An Introduction to Stata for Health Researchers:

Information to users of Stata 8 to 13

An Introduction to Stata for Health Researchers, first edition[1] was written for Stata release 9; the second edition[2] was written for Stata 10; the third edition[3] for Stata 11, and the fourth edition[4] for Stata 13. We refer to the different editions of the book as ISHR1, ISHR2, etc.

There may be some discrepancies between your edition of the ISHR book and the release of your Stata program, and in the following we point to the most important changes, organized by the structure of the book. We are only pointing, and you will need Stata's help facilities to get more information. Keep your Stata program updated to the latest version.

To see all changes, for example from release 12 to 13, type (it only works in Stata 13):
    . help whatsnew12to13

Stata introduced new file formats in releases 8, 10, 12, and 13. Your Stata program reads datasets from previous releases without trouble, but the opposite is not the case.[5] [6] If you want to create a dataset that can be read by a previous Stata version, you may need to use the **saveold** command; see the following table.

Your version of Stata:	**save** generates data with the format of:	**saveold** generates data with the format of:
Stata 13	Stata 13	Stata 12
Stata 12	Stata 12	Stata 8/9
Stata 10 or 11	Stata 10/11	Stata 8/9
Stata 8 or 9	Stata 8/9	Stata 7

If you from Stata 13 want to create a dataset that can be read by Stata prior to version 12, it cannot be done directly. One possibility is to let Stat/Transfer (release 12) import a version 13 dataset and export it as a version 9 dataset.

Some date functions and formats changed from Stata 9 to Stata 10. With Stata 10+, you can still use the Stata 9 syntax using version control:
    . version 9: generate bdate = date(sbdate, "dmy")

Svend Juul, Morten Frydenberg, March 2014

---

[1]   Juul S. *An Introduction to Stata for Health Researchers*. College Station, TX: Stata Press, 2006

[2]   Juul S. *An Introduction to Stata for Health Researchers*. 2nd ed. College Station, TX: Stata Press, 2008.

[3]   Juul S, Frydenberg M. *An Introduction to Stata for Health Researchers*. 3rd ed. College Station, TX: Stata Press, 2010.

[4]   Juul S, Frydenberg M. *An Introduction to Stata for Health Researchers*. 4th ed. College Station, TX: Stata Press, 2014.

[5]   Stata 11 users: An update to Stata 11, released in September 2011, allows Stata 11 to read Stata 12 datasets. To use it, you need to update your Stata 11.

[6]   Stata 9 users: The unofficial **use10** command reads Stata10/11 datasets. Get if by **ssc install use10**.

Changes from Stata 12 to Stata 13

The information is organized by the chapters in ISHR3. In the Stata 12 column we show changes from the description in ISHR3. (ISHR3 describes Stata 11; on changes from Stata 11 to Stata 12, see next page).

Section	Stata 13	**Stata 12** (as described in ISHR3)
1.1	The *installation directory***ado\updates** folder no longer exists. Updates are made by replacement in the **\ado\base** folder.	*Installation directory***ado\base** contains the originally shipped ado-files; updates are put in **\ado\updates** which Stata searches before **\ado\base**.
1.4	In the Do-file Editor, matching parentheses are highlighted automatically.	In the Do-file Editor, use *Ctrl+5* to see a matching parenthesis.
2.1	The manuals only exist in PDF versions.	The manuals exist in paper and PDF versions.
2.2.	**search** searches both in official Stata and the Internet. **findit** remains available, but undocumented.	**search** searches in official Stata; **findit** also searches the Internet.
2.2	If **help** does not find a result, **search** is invoked automatically.	**search** is a separate command to be invoked by the user.
2.2	**help videos** lets you see a number of instructive video tutorials.	
5.7	String variables can be up to 2,000,000,000 characters. If larger than 2,054 characters, the storage type is **strL**.	String variables can be up to 244 characters.
6.3	Use **import delimited** to read tab- or comma-separated data. **insheet** remains available, but undocumented.	Use **insheet** to read tab- or comma-separated data.
6.4	Use **export delimited** to write tab- or comma-separated data. **outsheet** remains available, but undocumented.	Use **outsheet** to write tab- or comma-separated data.
13.3	The value labels, if any, of factor variables are used to label estimation output. The **nofvlabel** option turns this off.	The values of factor variables are used to label estimation output.
16.2	New **power** command for sample size and power analysis. **sampsi** remains available, but undocumented.	Use **sampsi** for sample size and power analysis.

