11.220 Quantitative Reasoning & Statistical Methods for Planners I

Spring 2009

Computer lab #2 — Mar 6th, 2009

Modify data: add labels, create new variables

To use flash drive on Linux system

```
type "add consult" in the terminal
type "tellme root" and pay attention to the password it gives you
type "attach-usb" and then enter that password
The path will be "/mnt/usb/foldername"
type "detach-usb", and give the same password to detach f-drive
```

STATA commands used in today's class

log	Keep a log file to track your operation and outcomes
codebook	Show codebook information for file
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
encode	Encode string into numeric 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

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

```
cd E:\MIT\09Spring\STATALAB\DATA   (change this part to your own local directory)
use hs0, clear
log using log1, text replace

/// label the variable and value of "schtyp"
order id gender
label variable schtyp "The type of school the student attended."
label define scl 1 public 2 private
label values schtyp scl
codebook schtyp
list schtyp in 1/10
list schtyp in 1/10, nolabel

/// create a new numeric version of the string variable "prgtype"
```

```
encode prgtype, gen(prog)
label variable prog "The type of program in which the student was
enrolled."
codebook prog
codebook prgtype
list prog in 1/10
list prog in 1/10, nolabel

/// replace the missing value in science score
list science if science ==  .
recode science  . = 50
list science if science == 50
notes science: missing values are replaced by 50
notes

rename socst social /// change the name of socst to social

///generate a new variable to show the total score
gen total = read + write + social
summarize total

///recalculate the value for total score to include all classes
replace total = read + write + math + science + social
summarize total
label variable total "total score"
codebook total

///Change the scores into grades
recode total (0/150=1 F) (150/200=2 D) (200/250=3 C) (250/300=4 B)
(300/500=5 A), gen(grade)
codebook grade
label variable grade "combined grades for all classes"
list read write social math science grade in 1/10
list read write social math science grade in 1/10, nolabel

///Create standardized version of variables
egen zread = std(read)
summarize zread
list read zread in 1/10

///Calculate subgroup means and assign to each observation
egen rmean = mean(read), by(race)
list read race rmean in 1/10
egen mread = median(read), by(prog)
list read prog mread in 1/10

save hs1,replace
```

Exercises

1: Label the value of gender, male as 0, female as 1. List your results.

2: Calculate the total score of read, write and social, then change the scores into grades using the following rule:

(0/80=0 F) (80/110=1 D) (110/140=2 C) (140/170=3 B) (170/300=4 A)

List your results.

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