.5</td>
<td>47.2</td>
<td>49.0</td>
<td>52.5</td>
<td>53.0</td>
</tr>
<tr>
<td>1</td>
<td>40.0</td>
<td>42.0</td>
<td>44.0</td>
<td>45.5</td>
<td>46.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>49.5</td>
<td>50.1</td>
<td>51.4</td>
<td>52.4</td>
<td></td>
</tr>
</tbody>
</table>

You may name a data file any legal DOS name (i.e., 1-8 characters long - see your computer/DOS manual for complete details) but the extension must be .ASC. For example anyname.asc. If you already have a data file on disk, use the DOS RENAME utility to append the extension. Remember, you may only have data in ASCII format in your file. When preparing data for some programs, you may have to use a word processor to strip dates, ID numbers/names, and non-ASCII codes from the data files before you can read them into these programs.

To convert an ASCII file to a GAUSS file, type `copy data_file_name c:\directory_name <return>`, where `data_file_name` is the name of your data file and `directory_name` is the directory you create when you install the program. For most users, this will be \LDA. Then type `cd \ directory_name <return>` to move to that directory. Finally, type `atog386 data_file_name <return>` to invoke the ATOG.EXE utility. This will create the two files, `data_file_name.dat` and `data_file_name.dht` needed for the GAUSS format.

Note: It is not necessary to convert ASCII files to GAUSS files. The programs will accept ASCII files without modification. We have included the ATOG.EXE utility because GAUSS files run somewhat faster than ASCII files.

VI. RUNNING THE PROGRAMS

To run any of the programs, simply move to the directory you installed them in (again, for most users this will be \LDA) and type `GSRUNI program_name`. The use of the .CGG extension is optional (i.e., GSRUNI KTRK works just as well as GSRUNI KTRK.CGG).
VII. EXAMPLES

Three sample data files have been included with the distribution disks, RAMUS.DAT, JRB12.ASC, and ELS_GRP1.ASC.

1. RAMUS.DAT is a GAUSS format data file containing 5 measurements (time points) of the Mandibular ramus (ascending portion of lower jaw) height of 12 rhesus monkeys. (12 subjects X 5 time points). A second file, ramus.dht, must also be present to read this file into GAUSS. Both files are unreadable.

2. JRB12.ASC is an ASCII data file containing 4 measurements (time points) of the height of 12 Central American school children. (12 subjects X 4 time points)

3. ELS_GRP1.ASC is an ASCII data file containing 13 measurements (time points) of the head circumference of 11 achondroplastic children taken during the first year of life. Most of the subjects have missing data, some with as few as 4 measurements.

To illustrate the use of the package, we will step through two examples.

EXAMPLE 1

The first example uses the Foulkes - Davis tracking program. From the directory containing the programs, type GSRUNI FDTRK <return>.

1. After a brief moment, the initial screen will appear for 3 seconds followed by the main menu screen. The main menu consists of a series of prompts (in this case 5).

2. NUMBER OF DATA SETS: You may, if you wish, process more than one data file per run. In this case, however, chose 1 <return>.

3. NAME OF DATA SET: You are prompted to indicate if the data files are on the current directory. (If they are not, you are asked to provide the drive/directory containing the data files.) Type Y <return>. The window at the bottom of the screen now displays the names of all files ending in the extension .asc or .dht, 15 at a time. (This is why you must append the extension .asc to your data files. The ATOG.EXE utility will append the extension .dat to GAUSS files when invoked.) If you have more than 15 data files, you may use the PgUp and PgDn keys to search for more files. Use the arrow keys to highlight the correct data file. In this case, only RAMUS.DAT, JRB12.ASC, and ELS_GRP1.ASC appear, as they are the only data files on the current directory. Highlight JRB12.ASC and press <return>.
4. ASCII OR GAUSS: Enter A <return>. If this were RAMUS.DAT, you would enter G for GAUSS. A window now opens at the bottom of the screen and asks you for the number of subjects in your data file. Enter 12 <return>.

5. ARE DATA COLLECTED AT EQUAL TIME POINTS: Were the data (time points) collected at equal intervals? If not, you will be prompted for the exact times. In this case, type Y <return>.

6. NUMBER OF TIME POINTS: Type 4 <return>. A window will open at the bottom of the screen to inquire if you want the program to space the time points. Press Y <return>.

The program will now commence and you may answer the prompts as they appear.

EXAMPLE 2

From the DOS prompt in the directory containing the programs (usually C:\LDA) type GSRUNI ZRTE <return>.

1. After the initial startup screen, a message requesting information on your data file appears. After reading it, press <return>.

2. After the initial screen you will be asked if your group indicator is in the first column. It can be any in column, but here enter Y <return>.

3. You will next be asked for the number of groups in your data. You must have at least 2 but no more than 5. In this case enter 2 <return>.  
   Note: This data set is ELS_GRP1.ASC used in example data set #2 (see above).

4. NAME OF DATA SET: The standard input menu screen now appears. Type Y <return> to open the window with the data files. Highlight ELS_GRP1.ASC and press <return>.

5. ASCII or GAUSS: Enter A <return>. A window opens and asks the number of subjects. In this case enter 11 <return>.

6. ARE DATA COLLECTED AT EQUAL TIME POINTS: If you answer no, you will be prompted for the exact times. In this case, enter Y <return>.

7. NUMBER OF COLUMNS IN YOUR DATA SET: Enter 14 <return>. (There are 13 columns of data and 1 column of group indicators.) A window opens and asks if you want to start at time point 1. If not, you will be asked for a starting time point and an inter-time point increment. Enter Y <return>.
8. The program commences. You will have various options to choose from during the program.

VIII. OUTPUT

Each program writes its results to an output file on disk in ASCII format which is named program_name.out (i.e., KTRK.OUT, RAO.OUT, etc.). These files may be printed using most word processors or the DOS [PRN] utility. They may also be annotated using a word processor.

IX. ACKNOWLEDGMENT and CITATION

The development of this software is principally supported by NIH-NIDR Grant DE08730. Please consult and cite the appropriate publications listed above in Section IV when publishing results that were generated with these methods and programs.

X. USER ASSISTANCE

If you need assistance, find any errors, or have comments, please notify:

Dr. Emet D. Schneiderman        TEL 214-828-8432 or 828-8299
Stephen M. Willis              TEL 214-828-8905
                                  FAX 214-828-8346
                                  INTERNET: 71742.2362@compuserve.com

Dept. of Oral and Maxillofacial Surgery
Baylor College of Dentistry
3302 Gaston Ave.
Dallas, TX 75246