140.641 Survival Analysis Computer Lab I Feb 4

By Qing Li qli@jhsph.edu

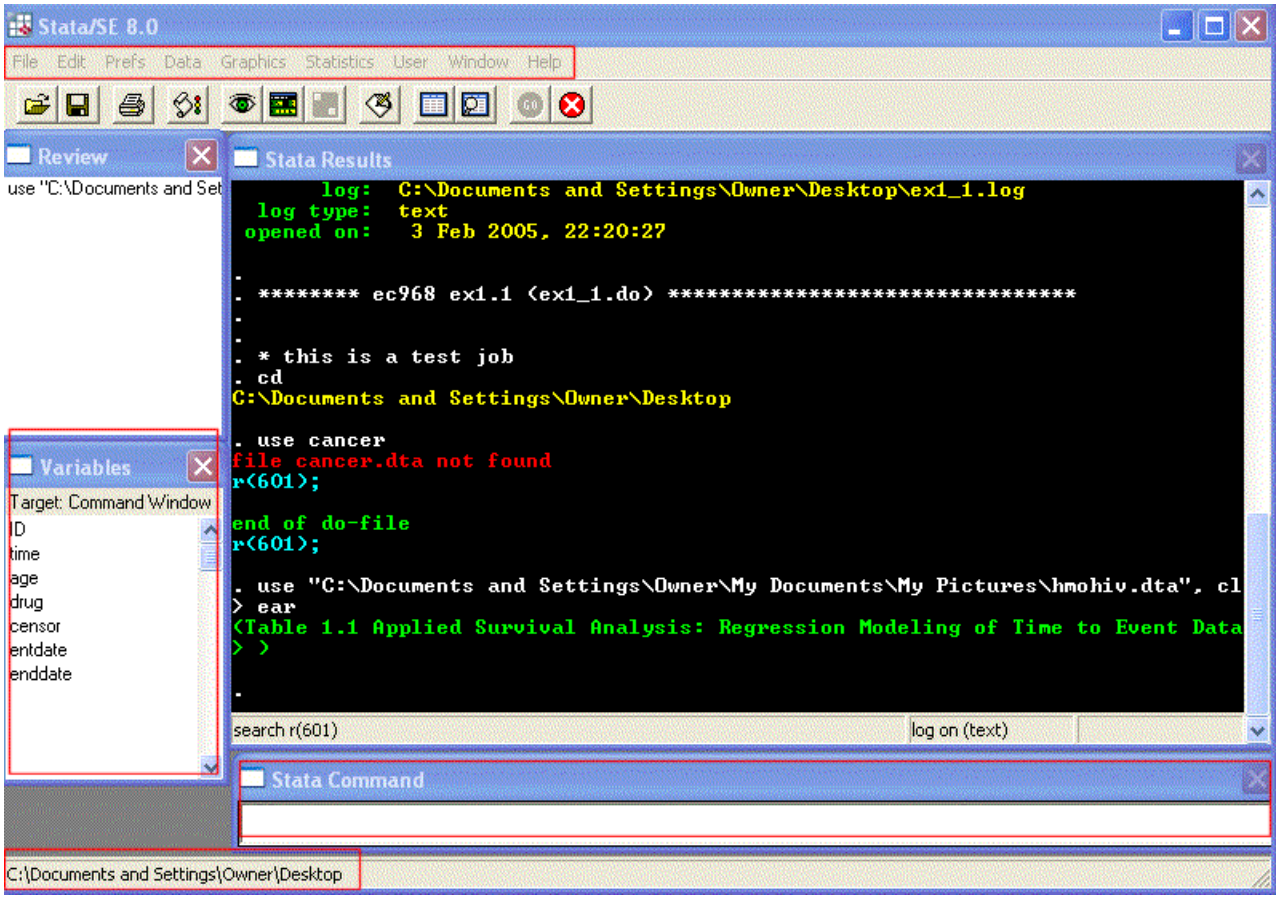
Objective:

- Brief introduction on statistical software STATA and R
- Programming building block and tips
- How to get help?

STATA:

Learn STATA the same way you learn about Microsoft Word

- By click, click, click (look though each item on menu bar)
- Communicate with your program
- KISS (Keep it stupidly simple), try to build on your knowledge
- Keep your favorite reference handy
- Use help function often



Basic STATA command[1]

Searching for help and programs (official and user-written)

help, **findit**, **ssc**

Utilities

log, **cd**, **dir**, **erase**, **mkdir**, **copy**, **display**, **delimit**

Reading data in, and saving it in Stata format

use, **insheet**, **infile**, **compress**, **save**

Setting memory size

memory, **clear**, **set memory**, **set matsize**

Inspecting, summarising and describing data

describe, **list**, **inspect**, **summarize**, **tabulate**, **correlate**, **sort**

Creating new variables and re-organising data

generate, **replace**, **egen**, **recode**, **label**, **keep**, **drop**, **rename**, **expand**, **by**

[see also **help functions**, **help exp**, **help operators** about functions and expressions]

