

### Assignment 3

Due Tuesday, Oct. 10, 6 pm.

Question 1 log file should be posted on QuestromTools→Assignments

Question 1 data file should be uploaded to <https://tinyurl.com/qm222a1>

Answers to Question 2 along with a revised Project Description questions 1-5 (if changed) should be handed to Prof. Kahn during class or put under Prof. Kahn's door (518C) or in her mailbox (room 531)

This is an important assignment to do fully and correctly.

(Worth 1½ points if completed and handed in on time; late assignments lose 5% per day)

1. After you have your data in a Stata data set, prepare this data set so that you can use it for your analysis by doing the following:
  - a. Open a TEXT log file to capture your commands and the Stata output you generate in preparing your data set in steps b-f below. Put YOUR name and "assignment3" in its title (for instance: **log using jsmithassignment3, text**).
  - b. Clean your data set, ensuring all missing values of the numerical variables that you plan to use are a "." in Stata (if they weren't already). Makes the missing values of string variables ""
  - c. Generate any variables that you know you will need that are combinations of other variables.
  - d. Save this Stata dataset (for instance, **save dataset1**) on your own computer.
  - e. Then, if your dataset is so large that you have trouble using it, get rid of any *observations* that you know for certain you will never need, and any *variables* that you know **for certain** you will never need. However, it is much more difficult to retrieve erased variables than to keep extra variables. Save this as a new Stata dataset under a new name (e.g. **save dataset2**).
  - f. Name or rename the variables you will use so others understand what they represent.
  - g. Close your log file (**log close**).
  - h. Post this log file on QuestromTools→ Assignments→Assignment 3 log file as an attachment.
  - i. Rename your current dataset with your name (e.g. janedoe.dta). Then post it on the google drive folder: <https://tinyurl.com/qm222a1>
2. After you have done the above, hand in hard copy answers to the following questions:
  - a. How many observations in total are there in your data set? (Hint: use command **count**)
  - b. How many observations in total are there in your data set that have non-missing data for your main dependent variable? (use **count if .....**)
  - c. How many observations in total are there in your data set that have non-missing data for *both* your main dependent variable and the main explanatory variable(s) you will focus on? (At this point, this is the same as the Project Description's question 5 "In your data, how many usable -with all necessary variables- observations are there?")
3. Finally, revise the answers to Questions 1-5 in the **Project Description** document if any part has changed and hand it in (hard copy) as well.

Pro-Tip: If you put all of the commands you use to input your data and make your variables into a single file, you can easily change or add to it if you change your mind or want to new commands. Stata calls files with commands **do-files**.