# Jeff Davis

# POLS 418: Quantitative Methods

Problem Set 2

***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)

Level of measurement: nominal

Unit of Analysis: individual

1. Days per week a person watches TV news (2 days; 5 days)

Level of measurement: interval/ratio

Unit of Analysis: individual

1. Self-rated health (good; fair)

Level of measurement: ordinal

Unit of Analysis: individual

1. Times per month one attends religious services (four; one)

Level of measurement: interval/ratio

Unit of Analysis: individual

1. Ethnicity (African American; Latino)

Level of measurement: nominal

Unit of Analysis: individual

1. Political views (somewhat conservative; very conservative)

Level of measurement: ordinal

Unit of Analysis: individual

***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?*

**Introduction:**

One of the objectives of political science research is to describe abstract ideas and concepts and the relationships between them.[[1]](#footnote-1),[[2]](#footnote-2) This objective has several inherent challenges because “you cannot see” a concept.[[3]](#footnote-3) However, researchers can determine a concept’s “measurable properties” and develop a “strategy” for measuring those properties “in the real world.”[[4]](#footnote-4) Those measurable properties are formally identified as variables. And it is crucial that the instrument used to measure a variable is reliable and valid. The purpose of this essay is to explain what it means for an instrument to be both reliable and valid. Ultimately, the terms reliability and validity pertain to an instrument’s consistency and accuracy, respectively.

**Validity and Reliability**

Before turning to the concepts of reliability and validity, it is useful to briefly describe the process of moving from a concept to measuring its properties.[[5]](#footnote-5) To use an example, socioeconomic status is a concept defined as the “social standing or class of an individual or group.”[[6]](#footnote-6) The definition itself cannot be measured. But it is possible to identify “measurable properties” or “characteristics” of socioeconomic status.[[7]](#footnote-7) In this case, it is widely accepted that the concept’s measurable properties are an individual’s “income, occupation, and education.”[[8]](#footnote-8) In this example, income, occupation, and education are “measure[s]” of the original concept—socioeconomic status.[[9]](#footnote-9) Another way to think of these measures is to refer to them as variables, which can be measured by an instrument such as a survey question to determine the values for each variable.

In the process of identifying, defining, and measuring concepts, it is unlikely that a researcher’s decisions will be universally accepted.[[10]](#footnote-10) For this reason, social science strives to safeguard the quality of research by using standards such as reliability and validity to inform the decision-making process.[[11]](#footnote-11) In short, reliability is about consistency and validity is about accuracy.

As a research standard, reliability is concerned with how consistent an instrument produces similar results over time.[[12]](#footnote-12) For example, a vehicle’s speed is a variable, which can be measured with a speedometer—the instrument. If the same pressure on a gas pedal produces different readings “after repeated use,” the speedometer has a problem with reliability because it is not consistent.[[13]](#footnote-13) However, a reliable instrument may not be valid. If, for instance, the speedometer is *always* 5 mph too low it is still considered reliable because the reported speed is consistent, but the instrument has a problem with validity.

Validity is concerned with how well an instrument captures the “intended characteristic” of the original concept that a researcher is trying to “measure.”[[14]](#footnote-14) Unlike reliability, validity is about accuracy not consistency per se.[[15]](#footnote-15) Although by some standards if an instrument is valid it is also reliable. This distinction aside, a thermometer that varies +/-2 degrees but on average reports the right temperature, has a problem with reliability but not necessarily validity. Although the thermometer roughly captures an accurate reading, it produces different results under similar conditions, and thus is not reliable. The thermometer is, however, a valid instrument due to its relative accuracy.