Changes from Stata 11 to Stata 12

The information is organized by the chapters in ISHR3. In the Stata 12 column we show changes from the description in ISHR3.

Section	Stata 12	Stata 11 (as described in ISHR3)
1.4	Different organization of windows, including a new Properties window. The default layout fits a wide screen; if you prefer a narrower layout, select: **Edit > Preferences > Load Preference Set > Combined Layout**	
1.4	Paste a variable name from the Variables window to the Command window with a double-click.	Paste a variable name from the Variables window to the Command window with a single click.
1.4	Improved data editor. In the Variables Window you can do quite a few things, including hiding and reordering variables.	
5.8	Stata 12 automatically calculates the memory needs, and the **set memory** command is no longer needed.	**set memory** allocates Stata memory.
6.4	Stata 12 now imports and exports Excel and SAS data directly. See: **help import excel** **help import sas**	Import and export SAS datasets using the **fdause** and **fdasave** commands.
9.2	New, flexible **rename group** command. See: **help rename group**	
13.3	The **contrast** command may be used as a handy replacement for **testparm**. See: **help contrast**	
13.3	The output from regression commands (including **regress**, **logistic**, **stcox**, and **poisson**) with factor variables will always display the base level if you, just once, give the command: **set showbaselevels on, permanently** and we recommend doing that to enhance transparency.	By default, base levels are omitted from the output. The **set showbaselevels** facility was introduced in Stata 11.1, June 2010.
13.5 14.7	Baseline odds, risk, and rate estimates are displayed by **eform** commands, like: **logistic** and **logit, or** **binreg, rr** **poisson, irr**	Baseline odds, risk, and rate estimates are not displayed by **logistic** or any other **eform** command.
17.12	You can export a graph to a PDF file. See: **help graph export**	

Changes from Stata 10 to Stata 11

The information is organized by the chapters in ISHR2. In the Stata 11 column we show changes from the description in ISHR2.

Section	Stata 11	Stata 10 (as described in ISHR2)
1.4	Data Editor/Browser improved. You need not close it while doing other things in Stata	Data Editor/Browser must be closed while doing other things in Stata
1.4	Do-file Editor improved. It now has syntax highlighting	
1.5	New Variables Manager	
2.2	PDF version of manuals included	Only printed version of manuals
5.3	New **misstable** command describes missing value patterns. New **mi** family of commands performs multiple imputation of missing values.	

9.3	Order variables alphabetically by:
	    order *varlist* **, alphabetic**
		Order variables alphabetically by:
		    aorder

9.5	New **merge** syntax, e.g.:
	    use filea.dta
	    merge 1:1 id using fileb.dta
	
	merge automatically sorts the datasets.
	
	The old **merge** syntax still works.
		Old **merge** syntax, e.g.:
		    use filea.dta
		    merge id using fileb.dta
		
		The datasets must be sorted by the matching key before merging

13.1	Factor variables is a new construct in Stata 11; it partially replaces the **xi:** prefix in regression commands – but **xi:** still works.	Use the **xi:** prefix for categorical variables and interactions in regression analyses.
13.3	Stata 11.1 and later (June 2010): The output from regression commands (including **logistic**, **stcox**, and **poisson**) with factor variables will always display the base level if you, just once, give the command: **set showbaselevels on, permanently** and we recommend doing that to enhance transparency.	By default, base levels are omitted from the output.
14.3	You need not specify the **schoenfeld()** and **basesurv()** options to **stcox** to use the **estat phtest** and **stcurve** postestimation commands.	You must specify the **schoenfeld()** and **basesurv()** options to **stcox** to use the **estat phtest** and **stcurve** postestimation commands.

16.1	New random number function names:
	    runiform()
	    rnormal()
		Corresponding Stata 10 functions:
		    uniform()
		    invnormal(uniform())

Changes from Stata 9 to Stata 10

The information is organized by the chapters in ISHR1. In the Stata 10 column we show changes from the description in ISHR1.

