Problem Set 3

POL 501

November 4, 2024

Instructions

Contents Covered

- *From the textbook*: Chapters 5 and 6.1 (Inference for Proportions).

- *R Lessons*: Previous Lessons.

- See the updated PDF file "*Glossary of Terms and Concepts*" for a review of key theoretical concepts.

General Instructions

- *Justify all of your answers.* Failing to do so will have score deductions.

- If you use formulas, always write them down and justify their appropriateness in each case.

- **Unless otherwise instructed**, you can use software to compute your answers (that is, to come up with the accurate calculation of a given arithmetic operation).

- Yet, you **must clearly describe how you arrived at each result by stating the appropriate formula**. Correct "numbers" in answers, but without justification or computed with the wrong formula, will have score deductions.

- Your code justifies your calculations but does not replace explanations and justifications.

Submission Instructions

- You must submit your answers to this problem set as a PDF or Word file **generated by R markdown**.

- You must use the R Markdown template I uploaded in Brightspace to write your answers.

- You must upload your PDF or Word File in the corresponding assignment in Brightspace.

- You must upload your RMD file (the code used to generate the document).

- **Due Date**: November 10th.

Question 1 (2.5 points)

Instructions: Using the **2024 National Public Opinion Reference Survey (NPORS)** dataset, answer the following questions. The dataset was collected by Pew Research Center between February 1, 2024, and June 10, 2024, using a sample of **5,626 respondents** through a multimode approach (online, paper, phone). Respondents were asked questions about political affiliation, internet usage, radio listening habits, perceptions of economic conditions, government spending, crime, and gun ownership. The data excludes non-valid responses (coded as **99** or **98**) across all variables used.

For this question, you will use the following variables:

- **PARTY**: Political affiliation.

- **CRIMESAFE**: Perception of community safety in terms of crime.

Please answer the following questions. Use R for your calculations. See the appendix for the variable's levels and labels (taken from the survey questionnaire). **Refer to the provided RMD template for specific code chunks to complete each part of Question 1.** Follow the steps and complete the missing parts in the R chunks indicated.

(1.a) Calculate the 95% Confidence Interval for the Mean of `CRIMESAFE`. Use the `confidence_interval` function to compute the mean of `CRIMESAFE` with a 95% confidence level. Interpret the confidence interval appropriately.

(1.b) Calculate the 99% Confidence Interval for the Mean of `CRIMESAFE`. Use the `confidence_interval` function with a 99% confidence level. Interpret the confidence interval appropriately.

(1.c) Filter Data by Political Affiliation. Use `dplyr` to create two separate data frames for respondents who identify as Democrats and Republicans.

(1.d) Compute 95% CI for `CRIMESAFE` for Each Group. Calculate the sample mean and 95% confidence intervals for `CRIMESAFE` for the Democrats and Republicans subsamples. Describe the results.

(1.e) Interpret and Compare the confidence intervals calculated in (1.d) for Democrats and Republicans. Visualize the intervals with a plot to interpret the results (the plot commands are already included in the chunk from the template). What might the confidence intervals suggest about the difference between the true means between the two groups?

R Chunk: {r q1-e}

Question 2 (1.5 points)

Instructions: Please answer the following questions. Use R for your calculations. Complete each sub-question and use the functions and methods indicated in the RMD template chunks. This question employs the data frame from Question 1.

(2.a) Create a Binary Variable for Perception of Safety, Compute the Sample Proportions, and Interpret:

In the `df_clean` dataset, create a new binary variable, `notsafe_binary`, that equals 1 if a respondent selected either "Not too safe" or "Not at all safe" for the `CRIMESAFE` variable, and 0 otherwise. Recreate the data frames for Democrats and Republicans to include this new variable. Compute the sample mean of `notsafe_binary` for each group and interpret these means as proportions. Explain what each proportion tells you about the perception of safety within each group.

R Chunk: {r q2-a}

(2.b) For the Democratic subsample, test if the true proportion of Democrats who feel unsafe (as indicated by `notsafe_binary`) is different from the sample proportion observed for Republicans. Test the null hypothesis of no difference at the 5% and 1% significance levels. What can you conclude from this analysis?

R Chunk: {r q2-b}

(2.c) For the Republican subsample, test if the true proportion of Republicans who feel unsafe (as indicated by `notsafe_binary`) is different from the sample proportion observed for Democrats. Test the null hypothesis of no difference at the 5% and 1% significance levels. What can you conclude from this analysis?

R Chunk: {r q2-c}

Question 3 (2.5 points)

Instructions: In this question, you will conduct a hypothesis test for a population proportion using a mock dataset. Hypothesis testing allows us to assess whether a sample proportion is significantly different from a hypothesized population proportion. Here, you will test if the proportion of individuals who support a specific policy differs significantly from a hypothesized population proportion of 0.43. This will be a two-tailed test conducted at a 1% significance level, meaning we are testing if the sample proportion is either significantly higher or lower than 0.43.

To complete this question, we will generate a mock dataset representing 35 survey respondents, each either supporting (1) or not supporting (0) a specific policy. The data is generated with a true underlying proportion of 0.44, and a seed is set to ensure reproducibility (see **R Chunk: {r q3-setup}** for details).

(3.a) Calculate the sample proportion of respondents supporting the policy and the standard error of this proportion. Describe and interpret your results.

(3.b) Define the null hypothesis and alternative hypothesis for this test.

(3.c) Compute the Test Statistic. Calculate the z-test statistic based on your sample proportion, hypothesized population proportion, and standard error computed assuming the null hypothesis. Describe and interpret your results.

(3.d) Determine the critical value for a two-tailed test at the 1% significance level and calculate the p-value for the test statistic. Describe and interpret your results.

(3.e) Draw a Conclusion. Based on the test statistic, critical value, and p-value, determine if the null hypothesis can be rejected. Describe and interpret your results.