

# Introduction to Stata

**Mr. Kongmany Chaleunvong**

GFMER - WHO - UNFPA - LAO PDR

Training Course in Reproductive Health Research

Vientiane, 22 October 2009

# About



- **STATA is modern and general command driven package for statistical analyses, data management and graphics.**
- **STATA provides commands to analyze panel data (cross-sectional time-series, longitudinal, repeated-measures, and correlated data), cross-sectional data, time-series data, survival-time data, cohort study ...**
- **STATA is user friendly.**
- **STATA has an extraordinary set of reference books.**
- **STATA has internet capabilities (installing new features, updating).**

# Starting and stopping Stata

- On desktop computers at CPC, Stata is available only through the Start menu at the bottom-left corner of your screen. Click on Start, Software Run, Statistics, Stata 11 Run. Click on the "X" in the upper right corner to stop Stata. (To get started using Stata on a CPC Unix platform, contact [cpchelp@unc.edu](mailto:cpchelp@unc.edu).) If you have data in memory, and you've changed the data in any way (even sorting), Stata won't let you quit without either saving the data to a permanent file or clearing data from memory. If you're sure you don't want to save the data in memory, you can type:
  - `exit`, `clear` to get out of Stata.

# Opening Stata data files

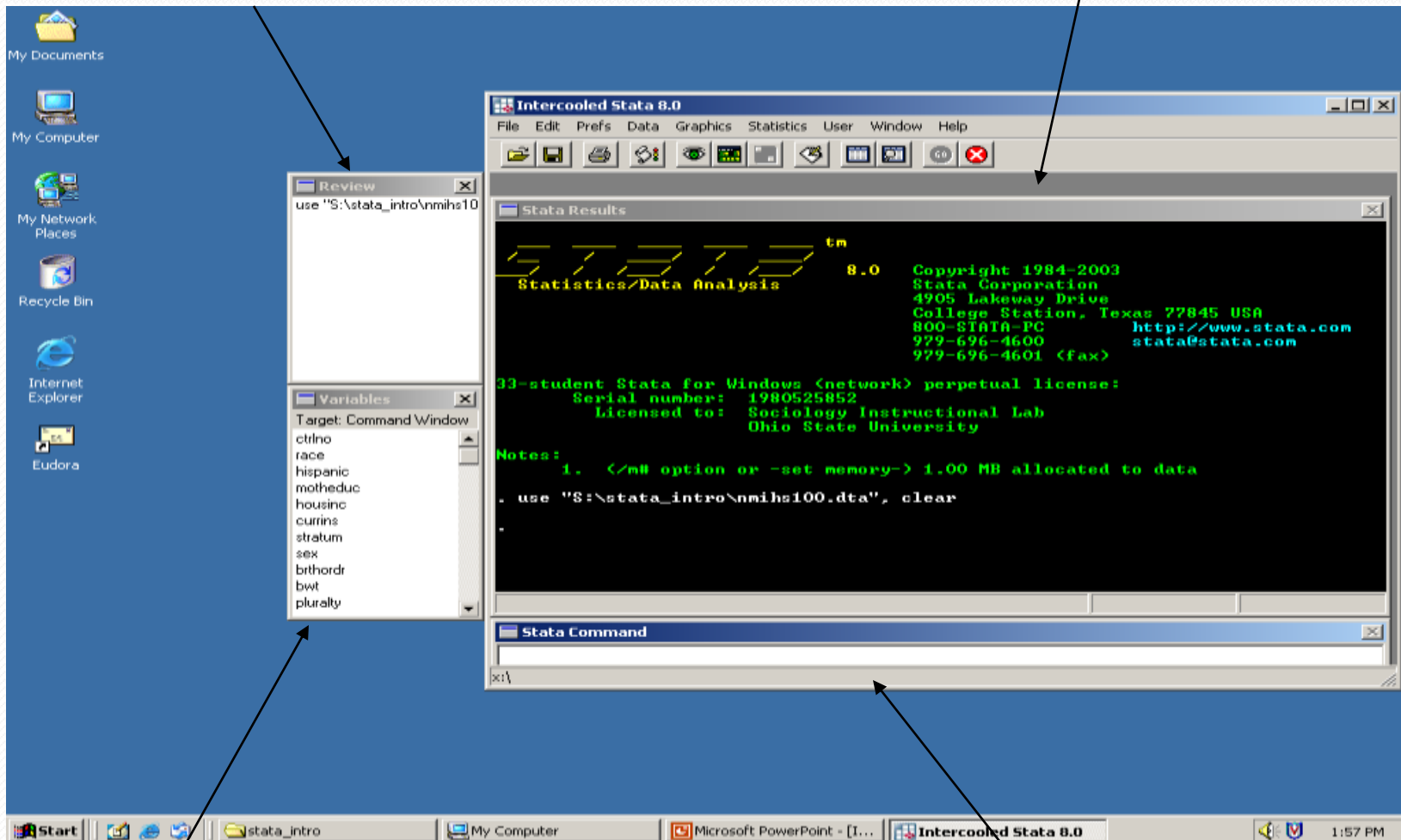
- Stata has its own format for data files
  - extension \*.dta
- Choose File→Open
  - Go to S:\stata\_intro\nmihs100.dta
    - S: in the SIL is not the same as S: in the SRL

