**EPIDEMIOLOGY 340.600: STATA PROGRAMMING**

**Lab 5**

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, open your do-file editor, write the header, and load transplants.dta.
2. Let’s merge this dataset with the donor dataset. First, merge with donors\_recipients.dta, and then with donors.dta, without specifying any options.

merge 1:1 fake\_id using donors\_recipients

1. What does Stata say? Interpret the output.

. merge 1:1 fake\_id using donors\_recipients

Result # of obs.

-----------------------------------------

not matched 4,000

from master 0 (\_merge==1)

from using 4,000 (\_merge==2)

matched 6,000 (\_merge==3)

-----------------------------------------

From the bottom: 6000 observations were successfully merged. 4000 observations from the using dataset (= donors\_recipients.dta) were NOT matched with the master dataset (= transplants.dta) but brought in anyways. 0 observation from the master dataset were not matched (i.e., all observations from the master dataset were matched.)

1. transplants.dta is our study population. We don’t want to bring in extra observations by merging. Use the option keep and make sure we don’t bring in extra observations from donors\_recipients.dta.

use transplants, clear

merge 1:1 fake\_id using donors\_recipients, keep(master match) nogen

1. Let’s move forward and merge with donors.dta.

merge m:1 fake\_don\_id using donors, keep(master match)

1. Now we want to calculate the mean age and the number of patients at each center. Preserve the dataset and collapse it by ctr\_id. Explore the collapsed dataset using list.

preserve

collapse (mean) age (count) n=fake\_id, by(ctr\_id)

1. Restore the dataset. The plan has changed. We want to calculate these statistics in ECD cases and non-ECD cases separately (use the variable don\_ecd). Calculate the mean age and the number of ECD patients and non-ECD patients at each center.

restore

collapse (mean) age (count) n=fake\_id, by(ctr\_id don\_ecd)

1. After the collapse, each center has two observations. One for ECD cases and another for non-ECD cases. Reshape the dataset into a wide format (i.e., each center has only one observation).

reshape wide age n, i(ctr\_id) j(don\_ecd)

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