CHAPTER 3. DATA FILES AND FORMATS

This chapter shows how to open or close a data file and how to save model setups. It also describes the four alternative data file formats that can be used.

Opening a Data File

The standard program accepts input data:

1) saved as an SPSS (.sav) system file,

2) saved as an ASCII text (rectangular) file,

3) saved as a special array file format, or

4) referenced in a previously saved Latent GOLD (.lgf) file. For detailed descriptions about the formats for The DBMS/COPY option allows the program to import data from more than 80 additional file formats, such as Excel, SAS and HTML files, using the File Import menu option. For a full list of additional formats, see http://www.statisti-calinnovations.com/products/latentgold_80formats.html

See below for details about the optional File Import menu option.

To open a data file using one of the 4 standard formats:

▷ From the menus choose: File → Open
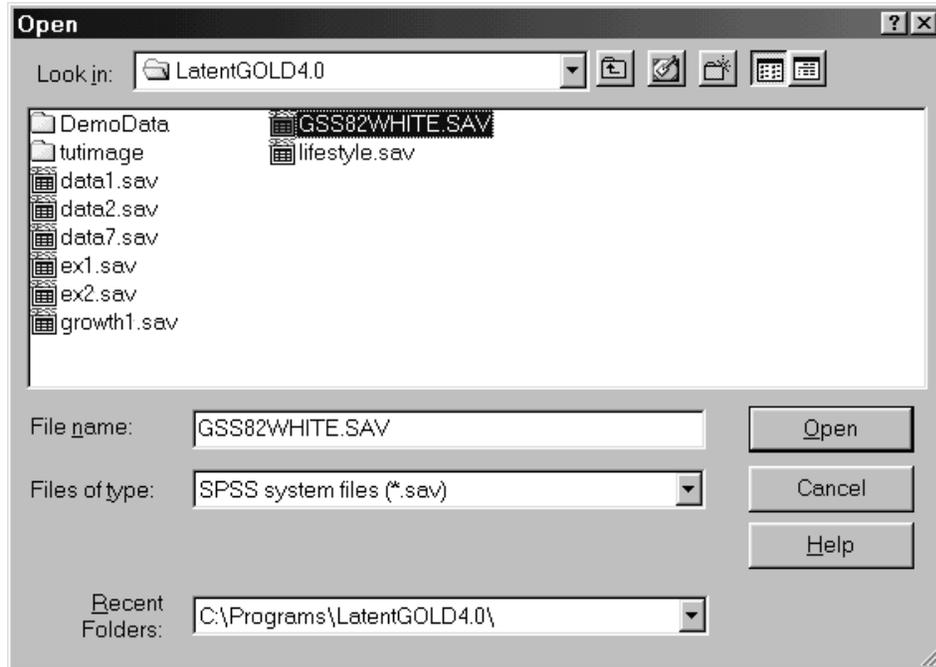
or select the Open icon 📂 from the toolbar.

Figure 3-1: File Open Dialog Box

▷ In the Open dialog box, select the type and name of the file you want to open.

FILE OPEN OPTIONS

Look in: Use this option to change drives and directories.

File Name. You can type in a filename, a path and filename or a wildcard search. By default, Latent GOLD looks for all files in the current directory with the same extension as the last file that was opened and displays them in the dialog box. If you type wildcards, all files that match the wildcard search in the directory will appear in the dialog box.

Files of Type. You may specify the data file format to be one of the following:

- Latent GOLD files (*.lgf). Lists previously saved Latent GOLD analysis files.
- Arrays (*.ar*). Lists ASCII array files. These files contain frequency counts for each cell in a multi-way table.
- Text files (*.dat, *.csv, *.txt). Lists text files with .dat, .csv or .txt extensions. These files are

often referred to as rectangular or ASCII text files.

- SPSS system files (*.sav). Lists SPSS system files with .sav extensions.

- All Files (*.*). Lists all files in the current directory.

