**Stata Assignment 1: Cumulative Stata Assignment**

**Sarah Pickersgill**

At the end of some Stata lectures, there will be a few short exercises designed to give you practice with the day’s lecture material and practice with commands and concepts that are not included in the regular homework assignments. All of these exercises must be uploaded as one assignment at the end of the 5-week session.

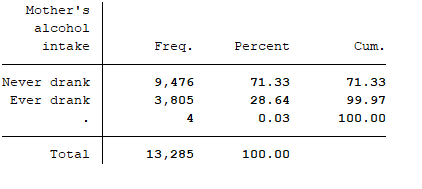
For the 1st 2 exercises, please turn in your coding (these exercises accompany the lecture “Categorizing, recoding, creation of new variables” and are based on the the vipcls dataset):

1. Create a new variable (anyetoh) that combines alcohol information for the trimesters. The variable should = 0 if the patient never drank during the 1st 2 trimesters and =1 if the patient drank at all during the 1st 2 trimesters.

gen anyetoh=.

replace anyetoh=0 if etoh1==6 & etoh2==6

replace anyetoh=1 if etoh1<6 | etoh2<6



1. Create a new variable called enrdate that combines enrmo, enrdy, and enryr into a date.

// creating date variables //

gen enrdate=mdy(enrmo, enrdy, enryr)

format enrdate %d

1. Practice with merging datasets\* (these exercises accompany the lecture “Dataset manipulation in Stata”):
   1. Download wine\_id.dta and wine\_smoke.dta from the course website.
   2. Open wine\_id.dta and use merge with the ,update replace option to merge this dataset (matching on idno) with wine\_smoke.dta.
   3. Answer the following questions:
      1. Were any observations updated in wine\_id.dta after merging?

Yes

* + 1. How many?

One observation

* + 1. Which idno(s) had their data updated? Which variable(s) were updated?

Idno 8 was updated. The variable updated was wine. The value was updated from 0 to 1.

\*You do not need to turn in any code or Stata output for this question.

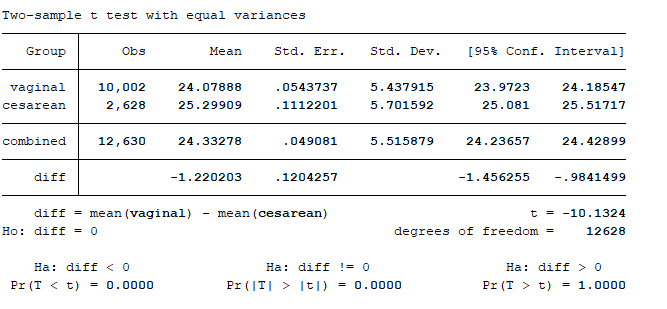
(Just recording here for my own records:)

use wine\_id.dta

merge 1:1 idno using wine\_smoke, update replace

1. Using a t-test, determine if mother’s age differs for mothers who delivered vaginally vs via c-section (in the vipcls dataset). Include your Stata output and use words to report your results in one sentence. (This exercise accompanies the lecture “Stata analysis and regression techniques.”)

ttest momage, by (deltype)



The t-test provides strong statistical evidence (p-value < 0.0001) that the mean mother’s age is not equal for the two groups (vaginal vs. C-section). The t-test estimates that mothers who delivery vaginally are, on average, 1.22 years younger than mothers who deliver via C-section (95% CI: [-1.46 , -0.98]).

1. Using the longitudinal dataset sample\_hpv.dta, answer the following questions (this exercise accompanies the lecture “working with longitudinal data in Stata”):
   1. How many unique individuals are included in this dataset?

There are 30 unique individuals in this data set

* 1. What is the maximum number of visits per subject?

The maximum number of visits for an individual is 18.

* 1. How many subject had exactly 7 visits?

Six subjects had exactly 7 visits.

\*You do not need to turn in any code or Stata output for this question.

(Just recording here for my own records:)

codebook idno

egen visnum=rank(\_n), by(idno)

egen vistot=count(\_n), by (idno)

ta vistot if visnum==7

1. Practice with the reshape command (this exercise accompanies the lecture “working with longitudinal data in Stata”):
   1. Using the same dataset (sample\_hpv.dta), keep only the first visit for each subject.

Save sample\_hpv2

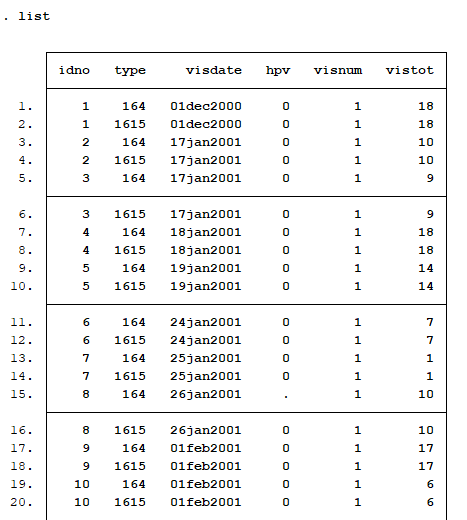
clear

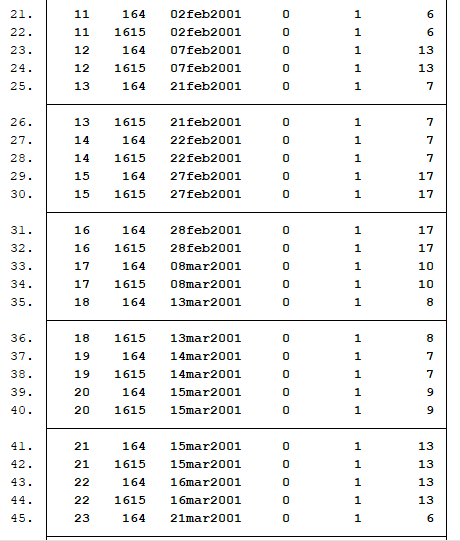
use sample\_hpv2 if visnum==1

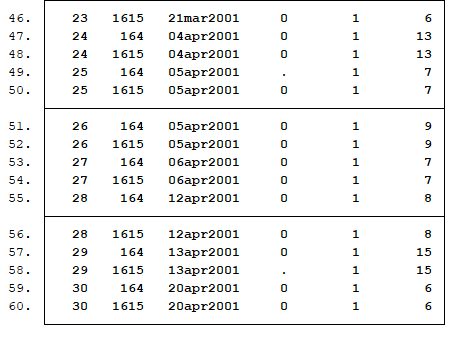
save sample\_hpv1

* 1. Using reshape, convert the data to long format, and turn in your reshape output.

reshape long hpv, i(idno) j(type)







**STATA CODE:**

version 15

cd "C:\Users\Sarah\Epi510\Stata"

use vipcls

// Missing Data: changing all -1 values to (.) //

recode \_all (-1 = .)

// Create variable anyetoh that combines alcohol info //

gen anyetoh=.

replace anyetoh=0 if etoh1==6 & etoh2==6

replace anyetoh=1 if etoh1<6 | etoh2<6

// Label new alcohol variable //

label variable anyetoh "Mother's alcohol intake"

label define l\_anyetoh 0 "Never drank" 1 "Ever drank"

label values anyetoh l\_anyetoh

// creating date variables //

gen enrdate=mdy(enrmo, enrdy, enryr)

format enrdate %d

// t-test for mother's age & deltype //

//recoding to 0/1 //

recode deltype 1 = 0

recode deltype 2 = 1

label variable patid "Patient ID"

label define l\_deltype 0 vaginal 1 cesarean

label values deltype l\_deltype

recode momage 15 = .

label variable momage "Mother's Age"