# Importing data

- Stata can also read tab-delimited ASCII text files
- Most other software (e.g., Excel) can write tab-delimited ASCII text files
- Let's get data from Excel....
  - From Windows Start button
    - choose All Programs → Office Productivity → Microsoft Office → Excel
  - In Excel
    - choose File → Open
      - find S:\stata\_intro\example1.xls
    - choose File → Save As... → Save as type: Text (tab delimited)
      - Save as X:\example1.txt
  - In Stata
    - In command window, type “clear”  
—gets current data out of memory
    - Choose File → Import → ASCII data created by a spreadsheet
      - Find X:\example1.txt

*Review window:*  
Past commands.  
Click to paste in Command window

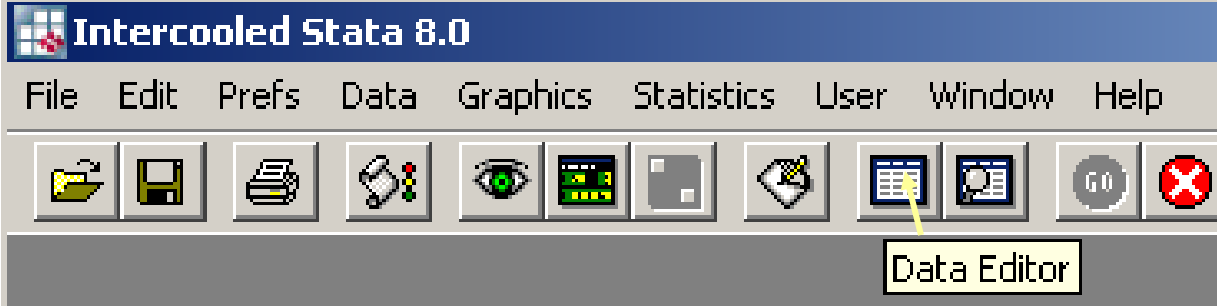
*Results window:*  
Output and past commands



*Variables window:*  
List of variables in open data set.  
Click to paste in Command window.

*Command window:*  
Current command

# Examining data



Undo changes since last "save"

"Save"

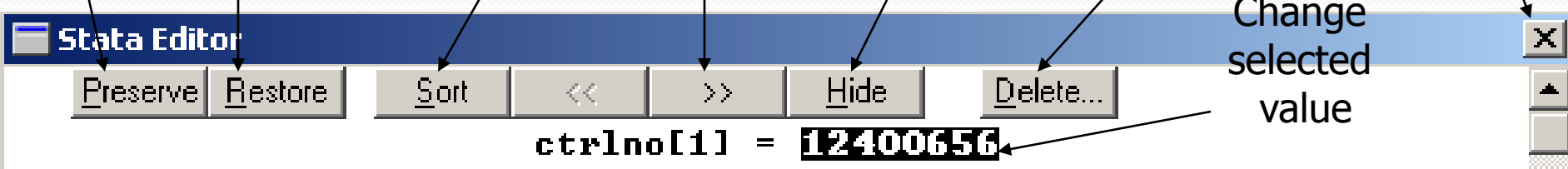
Sort by selected column

Move selected column to the end

Hide selected column

Delete selected columns or rows

Close editor



`ctrlno[1] = 12400656`

	ctrlno	race	motheduc	stratum	bwt	ges
1	12400656	1	12	6	3941	
2	13102530	1	5	6	3572	
3	10510241	1	4	6	3856	
4	12601295	2	12	3	2977	7