SPSS, TEXT AND ARRAY FILES

For SPSS system files, Text files and Arrays, once you have selected a file, the data file name appears in the Outline pane with a default model name ('Model1') below it.

▷ Either double or right click 'Model1' to open the Model Selection menu.

File Import Option

IMPORTING SAS, EXCEL, AND OTHER DATA FILE FORMATS

If you have licensed the DBMS/COPY add-on, an additional option called "Import" appears in the File menu. You can use this option to open data files saved in any of these 80 formats. Selection of File Import opens a dialog box as shown in Figure 3-2.
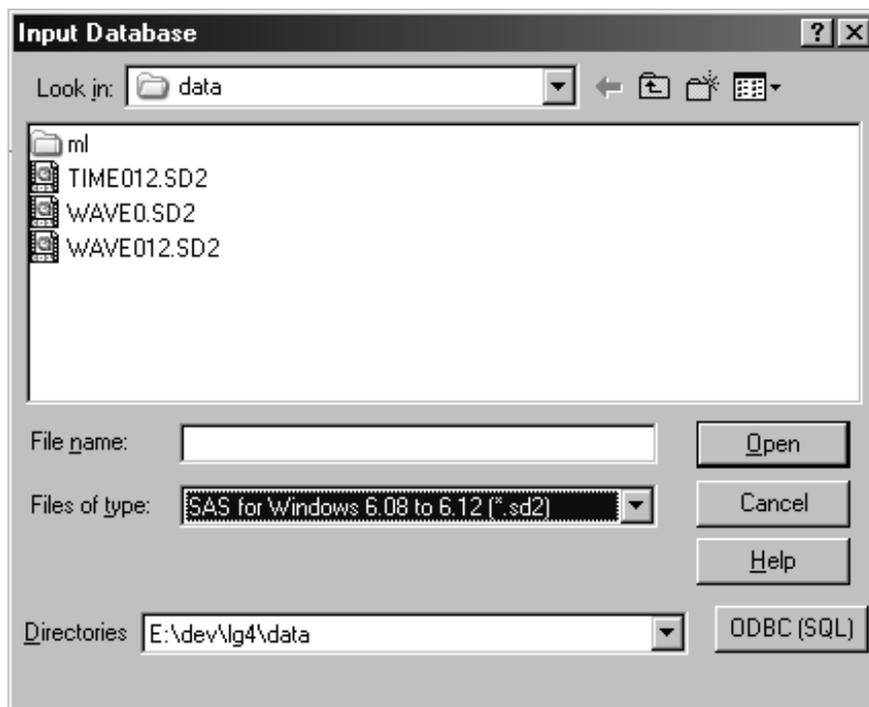

Figure 3-2: File Import Dialog Box

Files of Type: shows available file types. The database file type displayed by default is the type last used.

Directories: shows the last 10 directories that you have used. You can click on one and it becomes the current directory.

ODBC (SQL): this button invokes the Microsoft Open DataBase Connectivity/SQL (Structured Query Language) subsystem. This is how you get to Microsoft Access, Oracle and others. From here you will select an ODBC driver and use the Query Builder to create an SQL query.

Note: The option to output data to a file using the ClassPred Tab is not available when the data file was input using the File Import option.

Latent GOLD Files

If you have previously saved the settings for one or more models in a Latent GOLD definition (.lgf) file using the File Save Definition menu option (see 'Saving Model Settings & Output' below), upon opening this file, a list of model names associated with the saved model settings appears in the outline pane.

To estimate one of these models:

▷   Double click a model name to open the Analysis dialog box for that saved model setup

▷   Make any desired changes to the settings (using the saved setup as the starting point).

▷   Click Estimate.

or

▷   Click on the toolbar to estimate the model.

Alternatively, to estimate all models that have not yet been estimated.

▷   click on the data file name in the Outline Pane

select 'Estimate All' from the Model Menu

RE-OPENING DATA FILES QUICKLY

Recent Folders: the software retains a list of recently used folders that you have used: these can be quickly selected from this list.

