

ESRC Researcher Development Initiative

Longitudinal Data Analysis for Social Science Researchers

Training workshop

www.longitudinal.stir.ac.uk

Guide to Lab exercises 0,1,2, and 3: Longitudinal Data Analysis in SPSS and STATA

-

Stirling workshop 2-6th September 2006

The Longitudinal Data Analysis for Social Science Researchers (LDA) training materials include a collection of extended example sessions in which SPSS and STATA commands are used to illustrate techniques of longitudinal data management and data analysis. This document acts as a guide to the content of those sessions. The sessions hinge upon example command files in the SPSS and STATA languages. Those files are available for download from the LDA Web site (and/or on discs during taught workshops). However, those command files also draw upon further data and programme files, and it is also necessary that users ensure that those relevant files are stored somewhere appropriately on the machine being used.

1. Using the SPSS and STATA command files

The command files in the SPSS and Stata languages are written as ‘interactive syntax’ files (‘syntax’ files to SPSS, and ‘do’ files to Stata). These involve writing out textual commands which call the relevant processes in each package. The command files are opened up in specialist windows within each package (the ‘syntax editor’ in SPSS, and the ‘Do-file editor’ in Stata). Users then process the component commands by moving through the syntax files, highlighting relevant lines or groups of lines, and asking the package to process the highlighted text through their respective ‘run’ commands. (Some more notes on getting started with SPSS and Stata are available from the ‘software’ link of the LDA Web site).

The relevant SPSS and Stata command files associated with the LDA training sessions are:

SPSS syntax files*	Stata Do-files*
<i>Session 0: Re-introducing survey data</i> lab0.sps	lab0.do
<i>Session 1: Five approaches to Longitudinal Data Analysis</i> lab1.sps	lab1.do
<i>Session 2: Panel survey data management and data analysis</i> lab2.sps	lab2.do Bell_PanelModels_GreeneC14.do
<i>Session 3: Event history data management and data analysis</i> lab3.sps	lab3.do

**The files are written for SPSS version 14, and Stata version 9. There is almost complete compatibility between the commands used on these versions and earlier versions of the software, but there may be one or two small segments which do not run successfully on an older version of either package. However, users of Stata 8 or earlier should note that the length of the do files for lab0 and lab2 exceeds that which used to be permitted as the maximum length of the do file editor in those versions. To process these commands it is therefore necessary to split the do-files into two or more components.*

To undertake the LDA training labs, users are expected to open up the relevant syntax files, and work their way through their contents. Relevant commands typically have some explanatory text associated with them, whilst, in addition, processing the commands will generate output in the SPSS and Stata packages, which users should inspect in order to understand what the commands have just done.

All of the command files are relatively long text files. At the top of all files is a listing of contents, which correspond to the different segments of the syntax files below (searching for the section titles is one quick way to navigate these command files). The contents of respectively the SPSS and Stata files cover in most instances the same examples using the same data files, although there are some limitations to this compatibility, most noticeably where there is no equivalent SPSS functionality for some of the specialist panel and event history models available in Stata.

As described in section 2 below, all command files used in these sessions make, in addition, use of further associated files (such as data files) for their processing: it will not be possible to run the training session example command files successfully without also ensuring that the relevant associated files are also stored at an appropriate location on your machine.

SPSS or Stata? *Most of the LDA workshop sessions offer example applications in both SPSS and STATA. The range of commands illustrated for each package are similar, but not completely equivalent. Users may prefer to implement the exercises in either one or the other, or both.*

*Working with SPSS alone may suit users with no prior experience of other packages, and may have the benefit of greater compatibility with other researchers. However, SPSS does not have the full functionality for the more advanced longitudinal data analyses described in the later exercises. Working with Stata, on the other hand, has the benefit of much greater functionality – though Stata is still not as widely used in the UK as it might be. Lastly, **working jointly with SPSS and Stata** offers a considerable amount of training, but you should be aware that it can be cognitively (and logistically) difficult to switch frequently back and forth between alternative software package operations.*

Selected SPSS syntax basics:

- ✓ New commands begin on a new line.
- ✓ Command lines can spread over several lines, but commands must always end with a full stop. When a command line is spread over several lines, it is conventional, though not essential, to indent the 2nd, 3rd etc lines by one or more spaces.
- ✓ To process a syntax command, highlight part of the relevant line (or lines) within SPSS 'syntax editor' window, then either click the 'run' arrow (black triangle), or press 'ctrl'- r.
- ✓ Several adjacent commands can be processed at once, by highlighting the relevant segment of the syntax command file and pressing 'ctrl' – r .
- ✓ Ensure that your SPSS output ('viewer') has its option set to 'show commands in the log' – this means the syntax command will be echoed to the output screen, making it much easier to follow your outputs (see description on LDA Web site)
- ✓ Any line beginning with a '*' symbol is ignored by the SPSS processor and is regarded as a 'comment' (it is good practice to write explanatory comments into syntax files purely for your own benefit).
- ✓ The sequence of SPSS commands is important – errors often occur when users try to run things out of sequence. It usually helps to start your analysis sessions from 'first principles', by beginning with an appropriate 'get file' syntax command
- ✓ There are several ways to learn new syntax commands. The best, when available, is to copy and paste from somebody else's previous examples – such as these template example files. Another option which some people favour, is to set up new processes through the GUI commands (windows), then click 'paste' rather than 'ok' to send the syntax command equivalent to your currently active SPSS window
- ✓ Quick help on the text of a syntax command may be obtained by clicking on the 'syntax help' icon of SPSS's syntax editor window, whilst your cursor is positioned on the relevant command that you want to query
- ✓ Links to further support materials on working with SPSS are available from the LDA website at http://www.longitudinal.stir.ac.uk/SPSS_support.html