# Saving data

- “Preserve” saves only a temporary copy of the data file.
- The original data file is unaffected.
- To save a permanent data file,
  - Choose File→Save As...
  - Navigate to your X: drive
    - X: is where you should save things
    - X: in the SIL is not the same as X: in the SRL
  - Save as “my\_example1.dta”



# 4 windows

- Stata gives you 4 windows: Command, Results, Review, and Variables. Command: type a command here and press Enter
- Results: the results of your command are displayed here
- Review: each command you type is displayed here
  - Click on a command to put it into the command window for editing
  - Double-click on a command to execute it directly
- Variables: lists the variables in memory
  - Click on a variable name to put it into the command window
- You can resize these 4 windows independently, and you can resize the outer window as well. To save your window size changes, click on Edit, Preferences, Save Preferences Set

# Menu Bar

File Edit Data Graphics Statistics User Window Help

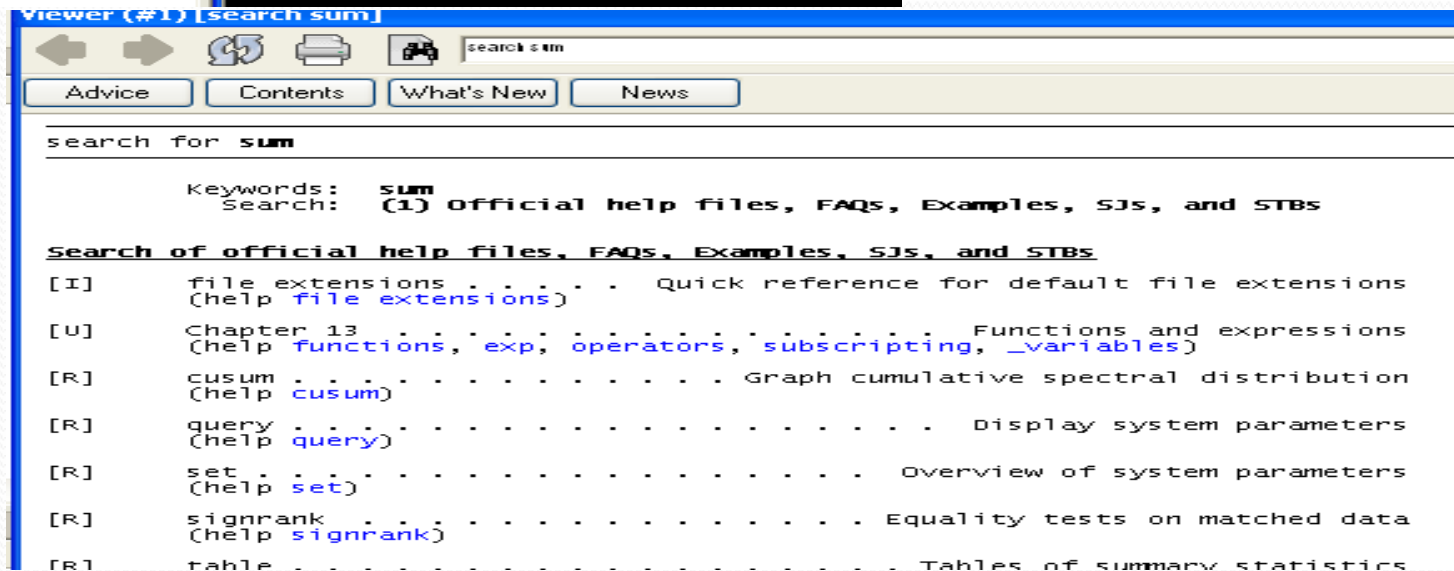
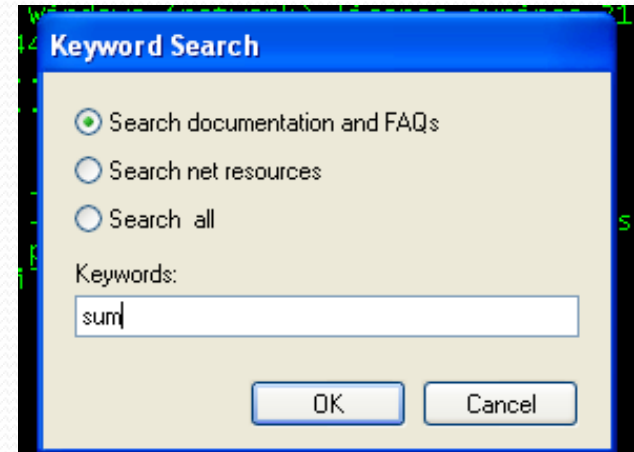
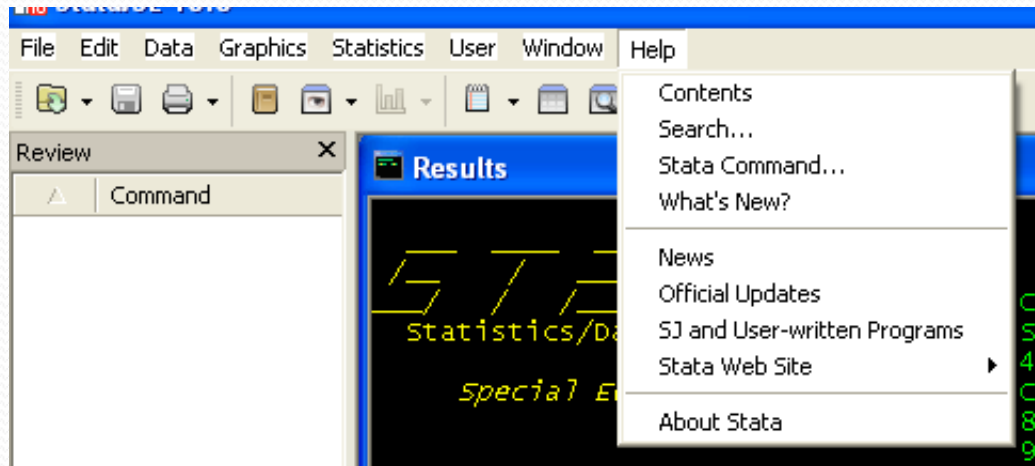
- Stata displays 8 drop-down menus across the top of the outer window: File
  - Open: open a Stata data file (use)
  - Save/Save as: save the Stata data in memory to disk
  - Do: execute a do-file
  - Filename: copy a filename to the command line
  - Print: print log or graph
  - Exit: quit Stata
- Edit
  - Copy/Paste: copy text among the Command, Results, and Log windows
  - Copy Table: copy table from Results window to another file
  - Table copy options: what to do with table lines in Copy Table
- Data, Graphics, Statistics - build and run Stata commands from menus
- User - menus for user-supplied Stata commands (download from Internet)
- Window - bring a Stata window to the front
- Help - The Stata manual set in PDF format plus Stata command syntax and keyword searches

