11.220 Quantitative Reasoning & Statistical Methods for Planners I
Spring 2009

Computer lab #1   **Feb 20th, 2009**

Input data into STATA & Descriptive Statistics

STATA commands in today's class

cd	Change directory
dir or ls	Show files in current directory
insheet	Read ASCII (text) data created by a spreadsheet (.csv)
infile	Read unformatted ASCII (text) data (.raw or .txt)
infix	Read ASCII (text) data in fixed format (.fix)
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
use	Load a Stata-format dataset (filename.dta)
count	Show the number of observations
list	List values of variables
clear	Clear the entire dataset and everything else
memory	Display a report on memory usage
set memory	Set the size of memory
browse	Check the input to your dataset without editing
run	Run a do. file
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

Scripts we use in the real Command Window   Note: STATA is case-SENSITIVE!

cd *E:\MIT\09Spring\STATALAB\DATA*   (change this part to your own local directory)
dir
insheet *using* hs0.csv
insheet gender id race ses schtyp prgtype read write math science socst **using** hs0_noname.csv,
clear
infile gender id race ses schtyp *str10* prgtype read write math science socst **using** hs0.raw**, clear**
infix id 1-2 a1 3-4 t1 5-6 gender 7 a2 8-9 t2 10-11 tgender 12 **using** schdat.fix**, clear**
describe
compress
save hsb10**,replace**
memory
set memory 5m

Descriptive Statistics and Basic Plotting

use hs0, clear
describe   /*talk about value label, white = 1, Asian = 2, etc*/
list
list gender-read
codebook
summarize
summarize read math science write
summarize write, detail
sum write **if** read>=60
sum write **if** prgtype=="academic"   / * == is not =*/ Do use quotation mark for string
sum write in 1/40   /specify the range, from 1 to 40/

tabulate prgtype**, summarize(**read**)**
tab prgtype**, summarize(**write**)**   /* note: tab is abbreviation of tabulate */

tabstat read write math, **by(**prgtype**) stat(**n mean sd**)**
tabstat write, **stat(**n mean sd p25 p50 p75**) by(**prgtype**)**

stem write
histogram write**, normal**
kdensity write**, normal**
graph box write
graph box write**, over(**prgtype**)**   /* here it is over() instead of by(), if use by there will be 3
separate graphs*/

correlate write read science
pwcorr write read science**, obs**
scatter write read
scatter write read**, jitter(2)**
graph matrix read science write**, half**
graph matrix read science write**, half** by(prgtype)

Note: With reference to Bruin, J. 2006. newtest: command to compute new test. UCLA:
Academic Technology Services, Statistical Consulting Group.