**EPIDEMIOLOGY 340.600: STATA PROGRAMMING**

**Lab 1**

This lab is optional; you are NOT required to complete these questions. Please use this lab as an opportunity to review the course material and prepare yourself for the homework questions. Sample responses to the lab questions will be provided separately.

1. Start Stata and open your do-file editor or your favorite text editor.
2. You will now write your first do-file. In the text editor or Stata do-file editor, write the header. That is, start logging, describe the purpose of this script as a comment, specify the version for which the script was written, and clear memory. Your homework scripts should start this way. (Your other scripts too!) Code for a .do file header was given in lecture 1.

capture log close

log using lab1.log

//The purpose of this script is just to give you an idea of what a proper header should look like.

//if you're using an earlier version of Stata,

//put "version 15" or whatever

version 16

clear all

set more off

1. We want to load transplants.dta. But before that, let’s check your working directory. See the bottom left side of the console (the main Stata window). You may also type pwd on the console. Is it where your transplants.dta is located? If not, use one of these two methods:

1) Next time, launch Stata by double-clicking on transplants.dta.

2) Type cd c:/your/actual/path on the console (NOT in your do file).

1. Now let’s get back to your do file. Load transplants.dta

use transplants.dta

1. How many observations does the dataset have?

//method 1

count

//method 2

disp \_N

1. How many adults (age>18) does the dataset have?

count if age>18

1. How many observations have missing bmi?

count if missing(bmi)

1. Generate a new variable called agecat. The value of agecat is 1 for patients younger than 18, 2 for those from 18 to 65, and 3 for those older than 65.

//method 1

generate agecat=.

replace agecat=1 if age<18

replace agecat=2 if age>=18 & age<=65

replace agecat=3 if age>65 & !missing(age)

//method 2

gen byte agecat = 1+(age>=18)+(age>65) if !missing(age)

1. What are the means of age and bmi?

sum age bmi

1. Now preserve your dataset.

preserve

1. Drop all patients who are younger than 18 or older than 65, or have missing value for age.

drop if age<18 | age>65 | missing(age)

1. Again, what are the means of age and bmi? Restore the dataset. Yet again, what are the means of age and bmi?

sum age bmi

restore

sum age bmi

1. What happened? Leave a comment on your do-file explaining what you just did. Remember, your homework (and all other) scripts should have 1-2 comment(s) per major task.

// This black magic enables time travel.

1. Lab 1 is almost over. Let Stata say “that was easy!”

di "that was easy!"

//Black magic always \*seems\* easy... until you pay the price

1. The last bit. Close your log file. Never forget to close log files!

log close

1. You have all your commands so far in your do file, right? Run your entire do file and make sure your do file does exactly the same thing.