\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* SOC7717 (S19-Fan). Instructions on opening a Stata data file \*

\* and running basic commands \*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

// program: soc7717\_IntroStata.do

// task: get to know Stata

// project: event history analysis and sequence analysis

// author: Wen Fan / 2019-01-17

\*\*\* 0. program setup

version 14.1

/\* Sometimes newer versions of Stata change the way in which a statistic is

computed. Therefore different versions may produce different results. \*/

clear all

set linesize 80

macro drop \_all

\*\*\* 1. open data

cd "ENTER WORKING DIRECTORY HERE"

/\* The 'cd' command tells Stata what directory you are working in. The name of

the current working directory is listed beneath the variable window. You can

display the current working directory by entering the 'pwd' command. \*/

capture log close

log using soc7717\_IntroStata\_Log, replace text

/\* The 'log' command tells Stata to start a log file. Log files

record everything that happens during a given session, including the

commands you entered and the results you obtained. \*/

use GSS2014.dta, clear

/\* 'use' tells Stata what file to open. Notice that if we had not named a

working directory earlier, we'd have to specify the full path name \*/

\*\*\* 2. basic commands

desc

/\* 'describe' summarizes the data set in memory \*/

codebook joblose

/\* the 'codebook' command lists information about your variables \*/

note: The General Social Survey (GSS) is a sociological survey used to ///

collect data on demographic characteristics and attitudes of residents of ///

the United States.

/\* Attach a note to the data set. \*/

sum age joblose

/\* 'summarize' prompts Stata to calculate descriptive statistics. The detail

option (e.g., 'sum age, detail') can be used to obtain information about

skewness, kurtosis, etc. \*/

tab sex

tab joblose sex

tab joblose sex, row col cell chi2

/\* The first 'tab' command creates a one-way frequency table. The second 'tab'

command cross-classifies sex and joblose. The 'row' and 'col' options

can be used to obtain percentages within rows and columns, respectively \*/

histogram age

/\* Creates a histogram of age \*/

kdensity age

kdensity age if sex == 2, addplot(kdensity age if sex == 1)

/\* Generates kernel density plots reflecting the distribution of age in the

GSS sample. The second command tells Stata to overlay two separate

density plots, where the first plot pertains to women (if sex == 2) and the

second pertains to men (if sex == 1) \*/

graph box educ

graph box educ, over(sex)

/\* Generates box plots of years of schooling, first for the entire sample

and then separately by gender \*/

gen insecure = (joblose < 3) if joblose < .

label var insecure "insecure about one's job: < 3"

recode educ (0/15 = 0) (16/20 = 1), gen(college)

label var college "indicator of college degree"

/\* Generate a binary indicator of job insecurity and of college attainment \*/

tab college

/\* Check to make sure the recode worked properly \*/

label define college 0 "0 < college" 1 "1 >= college"

label values college college

/\* Attach value labels to the college variable \*/

\*\*\* 3. variable manipulation

/\* create a variable indicating each case's deviation from the mean income,

call this variable dev\_pinc \*/

egen avg\_pinc = mean(pinc)

gen dev\_pinc = pinc - avg\_pinc

lab var dev\_pinc "deviation from mean income"

/\* load and install contributed commands; "findit" is more thorough and complete

than "ssc install" \*/

ssc install egenmore

\*findit egenmore

help egenmore

\*\*\* 4. introduction to OLS

/\* regress income on age for men; display stored results; store the results

under the name m1 \*/

regress ln\_pinc age if sex == 1

ereturn list

estimates store m1, title(Model 1)

/\* regress income on age and race for men; one way to enter race is to create

dummy variables manually, but Stata makes it easier by the i. and c.

operators; store the results under the name m2 \*/

tab race

regress ln\_pinc age i.race if sex == 1

estimates store m2, title(Model 2)

/\* compare the two sets of results, add fit statistics to the table \*/

esttab m1 m2, b(%10.4f) se scalars(N r2 F ll) mtitles

/\* write tabulated results to Excel (you could also write them to a simple text

file, Word, or LaTeX) \*/

esttab m1 m2 using "model comparison\_esttab.csv", b(%10.4f) se ///

scalars(N r2 F ll) mtitles title("Model Comparison of m1-m2")

outreg2 using "model comparison\_outreg", excel bdec(3) alpha(0.001, 0.01, ///

0.05) symbol(\*\*\*, \*\*, \*)

log close

exit