Section	Stata 10	Stata 9 (as described in ISHR1)
1.4	Select windowing preferences by: **Edit > Preferences > Manage Preferences >**	Select windowing preferences by: **Prefs > Manage Preferences >**
1.4	Save past commands to a do-file by clicking somewhere in the Review window. Press *Ctrl-A* to highlight all commands and *Ctrl-C* to copy them to the Windows Clipboard. Next open a Do-file Editor window and paste the commands by pressing *Ctrl-V*.	Save past commands to a do-file by right-clicking in the Review window and selecting Copy Review Contents to Clipboard. Next open a Do-file Editor window and paste the Review contents by pressing *Ctrl-V*.
3	Help-files now have the extension **.sthlp** (but Stata 10 still understands the **.hlp** extension)	Help-files have the **.hlp** extension.
5.5	New in Stata 10: Date-and-time variables with new format descriptors (old format descriptors still work).	Stata 9 has date variables, but no facilities for handling time of day. (Some unofficial commands and functions may be helpful).
5.5	In the **date()** function the sequence can be defined as **"DMY"**. (You can use the Stata 9 **"dmy"** specification with version control; see page 1).	In the **date()** function the sequence can be defined as **"dmy"**.
6.1	**save** saves a dataset in Stata 10/11 format; it cannot be read by Stata 8 or 9. To create a dataset that can be read by Stata 8/9, use the **saveold** command	To enable Stata 9 to read a dataset created by Stata 10/11's **save** command, install the unofficial **use10** command by: **ssc install use10**
6.4	Stat/Transfer 9 translates Stata 10/11 files, but Stat/Transfer 8 does not.	
11.10	New graph editor.	
13.3	New **exlogistic** command for exact logistic regression	
14.2	New **risktable** option to **sts graph** displays number at risk below the x-axis.	
14.5	New **expoisson** command for exact Poisson regression	
16.2	New **stpower** command (sample size and power for incidence and survival analysis)	

17.4	To create **r(varlist)**:
	   describe ... **, varlist**
		To create **r(varlist)**:
		   describe ... **, simple**

18.7	New **datasignature** command.	The **datasignature** command in Stata 9.2 is different from the version 10 command.

Changes from Stata 8 to Stata 9

The information is organized by the chapters in ISHR1.

Section	Stata 9 (as described in ISHR1)	Stata 8
1.4	You can have multiple Do-file, Viewer, and graph windows.	You can have only one Do-file editor window, one Viewer Window, and one Graph window
1.4	Data window: String variables are displayed in red, value labels in blue. You may toggle between displaying value labels and codes.	No visual distinction between string variables and value labels. The Data window displays value labels, unless you open it by: **browse, nolabel**
2.1	New Data management manual, [D].	Data management commands in [R].
2.3	New command in Stata 9.1: **adoupdate**	
6.4	New commands in Stata 9: **xmlsave** **xmluse**	
8.2	New function names (the old names still work in Stata 9, but not in Stata 10): **normal()** **invnormal()** **normalden()**	Old function names: **norm()** **invnorm()** **normden()**
8.3	New, less confusing **egen** function names (the old names still work): **total()** **rowmin()** **rowmax()** **rowtotal()** **rowmiss()** **rownonmiss()**	Old function names: **sum()** **rmin()** **rmax()** **rsum()** **rmiss()** **robs()**
9.5	New **sort** option to **merge**.	
10.1	New option to **codebook**: **codebook, compact**	No **compact** option to **codebook**. A useful alternative is the unofficial **summvl**: **findit summvl**
10.3	New command in Stata 9: **proportion**	
10.4	New command in Stata 9: **mean**	
11.7	Marker and line options simplified (the old option names still work)	See Stata 8 syntax in table below
13	**stepwise**	**sw**

Section	Stata 9	Stata 8
13.2	**estat gof**	**lfit**
13.4	**svy:**	Survey analysis commands changed syntax in version 9.
14.3	**estat phtest**	**stphtest**
17.1	**statsby:**	Changed syntax

17.3	New function name (the old name still works in Stata 9, but not in Stata 10):
	    strpos()
		Old function name:
		    index()

Stata 8: Options for defining the appearance of lines, bars and markers
(they are much simpler in Stata 9; see ISHR1 table 11.2):

		Lines		Markers	
Element	Color	Pattern	Width	Symbol	Size
Legend etc.	**color()**				
fill	**fcolor()**				
outline	**lcolor()**	**lpattern()**	**lwidth()**		
Bars, areas	**bcolor()**				
fill	**bfcolor()**				
outline	**blcolor()**	**blpattern()**	**blwidth()**		
Markers	**mcolor()**			**msymbol()**	**msize()**
fill	**mfcolor()**				
outline	**mlcolor()**		**mlwidth()**		
Connecting lines	**clcolor()**	**clpattern()**	**clwidth()**		