Problem Set 3

Quantitative Political Methodology (L32 PS 363)

Instructions

1. Print out and write your WUSTL ID at the top of **each** page, and complete each exercise in the space alloted. You may attach extra pages if the space provided is not sufficient, but please indicate that you have done so below.

   Pages attached:_____

2. Please **show your work** if possible. You may lose points by simply writing in the answer. If the problem requires you to use `R`, please include the code you used to get your answers. If you are not sure if work needs to be shown for a particular problem, please ask a TA or post a question on Facebook.

3. The various pages of your homework should be **stapled together** (no paper clips please). If pages are lost because of a lack of a staple, no credit will be granted for that portion of the homework.

4. This problem set is **due at the beginning of class on Wednesday, October 4, 2017**. No late assignments will be accepted.

5. Total available points for this homework: 60.

Question 1 (Total: 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.

Question 2 (Total: 10 points)

The shape of the t-distribution varies by a parameter call "degrees of freedom," or df for short.

(a) (1 point) When df is large, the t-distribution approximates what other distribution?

(b) (5 points) Use `R` to plot the standard normal distribution as well as three t-distributions with $df = 20$, $df = 3$, and $df = 1$. Present all plots on *the same set of axes*, print, and attach to your submitted homework. Also attach the code used to produce the plot. Give your plot a meaningful title and label your axes. Use a different color, shade of gray, or line type for each line so a reader can clearly see the difference. (You may find `R`'s help files useful. For example, try `?plot`, `?lines`, `?dt`.)

(c) (4 points) Describe what your plot shows about the t-distribution. With reference to your plot, explain how different sample sizes might affect your estimates of population parameters.

Please find the data for this question by using the following code:

```
install.packages("Zelig")
library("Zelig")
data("voteincome")
?voteincome
```

Make sure to show all your work for parts (b) and (d) either with `R` code you attach or by hand in the space provided. If you complete (b) and (d) in `R`, make sure to clearly label the code pertaining to each part of the problem.

Question 3 (Total: 16 points)

You would like to test the hypothesis whether the average age among American voters is different from 50.

(a) (2 points) State the null hypothesis and the alternative hypothesis.

(b) (6 points) Calculate the standard error, the z test-statistic and the p-value for your test.

(c) (2 points) What is your conclusion at $\alpha = 0.05$?

(d) (4 points) Calculate the 95% confidence interval for the mean age.

(e) (2 points) How are your answers for parts (c) and (d) related?

Please show all work for this problem in the space provided.

Question 4 (Total: 24 points)

A librarian would like to learn how many books their patrons purchase per year. In particular he wants to test the $H_0 : \mu = 10$ against $H_a : \mu < 10$. He randomly selects 16 patrons and asks them how many books they purchase per year. He finds that $\bar{y} = 9.5$ and $s = 1.2$.

(a) (2 points) Given the sample size and the fact that you do not know the population standard deviation, which test statistic would you use?

(b) (2 points) What additional assumption do you need to use the test-statistic indicated in part (a)?

(c) (6 points) Calculate the test-statistic, and the p-value. What is your conclusion at significance level $\alpha = 0.05$?

(d) (2 points) Assume that you know the population standard deviation $\sigma = 1.2$. Can you use a test statistic different from the one indicated in part (a)? If so, what is that test-statistic called?

(e) (2 points) What assumption (if any) do you need to use the test-statistic indicated in part (d)?

(f) (6 points) Calculate the standard error, the test-statistic and the p-value. What is your conclusion at significance level $\alpha = 0.05$?

(g) (4 points) Compare your conclusions in parts (c) and (f). Explain the difference (if any) in the conclusions.

Question 5 (Total: 5 points)

A recent poll of 698 decided voters in Pennsylvania showed 341 preferred Donald Trump and 357 preferred Hillary Clinton. Let π be the population proportion of decided Pennsylvania voters who prefer Trump.

(a) (2 points) If the voters are only given two options (Trump or Clinton) and the sample size of your survey is relatively large, what type of distribution is the population distribution? What type of distribution is the sampling distribution of your survey?

(b) (1 point) What is the value of $\hat{\pi}$, the estimate of π obtained from the survey?

(c) (1 point) What is the standard error of this estimate?

(d) (1 point) Give the 95% confidence interval for the value of π.