In the social sciences, meeting the standards of validity is particularly challenging due to the inherent obstacles in defining a concept, determining its measurable properties, and selecting research instruments. The process of determining the best way to measure concepts like “power” or “democracy” is considerably more complex than measuring concepts like speed or temperature.[[16]](#footnote-16)

Philip Pollock illustrates how the standard of validity is higher than the standard of reliability in the following example: A “researcher” wants to measure the extent that people support government spending on social programs by measuring a person’s weight. Aside from “random fluctuations in weight,” the scale used to measure the variable “weight” can be considered reliable. But a person’s weight bears no plausible relationship with support for government spending.[[17]](#footnote-17) Research on political tolerance in the 1950s provides another example of validity problems in the social sciences. In one study, tolerance was defined as the willingness to tolerate unpopular groups. The researchers designed a survey question that only included groups that were left leaning and thus unintentionally measured tolerance “toward leftist groups.”[[18]](#footnote-18) In this example, the survey instrument was invalid because it failed to “link…the concept of tolerance and the empirical measurement of the concept.”[[19]](#footnote-19)

**Conclusion:**

Reliability and validity are standards used to improve the overall quality of research and scientific knowledge. Ensuring that an instrument produces consistent readings is the hallmark of reliability, but it does not guarantee its validity. To meet the standard of validity an instrument must measure, at a minimum, “the intended characteristic” of a concept.[[20]](#footnote-20) Thus, an instrument is both reliable and valid if it is consistent and accurate.

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

1. The first step is to import the file “Political Science 434 Class Survey.xlsx” into STATA. You can do this by selecting ‘File>Import>Excel spreadsheet’ from the dropdown menu. Once you have done this you can specify the file you want to import. Here, select your Excel file. Make sure you check the “import first row as variable names” otherwise STATA will not know the first row is variable names and will instead treat them as values.

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

> row

1. ars, 385 obs)
2. As you will notice the variables all have very long names. This makes it difficult to determine they mean. We will fix this by renaming the variables. You can do this by selecting ‘Data>Data utilities>Rename groups of variables’ from the dropdown menu. I want you to rename the series of variables starting with “Whatisyourgender” through “Pleaselisttheprimarymediaso” You can tell what each variable means by looking at the variable label. You can rename one variable at a time or rename them all at once.

. 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)

1. Open up the data browser. One thing to pay attention to is if a variable’s values are red it means that STATA recognizes this variable as text opposed to a number. This is okay if the variable is supposed to be text (like the name of a state) but this is the problem if the variable is supposed to be a number. All of our variables are okay except for one—the variable originally titled “Whatisyourage”. If you scroll down, you will notice that there are a couple values that are non-numeric. One person responded, “I don’t identify with a particular age” while another responded, “Time is a flat circle” and yet another responded “Q”. These are obviously not serious answers. Additionally, a fourth person typed out “Twenty One”. We will handle these cases by replacing these values as missing. You can do this by selecting ‘Data>Create or change data>Change contents of a variable’ from the dropdown menu. Here, select the variable name for the variable whose values you wish to edit. From there, you need to tell STATA exactly what you want to replace. In the new contents line then type - "." if Yourvariablename=="Time is a flat circle" (Note: you will have to replace Yourvariablename with whatever you named the age variable). Here we are replacing the value “Time is a flat circle” with a period symbol, which is STATA’s way of indicating a missing value. The quotation marks around the period symbol and the ‘Time is a flat circle’ initial response tells STATA these values are text opposed to numbers. (Note: If you are replacing values for a numeric variable you don’t need the quotation marks). Repeat this process for the “I don’t identify with a particular age” and “Q” responses. Finally, repeat the process for the person who typed out “Twenty one” only here I want you to replace the value equal to “21” opposed to “.”. Once you have replaced this final value you will have replaced all the non-numeric values.

. 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)

4.) Now it is time to tell STATA that your ‘age’ variable is a number and not text. Select ‘Data>Create or change data>Other variable-transformation commands>Convert variables from string to numeric’ Here, type in the name of the variable you want to transform and select the “Convert specified variables to numeric” option. Then click OK. Now check the data browser. The values for your ‘age’ variable should be in black now, opposed to red. This means STATA is recognizing the ‘age’ variable as a number opposed to text. If it is still in red, something has gone wrong. Make sure you replaced all of the non-numeric values.