# Tool bar



- The buttons on the button bar are from left to right (equivalent command is in bold): Open a Stata data file: **use**
- Save the Stata data in memory to disk: **save**
- Print a log or graph
- Open a log, or suspend/close an open log: **log**
- Open a new viewer window (to view Help or a log file)
- Bring the graph window to the front (if you've created a graph)
- Open a do-file
- Edit the data in memory: **edit**
- Browse the data in memory: **browse**
- Open the Variables Manager
- Scroll another page when --more-- is displayed: **Space Bar**
- Stop current command or do-file: **Ctrl-Break**

# Sources of help



# Sources of help

- Help menu
  - Command: almost the full reference manual for each Stata command
  - Search: keyword search of the manuals, technical bulletins, and frequently asked questions
  - and lots more!
- Data, Graphics, and Statistics menus
  - build a command with the correct syntax for you
  - lead you to consider options that you might easily overlook
- Manuals

At CPC we no longer carry the printed manual set, since it's available in PDF in the Help menu. However, we do have other Stata and third-party books that focus on specific aspects of Stata programming

# Basic Operations

- **Entering Data**
- **Exploring Data**
- **Modifying Data**
- **Managing Data**

# Entering Data

- **Insheet:** Read ASCII (text) data created by a spreadsheet (.csv files only)
- **Infile:** Read unformatted ASCII (text) data (space delimited files)
- **Input:** Enter data from keyboard
- **Describe:** Describe contents of data in memory or on disk
- **Compress:** Compress data in memory
- **Save:** Store the dataset currently in memory on disk in Stata data format
- **Count:** Show the number of observations
- **List:** List values of variables
- **Clear:** Clear the entire dataset and everything else

