Problem Set 1
Quantitative Political Methodology (U25 363)
Due: February 6, 2018

Instructions

- Please show your work if possible. You may lose points by simply writing in the answer. If the problem requires you to execute commands in R, please include the code you used to get your answers. Please also include the .R file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask me during class, email me, or come to office hours.
- Your homework should be submitted electronically on the course Github page.
- This problem set is before the beginning of class on Wednesday, February 6, 2019.
- No late assignments will be accepted.
- Total available points for this homework is 90.

Question 1 (8 points)

Go to the GSS Web site, sda.berkeley.edu/GSS/ and click on "GSS – with 'No Weight' as the default (SDA 4.0)". By entering LDCGAP, find a summary of responses to the question, "Turning to international differences, do you agree or disagree: Present economic differences between rich and poor countries are too large."

a) (2 points) What is the mean response?
2.2
b) (2 points) What was the most common response?
Agree
c) (4 points) Is you answer in (b) a descriptive statistic or an inferential statistic? Explain. **Descriptive. This is because this statistics provides a numerical description of the data, and not an inference about a population, which would make it inferential.**

Question 2 (8 points)

In the 2016 presidential election, an exit poll sampled 1,941 of the 2,808,605 people who voted in Missouri. The poll stated that 57.23% of respondents reported voting for the Republican candidate, Donald Trump. Of all 2,808,605 voters, 56.8% voted for Trump.

a) (2 points) For this exit poll, what was the population? **2,808,605 people who voted in Missouri**
b) (2 points) For this exit poll, what was the sample? **1,941 sampled**
c) (2 points) Identify a statistic**. Of all 2,808,605 voters, 56.8% voted for Trump.**
d) (2 points) Identify a parameter. **57.23% of respondents reported voting for the Republican candidate, Donald Trump**

Question 3 (24 points)

Which scale of measurement (nominal, ordinal, or interval) is most appropriate for:

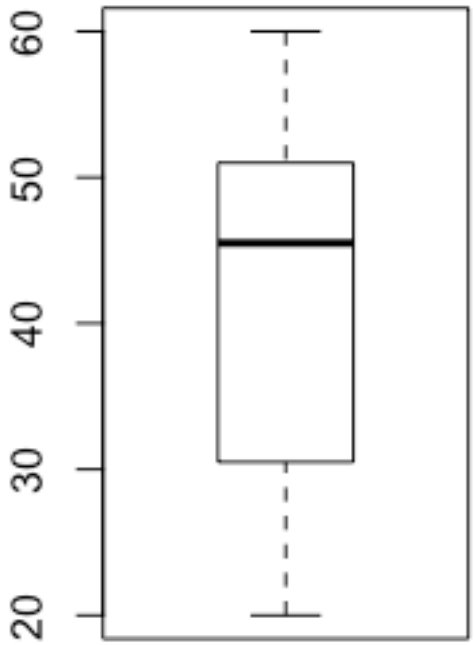
a) (2 points) Educational attainment (less than high school, some high school, high school, some college, college degree, graduate or professional degree)
NOMINAL

b) (2 points) Race (White, Black or African American, Asian, . . .)
NOMINAL
c) (2 points) Letter grades
ORDINAL
d) (2 points) Statewide murder rate (number of murders per 1000 population)
INTERVAL
e) (2 points) Distance (in miles) commuted to work
INTERVAL
f) (2 points) Hair color (Blond, Brunette, Red, Black)
NOMINAL
g) (2 points) Number of people you have known who volunteered for the Obama campaign
INTERVAL
h) (2 points) Partisan affiliation (Republican, Democrat, Green,....)
NOMINAL
i) (2 points) Zip code
NOMINAL
j) (2 points) Religious affiliation (Catholic, Protestant, Jewish, Muslim, Buddhist, other)
NOMINAL
k) (2 points) Government spending on environment (up, same, down)
INTERVAL
l) (2 points) GPA (4.00, 3.00, 2.00, etc)
ORDINAL

