**EPIDEMIOLOGY 340.600: STATA PROGRAMMING**

**Lab 6**

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. Get a 10% random sample of the dataset. Specifically, follow these steps. (1) Set a seed number. (2) Generate a variable that includes a random number between 0 and 1 following a uniform distribution. (3) Sort by the random variable. (4) Keep the first 10% observations and drop the rest. (5) Drop the random variable.

set seed 2021

gen rdm=runiform()

sort rdm

keep if \_n<=\_N/10

drop rdm

1. Clear and re- load transplants.dta.
2. Generate a variable called fake\_age which is a normally distributed random variable with mean and standard deviation equal to the mean and standard deviation of the actual age variable.

sum age

gen fake\_age=rnormal(r(mean), r(sd))

sum age fake\_age

1. Make a scatter plot of peak PRA by age in transplant recipients. Does it look like there's a relationship between peak PRA and age, and if so, what is the relationship?

graph twoway scatter peak\_pra age //full syntax

tw sc peak\_pra age //abbreviated syntax

1. The graph of proportion of ECD transplants by age from the lecture was a little messy. Remake the graph with the age rounded to the nearest ten years.

(From the lecture)



use transplants, clear

collapse (mean) don\_ecd, by(age)

graph twoway line don\_ecd age

(After rounding)



use transplants, clear

gen age10 = round(age, 10)

collapse (mean) don\_ecd, by(age10)

graph twoway line don\_ecd age10

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.