

Mini-Group Project 1

Project Overview

This project is designed to enhance your understanding of statistical concepts, data analysis, and critical thinking in the context of presidential approval ratings and voter demographics. You will work in teams to complete three main tasks, each focusing on different aspects of data science. After each task, one team member will prepare a 5-minute video summarizing the findings, submitted individually as part of their solo project-2.

General Instructions

- Work collaboratively to complete each task. Ensure equitable participation among all team members.
- Submit a **single PDF document** containing the team's responses. The written pdf report should only include your answers. Do not include the questions. Clearly label each section corresponding to the tasks and parts. All explanations must be in your own words. Plagiarism is strictly prohibited.
- Submit a **single Excel document** showing the work being performed.
- After each task, a different team member will create a 5-minute video summarizing the task's findings. Each assigned team member must submit their video individually as part of their solo project requirements. If a team has more than three members, multiple team members can work on the same task, but each must submit their own original work.

Dataset Provided

In this project, we will work with two datasets. For Tasks 1 and 2, we will use the Gallup Presidential Approval data for each President starting with Franklin D. Roosevelt. You can access the data from the following link:

[Presidential Job Approval--All Data | The American Presidency Project \(ucsb.edu\)](https://www.americanpresidencyproject.org/president-job-approval/).

The data in this project is compiled based on responses to the survey question, "Do you approve or disapprove of the way [enter President name] is handling his job as President?" The Excel file contains public approval and disapproval ratings, as well as the proportion of respondents who answered "Don't know" or had no opinion. We will focus on a subset of data covering the last four presidents.

The Gallup Job Approval data for the current President, Biden, is stored in the worksheet "**Biden**" and covers the period from January 21, 2021, to September 15, 2021. The data for President Trump is stored in the worksheet "**Trump**" and covers the period from January 20, 2017, to January 15, 2021. The data for President Obama is stored in the worksheet "**Obama**" and covers the period from January 19, 2009, to January 19, 2017. Finally, the data for President George W. Bush is stored in the worksheet "**GWBush**" and covers the period from February 1, 2001, to January 11, 2009.

Mini-Group Project 1

It's worth noting that the Presidential Job Approval rating data does not have the same frequency of dates for each presidential term. The data for President Biden is available on a monthly basis, while we have weekly data for the others. However, for the purposes of this project, we will ignore this fact and analyze the data as is for Tasks 1 and 2.

For Task 3, you will use a revised version of PEW September 2020 Survey data under "PEW" worksheet.

Task 1: Confidence Intervals and Interpretation

1. In each worksheet dedicated to a President, populate the columns named "Approving_adj_[PresidentName]" and "Disapproving_adj_[PresidentName]". Allocate the value from the "Unsure/NoData" column between the "Approving" and "Disapproving" columns in a proportional manner. Ensure that the sum of "Approving_adj_[PresidentName]" and "Disapproving_adj_[PresidentName]" totals 100. For instance, if we have Approving 34, Disapproving 61, and Unsure/NoData as 5 for President GW Bush, the "Approving_adj_Bush" cell will be 35.79 and "Disapproving_adj_Bush" will be 64.21, totaling 100. Determine which president has the highest average "Approving_adj_[PresidentName]" value and which one has the lowest, and include your findings in your submission.
 2. Explain in your own words what a 95% confidence interval is, without using statistical jargon. Describe how you would explain this concept to someone with no statistical background.
 3. Create a time-series plot showing the "Approving_adj_[PresidentName]" rate for each president (Biden, Trump, Obama, George W. Bush) in a single graph. Ensure all data is plotted on the same graph, starting from the beginning of each president's term. Include appropriate labels, legends, and a clear title to distinguish the approval rates of each president. Include the plot in your report and provide at least one descriptive observation about it.
 4. Calculate the 95% confidence interval for the "Approving_Adj_[PresidentName]" rate for each president and enter your results. Ensure that you justify any assumptions made during the calculation process.
 5. Identify the president with the narrowest and the widest confidence intervals. Analyze the factors that contribute to the width of these intervals, such as sample size and variability. Explain how these factors influence the precision of your estimates.
-