Up to eight data files that have been opened most recently in Latent GOLD will be listed at the bottom of the File Menu (just above the Exit option). To re-open these data files quickly, simply select the data file name from the File Menu.

Data File Formats

Latent GOLD will load data from input files in any of the following data formats:

- **SPSS system file (*.sav)**
- **ASCII text file (*.dat,* .csv or *.txt)**
- **ASCII array (array format) file (*.arr)**
- **Latent GOLD save file (*.lgf)**

Use of an SPSS .sav file (or the array format if all your variables are categorical) allows you to assign category labels to particular categories or values.

> *Note:* The option to output data to a file using the ClassPred Tab is not available when the input data is an array format.

The Latent GOLD save file can be generated only after estimating a model.

SPSS FILES

For information on creating an SPSS system file, see your SPSS manuals (programs such as DBMS Copy may be used to convert SAS and other data formats to the SPSS format).

- Latent GOLD recognizes SPSS variable names and category (value) labels (less than 40 characters), but ignores SPSS variable labels and date variables.

- A variable used in SPSS to weight cases is automatically placed in the Frequency box.

- If an SPSS system file contains multiple user missing designations for a numeric variable, Latent GOLD will combine them into one aggregate 'missing' category. Latent GOLD does not recognize user missing specifications for non-numeric variables.

- String variables with lower and upper case letters, such as 'f' and 'F', are distinguished as separate categories.

TEXT FILES

A text (rectangular) file can be created in any spreadsheet or word processor using the 'Save as..'' command to save the data as a text file with either a .dat, .csv or .txt extension.

- The first line of the file needs to contain the variable names separated by blank spaces or a tab. The remainder of the file contains the data.

- For each record, the data for each variable should be separated by either a space or a tab (no other delimiters, such as ",", are allowed).

- The data for any variable may be numeric (quantitative values only) or may be a string variable, containing some or all alphabetic characters (such as 'Female', 'Male'). For numeric data, do not use commas (such as 3,634). String variables containing lower and upper case letters, such as 'f' and 'F', are distinguished as separate categories.

- Missing data for a variable may be specified using '.'.

- Each data record should contain exactly the same number of data elements, one for each variable name.

Below is a partial listing of a text formatted data file. This file contains one record per case. Note that records 6 and 7 are identical (0 30 0 F).

IMPROVE	AGE	TREAT	GENDER
0	23	1	F
0	23	0	F
1	27	1	M
0	29	1	M
0	30	1	M
0	30	0	F
0	30	0	F
1	31	0	F
2	32	1	M
0	32	1	F
0	32	0	F
2	33	0	F
1	37	1	F
0	37	0	M

For data sets consisting of many observations that contain identical values on all variables, the inclusion of an optional integer value frequency count variable will reduce the number of physical records in the file. Below is a partial listing of another text formatted data file. For each data record, FREQ contains the count of observations having the specified values on each of the variables. For example, the value of FREQ for the first record, defined by the values specified as "1, 1, 1, 1, 1" is 3,634.

BACK	NECK	JOINT	SWELL	STIFF	FREQ
1	1	1	1	1	3634
1	1	1	1	2	73
1	1	1	2	1	87
1	1	1	2	2	10
1	1	2	1	1	440
1	1	2	1	2	89
1	1	2	2	1	106
1	1	2	2	2	75
1	2	1	1	1	295
2	2	2	1	2	162
2	2	2	2	1	44
2	2	2	2	2	176

ARRAYS

In the case that all variables are categorical, a special array file format may be used. The array file can be created in any word processor or editor program. It contains frequency counts for each cell in the multi-way table. The general ASCII format is:

```
ARRAY
<d1> <d2> <d3> ...
LABELS
<var>/<cat1>/<cat2>/<cat3>/...
...
SCORES
<var>/<scr1>/<scr2>/<scr3>/...
...
DATA
f1...11 f1...12 f1...13 ...
f1...21 f1...22 f1...23 ...
... fn...n
```

The array below re-expresses the text formatted dataset shown above in array format with category labels (1 = 'No', 2 = 'Yes').