Selected Stata Do-file basics:

- ✓ New commands begin on a new line.
- ✓ Commands *cannot* [by default] be spread over several lines (because the default command 'delimiter' is the return carriage). If it is required to spread a command over several lines, a quick 'fix' is to use the continuation symbols /// at the end of the unfinished first line, and at the start of the next line (see example file illustrations).
- ✓ To process an individual do file command, highlight part of the relevant line (or lines) within the Stata 'do file editor' window, then either click the 'do current file' icon, or press 'ctrl'-d .
- ✓ Several adjacent commands can be processed at once, by highlighting the relevant segment of the do file and pressing 'ctrl'-d .
- ✓ Any line beginning with a '*' symbol is ignored by the Stata processor and is regarded as a 'comment' (it is good practice to write explanatory comments into do file purely for your own benefit).
- ✓ The sequence of Stata commands is important – errors often occur when users try to run things out of sequence. It usually helps to start your analysis sessions from 'first principles', by beginning with an appropriate 'use' do file command
- ✓ There are several ways to learn new do file commands. The best, when available, is to copy and paste from somebody else's previous examples – such as these template example files. The Stata manuals, available online (www.stata.com/capabilities/search.html), or in a collection of text volumes, are also accessible and give invaluable advice on do-file commands.
- ✓ Links to further support materials on working with Stata are available from the LDA web site at http://www.longitudinal.stir.ac.uk/Stata_support.html

2. Other files necessary to carry out LDA training sessions

After downloading the relevant example command files for these sessions, it will also be necessary to ensure that a number of appropriate data files (and a small number of folder structures) are stored in an appropriate location on your machine. There are a considerable number of these files which are used during the training sessions (in most cases, it should be possible to transfer the files in large groups, rather than one at a time). The details of these files are also entered at the top of the relevant command syntax files.

Both the SPSS and Stata example command files call upon these associated data files by looking in specified 'paths' (directory locations) on your machine. In both cases, the relevant path declarations are made at the top of the command files (by using short macro definitions). Therefore, in order to undertake these research training exercises, users must download the relevant data files to consistent locations on their machines, then define the paths for those data files accordingly at the start of the relevant SPSS and Stata command files.

The table below shows, by session, which files are necessary for the successful conduct of the lab exercises, and the provenance from which the associated files may be obtained.

LDA Training sessions: Associated data files to lab exercises		
SPSS File details	STATA File details	Source
<u>Lab 0: Revising survey data analysis</u>		
Ghs95.sav SSA sample files, 1999-2002	Ghs95.dta SSA sample files, 1999-2002	LDA Web site UKDA
BHPS wave-on-wave component files	Div3s2004.dat BHPS wave-on-wave component files	LDA Web site UKDA Sn 5151
Workhours.xls	Wemp.dat / wemp_s2.dat	LDA Web site
Cleandat.sav		LDA Web site
Dirtydat.sav		LDA Web site
Itskills_entry.sps		LDA Web site
Gb91soc2000.por	Gb91soc2000.dat	LDA Web site
SPSS macro syntax files	Stata macro do files	LDA Web site (see SPSS/Stata – macros)
<u>Lab 1: Introducing longitudinal data formats and analysis techniques</u>		
BHPS wave-on-wave component files	BHPS wave-on-wave component files	UKDA Sn 5151
BHPS life history files	BHPS life history files	UKDA Sn 3954
BHPS derived income files	BHPS derived income files	UKDA Sn 3909

bh1to14_long.sav (BHPS file)	bh1to14_long.dta	Created in lab 2
bh1to14_wide.sav (BHPS file)	bh1to14_wide.dta	Created in lab 2
LFS sample files	LFS sample files	UKDA, Sn 2875, 3647, 4448
SSA sample files, 1999-2002	SSA sample files, 1999-2002	UKDA, Sn 4967
2364a.por (NCDS example file)	2364.dta (NCDS example file)	UKDA, Sn 2364
Gb91soc2000.por	Gb91soc2000.dat	LDA Web site
SPSS macro syntax files		LDA Web site (see SPSS – macros)
<u>Lab 2: Panel survey data management and data analysis</u>		
BHPS wave-on-wave component files	BHPS wave-on-wave component files	UKDA Sn 5151
BHPS life history files	BHPS life history files	UKDA Sn 3954
BHPS derived income files	BHPS derived income files	UKDA Sn 3909
bh1to14_long.sav (BHPS file)	bh1to14_long.dta	Created in lab 2
bh1to14_wide.sav (BHPS file)	bh1to14_wide.dta	Created in lab 2
SPSS macro syntax files		LDA Web site (see SPSS – macros)
	Panel.csv	LDA Web site
<u>Lab 3: Event history survey data management and data analysis</u>		
BHPS wave-on-wave component files	BHPS wave-on-wave component files	UKDA Sn 5151
BHPS life history files	BHPS life history files	UKDA Sn 3954
BHPS derived income files	BHPS derived income files	UKDA Sn 3909
cox_childcare.sav		LDA Web site

3. Data File Access: Stirling Workshop

The data files supplied for the lab exercises (e.g. on CD) are expressly for use at this workshop. **They MUST NOT be taken away.** The datasets used are all available from the UK Data Archive. Because you have registered with the UK Archive prior to this workshop, you will be able to download them from the UK Archive.

EOF