Problem Set 2

Quantitative Political Methodology (U25 363)

Due: February 27, 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 have plots, attach them as well within your written document. Make sure you label clearly which question the codes correspond to. If you are not sure if work needs to be shown for a particular problem, please ask me.

Your homework should be submitted electronically on the course GitHub page.

This problem set is due before the beginning of class on Wednesday February 27, 2019. No late assignments will be accepted.

Total available points for this homework is 100.

# Question 1 (5 points)

You would like to find the proportion of bills passed by Congress that were vetoed by the President in the last congressional session. After checking congressional records, you see that for the population of all 40 bills passed, 2 were vetoed. Does it make sense to construct a confidence interval using these data to answer your question? Explain.

No it does not make sense to construct a confidence interval because since you have all the available information (whole population) there is no need to construct a confidence interval because you can be 100% certain of the proportion of bills passed by Congress but vetoed by the President in the last congressional session.

# Question 2 (25 points)

The distribution of family size in a particular tribal society is skewed to the right, with *µ* = 5*.*2 and *σ* = 3. Those values are unknown to an anthropologist, who samples families to estimate mean family size. For a random sample of 36 families, she gets a mean of 4.6 and a standard deviation of 3.2.

1. Identify the population distribution. State its mean and standard deviation. Is the data skewed? Population mean = 5.2, population stdev. = 3, yes data is skewed to the right
2. Identify the sample data distribution. State its mean and standard deviation. Is the data skewed? Sample mean = 4.6, sample stdev = 3.2, skewed to the right
3. Identify the sampling distribution of ¯*y*. State its mean and standard error and explain what it describes. Standard error = σ/n^1/2,[Y-bar*=*4.6 and standard error of y - bar= 3/(36)^1/2 =3/6 = .5], this tells us the spread of the sample data using statistics.
4. Find the probability that her sample mean falls within 0.5 of the population mean. Standard error = .5, Z=(u+.5)-u/.5, z=(1,-1), z table of 1= .8413, .1587= .8413-.1587= **.6826**.
5. Suppose she takes a random sample of size 100. Find the probability that the sample mean falls within 0.5 of the true mean, and compare the answer to that in (d). standard error = 3/(100)^1/2 = 3/10=.3, z= (u+.5)-u/.3= .5/.3=1.67,-1.67 = .9525, .0475= .9525-.0475= **.905**
6. Refer to (e). If the sample were truly random, would you be surprised if the anthropologist obtained ¯*y* = 4. Why? Yes because within both samples of n=36 and n=100 it was likely that the mean fell between 4.7 and 5.7 therefore I believe that given a truly random sample it would surprise me to see a mean as low 4.

# Question 3 (10 points)

The GSS asks respondents to rate their political views on a seven-point scale, where 1= extremely liberal, 4=moderate, and 7=extremely conservative. A researcher analyzing data from 2011 has the following data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | *N* | Mean | St. Dev | SE Mean |
| Polviews | 1294 | 4.23 | 1.39 | 0.0387 |

1. Show how to construct a 95% confidence interval from the information provided.
   1. X-bar=4.23, s=1.39, standard error= .0387, 4.23 +/- (1.96x.0387)= (4.15,4.31)

Interpret the confidence interval you found in (a). We are 95% confident that the population mean of political ideology lies between 4.15 and 4.31

1. Would the confidence interval be wider or narrower (i) if you constructed a 90% confidence interval, (ii) if you found the 95% confidence interval only for those who called themselves *strong Democrats* on political party identification (PARTYID), for whom the mean was 3.50 with standard deviation 1.61? the confidence interval would be narrower if you constructed a 90% confidence interval and if you found the interval for those who called themselves strong democrats the interval would be narrower because you are taking from a smaller range of identifications as opposed from 1 to 7

# Question 4 (5 points)

For a normal distribution with *µ* = 50 and *σ*2 = 36, find the probability that an observation falls (Hint: type help(Normal) in R):. Z = (x-*µ)/ σ*

1. At or below the value 57.75(x </ 57.75)

pnorm(57.75, mean = 50, sd = 6) = .90147 = 90.147%

1. At or above the value of 50.45

1 - pnorm(50.45, mean = 50, sd = 6)= 0.4701074 = 47.01074%

1. Between the values of 52.4 and 59.4

pnorm(59.4, mean = 50, sd = 6) - pnorm(52.4, mean = 50, sd = 6) = .2859819= 28.59819%

# Question 5 (5 points)

R has a number of functions that make it simple to simulate from a variety of distributions.