- **ARRAY**
- **2 2 2 2 2**
- **LABELS**
- **BACK/No/Yes**
- **NECK/No/Yes**
- **JOINT/No/Yes**
- **SWELL/No/Yes**
- **STIFF/No/Yes**
- **DATA**
- **3634 73 87 10 440 89 106 75**
- **295 25 15 5 137 42 35 39**
- **489 37 23 7 255 116 71 65**
- **306 48 16 11 229 162 44 176**

The key words "**ARRAY**", "**LABELS**", "**SCORES**" and "**DATA**" must begin in column 1 and be written in upper case. Following the "**LABELS**" key word, exactly one label record must appear for each variable. The "**SCORES**" section of the file, which may be used to assign scores to the categories, is an optional section. If the keyword "**SCORES**" appears in the file, at least one score record must follow.

The frequency counts for each cell are entered on lines following the **DATA** keyword in free format, and must be entered in the order such that the first dimension varies least rapidly..., and the last dimension most rapidly. (The use of scientific notation for frequency counts is not supported.)

LATENT GOLD SAVE FILE (*.LGF)

This File, Save menu option can be used to save the variable selections and other option settings used for one or more previously defined models in estimating the model(s) in the format of a Latent GOLD save file (.lgf) for a particular model or series of models (see Saving Model Settings). This file can then be opened at a later time using the File, Open command. The advantage of an .lgf file is the ability to retrieve analysis settings for a particular model or series of models at a later time without having to re-specify these settings. You can also use a retrieved setting as a starting point for specifying a similar model on the same data. In addition, if a saved definition file

Saving Model Settings & Output

Latent GOLD allows you to save your model settings for

- **an individual model**
- **a series of models estimated from one data file.**
- **model output**

The File / Save menu option incorporates two distinct save features -- **File / Save Definition**, and **File / Save Results**.

> **File / Save Definition** creates an .lgf text file containing the model settings so that the model can be re-estimated at a later time. To save the settings for a model prior to that model being estimated, after selecting the desired model options click the Close button. Then, select Save Definition for the File menu. When creating an .lgf file for a model that has been estimated, the .lgf file contains the model settings together with the best start seed from the previously estimated model. When the best seed is obtained from Latent GOLD 4.0, this insures that re-estimation of the model using Latent GOLD 4.0 with this .lgf file will yield the same results as before. For further details on this see *Start Values Procedure*.

> **File / Save Results** creates a ASCII text or HTML file containing selected output from a previously estimated model. Either output from all estimated models (Save All Views), all output from a single model (Save Model Views), or any one selected output section (Save View) can be saved.

SAVE DEFINITION

This creates an .lgf text file containing the model settings for one or more models associated with a particular data file.

▷   Highlight a model name in the Outline pane,

or to save the settings for all models associated with a data file

▷   Highlight the data file name in the Outline pane

▷   From the menus choose: File → Save Definition

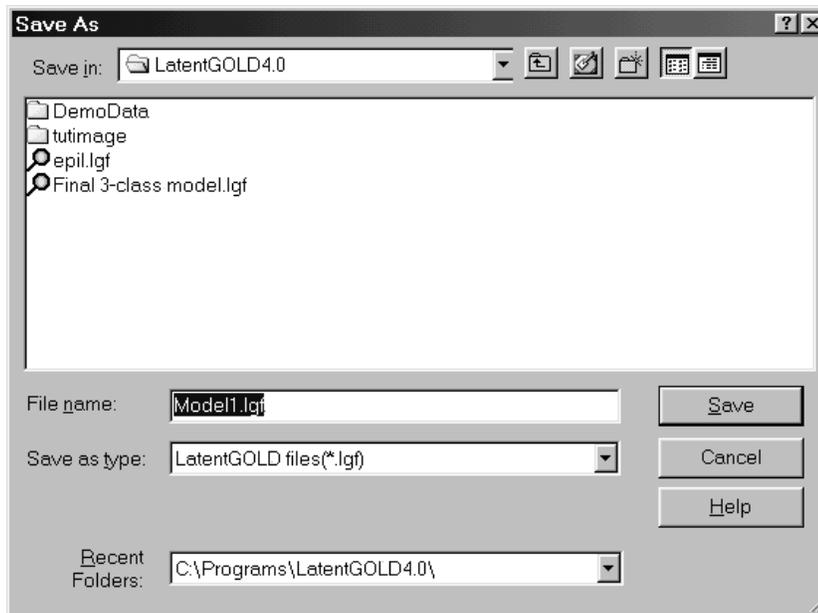(or select the Save icon 🖫 from the toolbar). The Save dialog
box will open.