# Exploring data

- **Describe:** Describe a dataset
- **List** List the contents of a dataset
- **Codebook:** Detailed contents of a dataset
- **Log:** Create a log file
- **Summarize:** Descriptive statistics
- **Tabstat:** Table of descriptive statistics
- **Table:** Create a table of statistics
- **Stem:** Stem-and-leaf plot
- **Graph:** High resolution graphs
- **Kdensity:** Kernel density plot
- **Sort:** Sort observations in a dataset
- **Histogram:** Histogram for continuous and categorical variables
- **Tabulate:** One- and two-way frequency tables
- **Correlate:** Correlations
- **Pwcorr:** Pairwise correlations
- **Type:** Display an ASCII file



# Modifying Data

- **label data:** Apply a label to a data set
- **Order:** Order the variables in a data set
- **label variable:** Apply a label to a variable
- **label define:** Define a set of a labels for the levels of a categorical variable
- **label values:** Apply value labels to a variable
- **List:** Lists the observations
- **Rename:** Rename a variable
- **Recode:** Recode the values of a variable
- **Notes:** Apply notes to the data file
- **Generate:** Creates a new variable
- **Replace:** Replaces one value with another value
- **Egen:** Extended generate - has special functions that can be used when creating a new variable

# Labeling variables

- To add a descriptive label to a variable
  - Data → Labels → Label variable
- Add these labels to these variables:
  - *bwt* : “Birth weight, in grams”
  - *smoke* : “Did mother smoke during pregnancy?”

# Labeling values

- Many variables are *dummy variables*
  - two values: 0 and 1
    - *e.g.*, “Did the mother smoke?” Yes (1) or no (0).
- To add labels to dummy values
  - **Data→Labels→Label Values→Define or Modify Value Labels**
  - Define label name: “dummy”
  - Add values
    - 1 means “yes”
    - 0 means “no”
- Now tell Stata that *smoke* is a dummy variable
  - **Data→Labels→Label Values→Assign value label to variable**
- Look at *smoke* in the Data Editor
  - and double-click it

# Generating and Recoding Variables

```
gen quality=0  
recode quality 0=1 if VA==1 or  
replace quality=1 if VA==1  
gen
```

# Creating a new variable

- According to the National Institutes of Health,
  - low birth weight (LBW)
    - < 2500 grams (5.5 pounds)
- Let's create a dummy variable for LBW
- Data
  - Create or change variable
  - Create a new variable

generate - Generate a new variable

Main if/in

Generate variable:  
LBW =

Contents:  
bwt<2500 Create ...

Generate variable as type: byte

Attach value label: dummy 21

# Managing Data

- **Pwd:** Show current directory (pwd=print working directory)
- **dir** or **ls:** Show files in current directory
- **cd** Change directory
- **keep if:** Keep observations if condition is met
- **Keep:** Keep variables (dropping others)
- **Drop:** Drop variables (keeping others)
- **append using:** Append a data file to current file
- **Merge:** Merge a data file with current file

# Syntax: Commands

Command	Recommended	Usage
<u>d</u> escribe	d	Describe data in memory
<u>g</u> enerate	gen	Create new variables
<u>g</u> raph	graph	Graph data
<u>h</u> elp	h	Call online help
<u>l</u> ist	l	List data
<u>r</u> egress	reg	Linear regression
<u>s</u> ummarize	sum	Descriptive statistics
<u>s</u> ave	save	Save data in memory
<u>s</u> ort	sort	Sort data
<u>t</u> abulate	tab	Tables of frequencies
<u>u</u> se	use	Load data into memory

# Do file

- *Do-files are created with the do-file editor or any other text editor. Any command which can be executed from the command line can be placed in a do-file*
- *To open a do file editor: **Window - Do-file Editor** or **Ctrl + 8***
- **set more off**
- **use hsb2, clear**
- **generate lang = read + write**
- **label variable lang "language score"**
- **tabulate lang**
- **tabulate lang female**
- **tabulate lang prog**
- **tabulate lang schtyp**
- **summarize lang, detail**
- **table female, contents(n lang mean lang sd lang)**
- **table prog, contents(n lang mean lang sd lang)**
- **table ses, contents(n lang mean lang sd lang)**
- **correlate lang math science socst**
- **regress lang math science female**
- **set more on**



# Do file – cont.

*Look at the commands in a do-file that contains:*

- **. type hsbbatch.do**

*To run the do-file.*

- **do hsbbatch**
- *From do file, choose **Tools - Do***