. destring Age, replace

Age: all characters numeric; replaced as byte

(4 missing values generated)

5.) Open the data browser. You will notice that there are a series of columns in the beginning of the dataset (EmailAddress through CustomData1) that are all missing values. I want you to get rid of these columns. In the command line type ‘drop’ and then the list of variables you want to get rid of. Make sure you do not include any variables besides these four.

. drop RespondentID CollectorID StartDate EndDate IPAddr

> ess EmailAddress FirstName LastName CustomData1

6.) Now we are going to generate a new variable. We are going to create a combined partisanship and ideology variable. This new variable will be a combined sum of a person’s self-reported ideology (1-7, with seven being very conservative) and partisanship (1-7, with seven being strong Republican). To do this, selection Data>Create or change data>Create new variable from the dropdown menu. Choose a name for this new variable and have it equal partisanship+ideology (note: you will have to use the variable names you chose for partisanship and ideology here). This new variable should range from 2 to 14.

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

> logy

(1 missing value generated)

7.) Let’s examine how these three variables co-vary. To do this, it helps to look at all three variables side by side in the data browser. To accomplish this, type ‘order’ into the command window followed by the names of these three variables (with spaces in between). Then hit enter. Now look at the data browser. These three variables should now appear as the first three columns.

. order Political\_Affiliation Political\_Ideology PID\_ID

8.) Now save your dataset. Select ‘File>Save’ from the dropdown menu. Give your file an name and then save it as a ‘.dta’ file (this is the extension for a STATA dataset).

Bibliography

Howard, Christopher. *Thinking like a political scientist: a practical guide to research methods*. University of Chicago Press, 2017.

Pollock III, Philip H., and Barry C. Edwards. *The essentials of political analysis*. Cq Press, 2019.

“Socioeconomic Status.” *American Psychological Association*. <https://www.apa.org/topics/socioeconomic-status/>

Appendix A

**Figure 1. Code for Problem Set 1**

. 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

1. Howard, Christopher. *Thinking like a political scientist: a practical guide to research methods*. University of Chicago Press, 2017: 38 [↑](#footnote-ref-1)
2. Pollock III, Philip H., and Barry C. Edwards. *The essentials of political analysis*. Cq Press, 2019: 12 [↑](#footnote-ref-2)
3. Ibid, 11 [↑](#footnote-ref-3)
4. Ibid, 13 [↑](#footnote-ref-4)
5. Howard: 38 [↑](#footnote-ref-5)
6. “Socioeconomic Status.” *American Psychological Association*. <https://www.apa.org/topics/socioeconomic-status/> [↑](#footnote-ref-6)
7. Pollock: 13 [↑](#footnote-ref-7)
8. Pollock: 18 [↑](#footnote-ref-8)
9. Pollock III, Philip H., and Barry C. Edwards. *The essentials of political analysis*: 18, 54 [↑](#footnote-ref-9)
10. Howard: 6 [↑](#footnote-ref-10)
11. Ibid, 29, 38 [↑](#footnote-ref-11)
12. Howard, Christopher. *Thinking like a political scientist*: 49 [↑](#footnote-ref-12)
13. Ibid [↑](#footnote-ref-13)
14. Pollock: 25, 26, 31 [↑](#footnote-ref-14)
15. Pollock III, Philip H., and Barry C. Edwards. *The essentials of political analysis*: 30 [↑](#footnote-ref-15)
16. Pollock: 18 and Howard: 39 [↑](#footnote-ref-16)
17. Pollock III, Philip H., and Barry C. Edwards. *The essentials of political analysis:* 29 [↑](#footnote-ref-17)
18. Ibid, 27 [↑](#footnote-ref-18)
19. Ibid, 28 [↑](#footnote-ref-19)
20. Pollock: 25, 45 [↑](#footnote-ref-20)