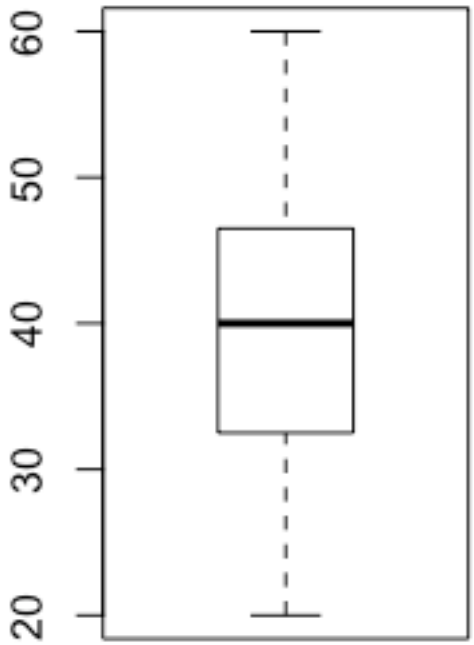
Question 4 (14 points)
Suppose you are comparing the age of 20 men and 20 women. For the men, you get the following responses (in years): 56, 60, 50, 26, 45, 35, 41, 43, 34, 42, 37, 39, 33, 28, 52, 48, 27, 20, 44, 32.

For the women, you get the following responses (in years): 47, 49, 20, 46, 43, 44, 45, 60, 57, 28, 55, 27, 25, 50, 52, 48, 23, 42, 33, 59.

a) (12 points) In R, make a box plot for both men and women. Make sure the mean and interquartile range are clearly marked for each box plot. Make sure to save the plot as a .pdf, and include the code and figure in your response.
Male
25%: 13.25
75%: 65.25
Mean: 36.9



Female
25%: 3.625
75%: 78.625
Mean: 42.65



Code: men <- c(56, 60, 50, 26, 45, 35, 41, 43, 34, 42, 37, 39, 33, 28, 52, 48, 27, 20, 44, 32)
women <- c(47, 49, 20, 46, 43, 44, 45, 60, 57, 28, 55, 27, 25, 50, 52, 48, 23, 42, 33, 59)
boxplot(men)
boxplot(women)

print("male data")
quantile(men,.25) - 1.5*IQR(men)
quantile(men,.75) + 1.5*IQR(men)

print("female data")
quantile(women,.25) - 1.5*IQR(women)
quantile(women,.75) + 1.5*IQR(women)

mean(men)

mean(women)

b) (2 points) Is this likely to be a random sample of the American female population? Why or why not?
    a. This is likely a random sample because the ages recorded span a great range, with a majority of the ages in the typical adult female age range.

Question 5 (24 points)
Identify each variable as discrete or continuous. If you think that more than one answer might be correct, justify your answer.

a) (2 points) Attitudes toward legalization of marijuana (favor, neutral, oppose) **Discrete**
b) (2 points) Number of political parties in a country **Discrete**
c) (2 points) Religious affiliation (Catholic, Jewish, Protestant, Muslim, ...) **Discrete**
d) (2 points) Choice of candidate a person will vote for **Discrete**
e) (2 points) Distance (in miles) commute to work **Continuous**
f) (2 points) Years of school completed (0, 1, 2, ...) **Continuous**
g) (2 points) Number of people you have known who volunteered for the Clinton campaign **Continuous**
h) (2 points) Partisan affiliation (Republican, Democrat, Green,....) **Discrete**
i) (2 points) Attitude towards the health care reform (favor, oppose, neutral) **Discrete**
j) (2 points) Political ideology (very liberal, somewhat liberal, moderate, somewhere moderate, somewhat conservative, conservative) **Either**
k) (2 points) Government spending on environment (up, same, down) **Discrete**
l) (2 points) GPA (4.00, 3.00, 2.00, etc.) **Continuous**

Question 6 (EXTRA CREDIT! 10 points)
Assume that the probability of an American high school graduate participating in sports is 0.45. Further assume that of those who participate in sports, about 34% play an instrument. Again, assume that the probability of a high school graduate attending college is 0.69 (which is the case for 2015 high school graduates according to Bureau of Labor Statistics). Estimate the following probabilities (show your work):

a) (2 point) that a randomly chosen American high school graduate does not participate in sports.
b) (4 points) that a randomly chosen American high school graduate both participates in sports and plays an instrument.
c) (4 points) that a randomly chosen American high school graduate participate in sports and attends to college.

Question 7 (12 points)
Find an article that uses a statistic in the news and attach it to your homework. Please comment on the following: What kind of underlying data is used to calculate the statistic?

Is it discrete or continuous? Is it nominal, ordinal, or interval? Is the statistic inferential or descriptive?

https://www.nytimes.com/2015/08/28/health/school-lunches-becoming-healthier-statistics-indicate.html

1. The underlying data is from percentage of schools that offer healthier options for their school lunches. The statistics range from sodium intake to how many vegetables are offered in lunches
2. Continuous
3. Interval
4. This statistic is descriptive because a main statistic stems from the stat about the percentage of meals