# Jeff Davis

# POLS 418/GEOG 418: Quantitative Methods

Analysis 2

**Variable Types, Reliability and Validity, and Data Cleaning**

**Instructions:** Please complete each task listed below. You should type your problem set, and turn in a hard copy to me. You do not need to include a cover page. I expect, however, that your first page will include your full name, the problem set number, the due date, and the date you submit the work. The final page(s) of your assignment ***must*** include the STATA commands you used, otherwise it is incomplete, and it will receive the grade of “I”.

This assignment is designed to familiarize you with basic commands in the software program STATA as well as the basics of how to organize a dataset (i.e. spreadsheet). This assignment asks you to 1.) import an excel sheet into STATA 2.) rename variables 3.) change variable values 4.) create new variables 5.) drop variables 6.) reorder variables. You can think of these as the basic data management skills you will need to complete the future assignments.

***Part 1: Types of Variables***

Each of the following lists a variable, and in parentheses indicates measured values for two respondents. For each variable, please indicate the level of measurement ***and*** the unit of analysis.

1. Attitude toward capital punishment (first respondent = in favor; second respondent = oppose)
2. Days per week a person watches TV news (2 days; 5 days)
3. Self-rated health (good; fair)
4. Times per month one attends religious services (four; one)
5. Ethnicity (African American; Latino)
6. Political views (somewhat conservative; very conservative)

***Part 2: Reliability and Validity***

What does it mean for a variable to be both valid and reliable? Does a car speedometer that is always five mph too low have a problem with reliability or validity? Does a thermometer that varies +/- 2 degrees but on average gives you the right temperature have a problem with reliability or validity?

***Part 3: Data Importing, Cleaning, and Management***

. import excel "/Users/jeffersondavis/Desktop/Political Science 434 Class Survey.xlsx", sheet("Sheet") first

> row

(35 vars, 385 obs)

. rename (Whatisyourgender) (Gender)

. rename (Whichraceethnicitybestdescri) (Race)

. rename (L) (Other)

. rename (Whatisyourage) (Age)

. rename (Whatisthehighestlevelofsch) (Education)

. rename (Haveyoueverservedinanybran) (Military\_Service)

. rename (Howoftendoyouwatchprofessio) (Sports\_Watching)

. rename (Generallyspeakingdoyouthink) (Political\_Affiliation)

. rename (Generallyspeakinghowwouldde) (Political\_Ideology)

. rename (Howdoyoufeelaboutprofession) (Sports\_Politics\_Opinion)

. rename (Doyoufeelthatplayerskneelin) (Kneeling\_Opinion)

. rename (Haveyoueverparticipatedina) (Protest\_Participation)

. rename (Doyouagreewiththefollowing) (Blacks\_Would\_Be)

. rename (W) (White\_People\_Have)

. rename (Pleaselisttheprimarymediaso) (Media\_Sources)

. replace Age = "." if Age=="I don't identify with a par

> ticular age"

(1 real change made)

. replace Age = "." if Age=="Time is a flat circle"

(1 real change made)

. replace Age = "." if Age=="Q"

(1 real change made)

. replace Age = "21" if Age=="Twenty one"

(1 real change made)

. destring Age, replace

Age: all characters numeric; replaced as byte

(4 missing values generated)

. drop RespondentID CollectorID StartDate EndDate IPAddr

> ess EmailAddress FirstName LastName CustomData1

. generate PID\_ID = Political\_Affiliation+Political\_Ideo

> logy

(1 missing value generated)

. order Political\_Affiliation Political\_Ideology PID\_ID