Mini-Group Project 1

Task 2: Hypothesis Testing and Critical Analysis

Part A: President Obama

1. Formulate a hypothesis to test if the proportion of people who approve of President Obama was higher in the first 1,000 days compared to the last 1,000 days of his presidency. Use Approving_adj_Obama column for the testing. You can use the following date ranges: First 1000 Days: January 19, 2009 – October 16, 2011. Last 1000 Days: April 27, 2014 – January 19, 2017. State the null hypothesis (H_0) and alternative hypothesis (H_1) for testing the difference in proportions.
2. Discuss the assumptions required for testing the difference between two proportions.
3. Evaluate whether these assumptions are met (e.g., independence, normality conditions).
4. Calculate the sample proportion of approval in the first 1,000 days and the last 1,000 days. Also, compute the point estimate of the difference in proportions.
5. Calculate the standard error of the difference between the two sample proportions.
6. Perform the hypothesis test for the difference in proportions using a 5% significance level. Compute the test statistic and determine the p-value associated with the test statistic
7. Interpret the test statistic and p-value.
8. State whether you reject or fail to reject H_0 .
9. Explain the conclusion in the context of Obama's approval ratings.

Part B: President Trump

Repeat steps 1–9 for President Trump, comparing the first 500 days and last 500 days, using a **1% significance level**.

Mini-Group Project 1

Task 3: Proportion Analyses

You will use the “PEW” worksheet to complete this task

Part A: Education level vs political leaning

1. Evaluate the claim that "Republican-leaning Americans are more educated than Democratic-leaning Americans.". Formulate a hypothesis about whether Republican-leaning Americans are more educated than Democratic-leaning Americans. State the null hypothesis (H_0) and alternative hypothesis (H_1).
2. The EDUCATION variable takes the value of “High-educated” for cases when a respondent has “Four year college or university degree/Bachelor.s degree (e.g., BS, BA, AB)”, “Postgraduate or professional degree, including master's, doctorate, medical or law degree (e.g., MA, MS, PhD, MD, JD)”, or “Some postgraduate or professional schooling, no postgraduate degree” and “Low-educated” otherwise. Discuss potential limitations or biases in this definition.
3. Before conducting hypothesis testing, first perform a descriptive analysis of the data to assess whether there is any support for the claim. Present data in a clear table or chart and discuss your initial observations.
4. Calculate the point estimate for the difference in proportions
5. Calculate the standard error for the point estimate.
6. Verify that normal approximation conditions are met.
7. Construct a 90% confidence interval for the difference in proportions.
8. Interpret the confidence interval in the context of the research question.
9. Explain what it suggests about education levels across political leanings.
10. Determine if there is statistically significant evidence of a difference.

Part B: Education Level Across Genders

1. Evaluate the claim that "Males are more educated than Females."
 2. Formulate a hypothesis about whether males are more educated than females. State the null hypothesis (H_0) and alternative hypothesis (H_1).
 3. Perform a descriptive analysis of the data. Present data in a clear table or chart. Discuss your initial observations.
 4. Calculate the point estimate for the difference in proportions
 5. Calculate the standard error for the point estimate.
 6. Verify that normal approximation conditions are met.
 7. Construct a 99% confidence interval for the difference in proportions.
 8. Interpret the confidence interval in the context of the research question. Explain what it suggests about education levels across genders.
 9. Determine if there is statistically significant evidence of a difference.
-

Mini-Group Project 1

Make sure your report includes only your responses and does not contain any questions or prompts.
A sample submission should look like this:

Team 01

Task 1:

1. Blah blah blah
2. Blah blah blah
3. Blah blah blah
4. Blah blah blah
5. Blah blah blah

Task 2:

Part A:

1. Blah blah blah
2. Blah blah blah
3. Blah blah blah
4. Blah blah blah
5. Blah blah blah
6. Blah blah blah
7. Blah blah blah
8. Blah blah blah
9. Blah blah blah

Part B:

1. Blah blah blah
2. Blah blah blah
3. Blah blah blah
4. Blah blah blah
5. Blah blah blah
6. Blah blah blah
7. Blah blah blah
8. Blah blah blah
9. Blah blah blah

Task 3:

Part A:

1. Blah blah blah
 2. Blah blah blah
-

Part B:

1. Blah blah blah
 2. Blah blah blah
-
-