Introduction to Stata:
Exercises

March 4, 2010

HINT: Use 'cd' (change directory) command to go to the directory with files and avoid typing paths to data (Answers to exercises assume that you are in correct directory). For instance:

cd C:\files

Exercise 1 : Import/Export

1. Load gss.tab (you can figure out from **.tab** extension that it is TAB separated values, but you would never know if it had extension **.dat** or **.raw**. To figure out delimiter open file in text editor and check out the delimiter). When you import data remember about 'clear' option– you cannot import data if there is already any data in Stata memory. Use 'describe', 'summarize', and 'edit' to see how the data look. Save data in Stata format as 'mydata.dta' on Desktop (remember about 'replace' option).

Exercise 2 : Names/Labels

1. Let's use gss.csv
2. If you say 'describe' it does not look nice: a bunch of v's there. Let's make it looking nice. Use 'rename' and 'label'. Here are current variable names, desirable variable names, and labels.

   v1, marital, marital status v2, age, age of respondent
   v3, educ, education
   v4, sex, respondent's sex
   v5, inc, income
   v6, happy, general happiness
   v7, region, region of interview

3. Now if you say 'describe' it looks much better.

Exercise 3 : Variables

Starting with this exercise let's have all the code in a new do-file. Let's name it mydofile.do and save it on Desktop.

1. For this exercise use gss.dta.
2. Generate age^2 from age.
3. Generate a divorced/separated dummy variable that will take on value 1 if a person is either divorced or separated and 0 otherwise

Exercise 4 : Descriptive Statistics

1. For this exercise use gss.dta.
2. Let's tabulate region with and without labels
3. Let's cross-tabulate region with gender and show gender percent by region
4. Let's produce a histogram of income

Solution to Exercise 1 : Import/Export

1. Load gss.tab (you can figure out from **.tab** extension that it is TAB separated values, but you would never know if it had extension **.dat** or **.raw** (remember about 'clear' option). To figure out delimiter open file in text editor and check out the delimiter). Use 'describe', 'summarize', and 'browse' to see how the data look. Save data in Stata format as 'mydata.dta' on Desktop (remember about 'replace' option).

```
insheet using gss.tab, clear
describe
summarize
browse
save "C:\Documents and Settings\ YOUR NAME\Desktop\mygss.dta", replace
```

Solution to Exercise 2 : Names/Labels

1. Let's use gss.csv
2. If you say 'describe' it does not look nice: a bunch of v's there. Let's make it looking nice. Use 'rename' and 'label'. Here are current variable names, desirable variable names, and labels.

   v1, marital, marital status
   v2, age, age of respondent
   v3, educ, education
   v4, sex, respondent's sex
   v5, inc, income
   v6, happy, general happiness
   v7, region, region of interview

3. Now if you say 'describe' it looks much better.

```
insheet using gss.csv, clear

ren v1 marital
ren v2 age
ren v3 educ
ren v4 sex
ren v5 inc
ren v6 happy
ren v7 region

la var marital "marital status"
la var age " age"
```

```
la var educ "education"
la var sex "respondent's sex"
la var inc "income"
la var happy "self-reported happiness"
la var region "region of interview"


des
```

Solution to Exercise 3 : Variables

Starting with this exercise let's have all the code in a new do-file. Let's name it mydofile.do and save it on Desktop.

1. For this exercise use gss.dta.
2. Generate age^2 from age.
3. Generate a divorced/separated dummy variable that will take on value 1 if a person is either divorced or separated and 0 otherwise

```
use gss.dta, clear


gen age2=age^2


tab marital
tab marital, nola


gen divsep=.
 replace divsep=1 if marital==3|marital==4
replace divsep=0 if marital<3 |marital==5
```

or use recode:

```
capture drop divsep /* see "help capture"*/
recode marital (1 2 5=0) (3 4=1), gen(divsep)
```

Solution to Exercise 4 : Descriptive Statistics

1. For this exercise use gss.dta.
2. Let's tabulate region with and without labels
3. Let's cross-tabulate region with gender and show gender percent by region

4. Let's produce a histogram of income

```
use gss.dta, clear
tab region, nola
tab region sex, row
hist inc
```