Bailey Williams
Problem Set 2
Due Date: July 10, 2025
Word Count: 811
Submitted: July 10, 2025

Problem Set 2: Variable Types, Reliability and Validity, and Data Cleaning

In this problem set, different variables are described, including identifying their levels of measurement and units of analysis. Reliability and validity are also defined, and examples of instances of high validity and low reliability and high reliability and low validity are discussed. Finally, STATA code is included in the Appendix, which includes importing an Excel file, renaming variables, replacing variable values, creating new variables, dropping variables, and recording variables before exporting the file.

Part 1: Types of Variables

a) Attitude toward capital punishment (first respondent = strongly in favor; second respondent = somewhat opposed)

    The variable measuring the attitude towards capital punishment, where respondent responses include strongly in favor and somewhat opposed, is an ordinal level variable that uses individuals as the unit of analysis.

b) Poverty rate (Virginia 12.7%; West Virginia; 21.2%)

    The variable measuring poverty rates as percentages is a ratio scale variable, where 0% is the complete absence of poverty and 100% signifies total poverty, and uses U.S. states as the unit of analysis.

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

    The variable measuring the number of days a week a person watches TV news is a ratio scale variable, as it measures the exact number of days per week with a meaningful zero point of zero days watching TV. This variable uses individuals as the unit of analysis.

d) Religious identification (Protestant; Greek Orthodox)

    The variable measuring religious identification is a nominal level variable because it communicates differences between religious groups and are unordered categories. This variable uses individuals as the unit of analysis.

e) Self-rated health (good; fair)

    The variable measuring self-rated health is an ordinal level variable, where an individual's self-rated health is classified into different values. This variable uses individuals as the unit of analysis.

f) Times per month one attends religious services (four; one)

The variable measuring the times per month one attends religious services is a ratio scale variable, as it measures the exact number of times per month with a meaningful zero point of zero days per month. This variable uses individuals as the unit of analysis.

g) Ethnicity (African American; Latino)

The variable measuring ethnicity is a nominal level variable because it communicates differences between ethnic groups and are unordered categories. This variable uses individuals as the unit of analysis.

h) Resting heart rate (77 BPM; 61 BPM)

The variable measuring resting heart rates is a ratio scale variable, as it measures the exact number beats per minute, with a meaningful zero point of zero beats per minute. This variable uses individuals as the unit of analysis.

i) Political views (somewhat conservative; very conservative)

The variable measuring political views is an ordinal level variable, where an individual's political views is classified into different values. This variable uses individuals as the unit of analysis.

j) Votes cast per precinct (522; 1456)

The variable measuring votes cast per precinct is a ratio scale variable, as it measures the exact number of votes cast, with a meaningful zero of zero votes cast per precinct, and precincts as the unit of analysis.

Part 2: Reliability and Validity

For a variable to be valid, it must record an accurate measurement of the intended characteristic and not measure other, unintended characteristics (Pollock III & Edwards, 2020, p. 17). Reliability, on the other hand, is based on how consistently the variable measures the intended characteristic (p. 16). An example of a valid but unreliable variable could be one measuring income based off a poorly written survey question. While it does measure income, making it valid, it is unreliable due to potential misinterpretation by participants. A car speedometer that is always five mph too low has a problem with validity,

as it records the speed consistently but inaccurately. A thermometer that varies +/- 2 degrees but on average gives you the right temperature has a problem with reliability, as it records an accurate measurement of the temperature, on average, but inconsistently.

Part 3: Data Importing, Cleaning, and Management

See Appendix for STATA code used.

Bailey Williams
Problem Set 2
Due Date: July 10, 2025
Word Count: 811
Submitted: July 10, 2025

References

Pollock III, P. H., & Edwards, B. C. (2020). *The Essentials of Political Analysis* (Sixth Edition). CQ

Press, A division of SAGE.

Appendix

```
import excel "/Users/baileywilliams/Downloads/Political Science 434 Class Survey.xlsx",
sheet("Sheet") firstrow

rename (Whatisyourgender ) (Gender)

rename (Whichraceethnicitybestdescri  Whatisyourage Whatisthehighestlevelofsch
Haveyoueverservedinanybran) (RaceEthnicity Age HighestLevelOfSchool MilitaryService)

rename (Howoftendoyouwatchprofessio Generallyspeakingdoyouthink
Generallyspeakinghowwouldde Howdoyoufeelaboutprofession ) (WatchProfessionalSports
PoliticalParty PoliticalIdeology PoliticsInProfessionalSports )

rename (Doyoufeelthatplayerskneelin Haveyoueverparticipatedina)
(PlayersKneelingDuringAnthem ParticipationInProtests)

rename (Doyouagreewiththefollowing W Pleaselisttheprimarymediaso)
(BlacksBetterOffWorkHarder WhitePeopleMoreThanDeserved PrimaryMediaSources)

rename (L) (RaceEthnicityOther)

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

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

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

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

destring Age, replace

drop EmailAddress FirstName LastName CustomData1

generate PartisanshipIdeologySum = PoliticalParty + PoliticalIdeology

order PartisanshipIdeologySum PoliticalParty PoliticalIdeology

save "/Users/baileywilliams/Desktop/ProblemSet2.dta"
```