One thing to note is that when sampling you want to set a seed in R. Setting the seed allows you to replicate your results. It doesn’t matter what it is set to. So, for the purposes of this question, type: set.seed(12345)

Suppose that salaries follow a normal distribution with mean 40000 and standard deviation 15000. We can sample from this distribution using the rnorm() command. Type the following into R to generate a sample with 10000 observations: salaries <- rnorm(n=10000,mean=40000,sd=15000)

Plot the distribution. Add a title to this plot. Save this plot as a .pdf file. Turned in separately

# Question 6 (10 points)

Plot probability density functions for the following normal distributions. Make all the plots on a single page. Make sure your plots have properly labeled titles and axes, and your axes are comparable across plots. Turned in separately

1. Normal Distribution with *µ* = 0 and *σ*2 = 0*.*4
2. Normal Distribution with *µ* = 0 and *σ*2 = 3
3. Normal Distribution with *µ* = 3 and *σ*2 = 3
4. Normal Distribution with *µ* = 3 and *σ*2 = 0*.*4
5. Normal Distribution with *µ* = −2 and *σ*2 = 5
6. Normal Distribution with *µ* = −2 and

# Question 7 (20 points)

Peake and Eshbaugh-Soha (2008) study drug policy coverage. Their data count the number of nightly television news stories in a month focusing on drugs, from January 1977 to December 1992. The dataset is in comma-separated format in the file named drugCoverage.csv. Download it from [Monogan (2015)’s Dataverse.](https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ARKOTI) The variables in the dataset are: a characterbased time index showing month and year (Year), news coverage of drugs (drugsmedia), an indicator for a speech on drugs that Ronald Reagan gave in September 1986 (rwr86), an indicator for a speech George H.W. Bush gave in September 1989 (ghwb89), the president’s approval rating (approval), and the unemployment rate (unemploy).

1. Draw a histogram of the monthly count of drug-related stories. Turned in separately
2. Draw two boxplots: One of drug-related stories and another of presidential approval. How do these figures differ and what does that tell you about the contrast between the variables? In the drug related stories boxplot there is a lot more variability in the data with many outliars but still a relatively low mean, the presidential approval boxplot has a lot less variability and with a relatively similar level of approval in respect to quartiles. This shows me that there is much more variability in the number of news stories about drugs and their responsiveness to speeches. While approval rating is much less responsive.
3. Draw two scatterplots:
   * In the first, represent the number of drug-related stories on the vertical axis, and place the unemployment rate on the horizontal axis.
   * In the second, represent the number of drug-related stories on the vertical axis, and place presidential approval on the horizontal axis.
   * How do the graphs differ? What do they tell you about the data? On the first graph the less drug related stories correlates to a higher unemployment rate and on the second graph the lower the number of drug related stories the higher the presidential approval rating. This tells me that the more the issue of drugs is addressed the lower the levels of unemployment but this also means that the presidents approval rating will be lower. Shows that two different variables may be correlated but this does not imply causation.
4. Draw two line graphs:
   * In the first, draw the number of drug-related stories by month over time.
   * In the second, draw presidential approval by month over time.
   * What can you learn from these graphs?separately, the max spike on both graphs are during the same five years which can suggest that the approval of the president may have shot up due to some address of drug related policy or situation.

# Question 8 (20 points)

For this question, you will work with W-NOMINATE data to trace the policy positions of members in the [U.S. House of Representatives.](https://drive.google.com/open?id=1kVn-Asz8KXPQhiLSQGEabdVJh9_8_wjf) With the data, you will learn about polarization (i.e. distance between the ideological positions of the Democratic Party and the Republican Party). You will also learn about the ideological cohesiveness of each party.

Answer the following questions:

1. Import data on the 88th and 107th Congresses. Then, create four subsets of the data by session and party (Democratic Party in the 88th session, Democratic Party in the 107th session, Republican Party in the 88th session, and Republican Party in the 107th session). In code file
2. For the Democratic Party, calculate the median W-NOMINATE scores for two Congresses. How did the median change over time? What does this mean? Median dem88= -.2365, median dem107 = -.4719, median has decreased over time, means that over time democrats have ideologically have become more polarized from republicans over time.
3. For the Republican Party, calculate the median W-NOMINATE scores for the two Congresses. How did the median change over time? What does this mean? Median rep88 = .4583, median rep107= .7179, median has increased over time, means that republicans have become more ideologically polarized away from democrats over time.
4. For the Democratic Party, calculate the standard deviation of W-NOMINATE scores for the two Congresses. How did the standard deviation change over time? What does this mean? Stdev dem88= .2894, stdev dem107= .2410, standard deviation reduced over time, means that over time there is less variance ideologically within the democratic party.
5. For the Republican Party, calculate the standard deviation of W-NOMINATE scores for the two Congresses. How did the standard deviation change over time? What does this mean? Stdev rep88= .1703, stdev rep107= . 1321, has reduced over time, means that within the republican party variance in ideology has reduced over time.
6. For the 88th Congress, create a plot that overlays two histograms. One histogram should plot the distribution of W-NOMINATE scores for the Democratic Party. The other histogram should plot the distribution of W-NOMINATE scores for the Republican Party. (Hint: to overlay two histograms, you can run two separate hist commands but include an add argument in the second hist one.) turned in separately
7. For the 107th Congress, create a plot that overlays two histograms. One histogram should plot the distribution of W-NOMINATE scores for the Democratic Party. The other histogram should plot the distribution of W-NOMINATE scores for the Republican Party.turned in separately
8. Based on what you have done so far, compare the 88th Congress and the 107th

Congress.

* + Did polarization decrease, increase, or stay the same? Are both parties responsible for this or is one party responsible? Polarization has increased and both parties are responsible.
  + For each party, what happened to the ideological cohesiveness of its members? Did it decrease, increase, or stay the same? For each party ideological cohesiveness increased in that democrats moved farther left and republicans farther right