Statistical analysis, estimation and graphics

regress, **logit**, **probit**, **tobit**, **cnreg**

predict, **test**, **testnl**, **lrtest**, **lincom** [also see **help est** about retrieving estimates]

ltable, **st**

stset, **stdes**, **streg**, **stcox**, **sts list**, **sts generate**, **sts graph**, **sts sts**, **stsplit**, **stgen**

xtlogit, **xtclog**

Graphics

graph

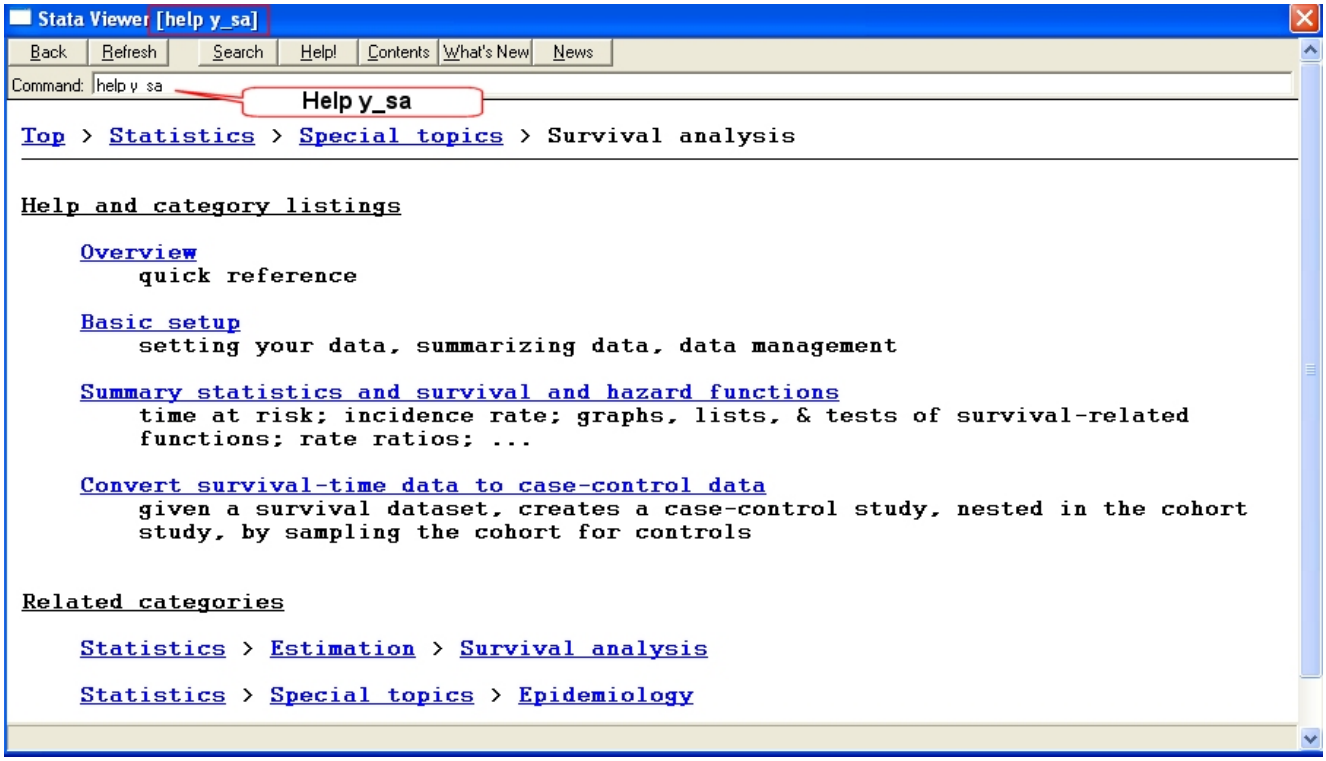
[NB standard graph options (for labelling etc) are typically applicable to graphs produced by other commands]

Survival Analysis

known as the st series of commands (**st** stands for survival time);
you must first **stset** your dataset

Help Command: **help y_sa**

[1] **(http://www.iser.essex.ac.uk/teaching/degree/stephenj/ec968/#_Toc520705917)**



R:

To learn R, we have to put ourselves into a programming mindset

- **Communicate with your program**
- **KISS (Keep it stupidly simple), try to build on your knowledge**
- **Keep your favorite reference handy**
- **You need to THINK about your code**

Building Block of Programming:

- **Data structure**

  Named data structure: vector, matrix, factor, list, data frame, function, object

- **Grouped expression**

  A = B, A<B, 3*c^2 …

- **Control sequence**

  if(){ }else{}, loops (for, while)

- **Function**
- **Package**
- **Program environment**

How to…? All about asking the right question:

- **How to read documentation**

  http://www.biostat.jhsph.edu/~qli/biostatistics_r_doc/library/base/html/00Index.html

- **How to access information**

  Get information about programming structure using

  1. **str() for any R object**
  2. **length() for vector, dim() for matrix, names() for named object**

  Get data stored in data structure

  1. **[3], [[3]] for vector, list**
  2. **var$name for named attribute**

- **How to assign value to an object**

  Is this as simple as using "<-" or "="?

- **How to transfer data from one data structure to another**
- **How to debug your program**

  print your value: print(), paste(),

- **How to use help(), source(), search engine for specific questions**

Reference:

R introduction site by Prof. Karl Broman
http://www.biostat.jhsph.edu/~kbroman/Rintro/

Statistical Computing Course website by Prof. Brian Caffo
http://www.biostat.jhsph.edu/~bcaffo/statcomp/index.html