Figure 3-3: Save Definition Dialog Box

▷   In the Save dialog box, specify the directory you wish to save to
and type in a filename (Latent GOLD will prompt you with a
default file name consisting of the model name and .lgf
extension.)

▷   Click Save. Save Options. File Name.

Enter the name of the document (Latent GOLD will prompt you with
a default name and a .lgf extension).

Save as Type.

Select Latent Gold files (*.lgf).

The specifications for a particular model that are saved are:

All settings from the Analysis dialog box, such as

- Variable settings (variables selected as indicators, covariates, frequency, dependent or predictors)
  and scale types set under the Variables Tab.
- Any restrictions or other settings chosen in any of the other tabs.

33

Save Results

> ▷ **This feature allows you to save the output file in either HTML or ASCII format for later viewings.**

▷   Highlight the model name in the Outline pane (If you select File, Save without highlighting a particular model name, Latent GOLD will save the model that corresponds to the output appearing in the Contents pane.)

▷   Highlight the data file name in the Outline pane

▷   From the menus choose: File → Save Results

      (or select the Save icon 💾 from the toolbar).
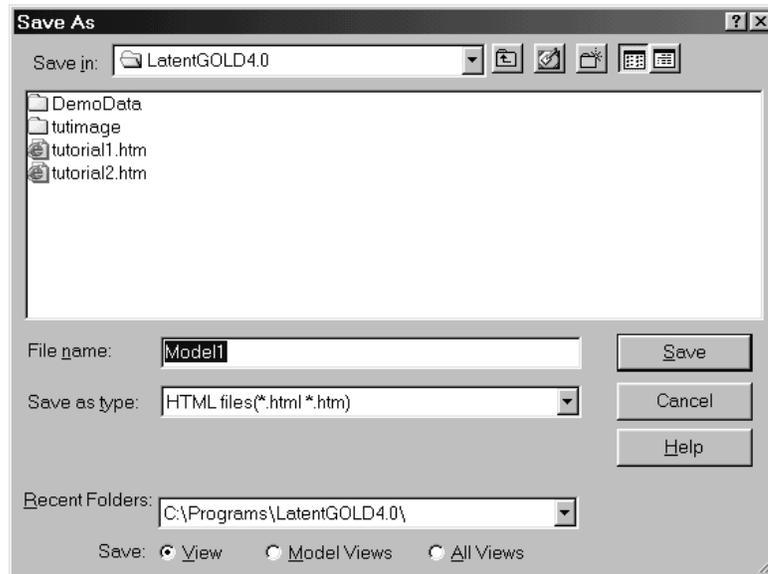   The Save dialog box will open.

Figure 3-4: Save Results Dialog Box

At the bottom of the dialog box you have several options:

    Save: View: Click to save only the selected output section from a model

    Save: Model Views: Click to save all output from a single model

    Save: All Views: Click to save all output from all estimated models

Closing a Data File

To close a previously opened Latent GOLD data file,

▷   Click on the data file name listed in the Outline pane.

▷   From the menus choose: File → Close

Latent GOLD will close all windows and dialog boxes (including output files) associated with the data file.