Final exam 2024

Due December 10, 2024, at 9pm

(Your name here)

NOTE: Start with the file `final_exam_2024.qmd` (available from the github repository at https://github.com/UChicago-pol-methods/IntroQSS-F24/tree/main/exams). Modify that file to include your answers. Make sure you can "render" the file (e.g. in RStudio by clicking on the `Render` button). Submit both the qmd file and the PDF via Canvas.

ADDITIONAL NOTE: Whenever possible, use the approach and code that we taught in this class: `mean()`, `var()`, `sd()`, `lm()`, `lm_robust()`, `modelsummary::modelsummary()`, etc. For example, if you are asked for a confidence interval or p-value, do not use `t.test()` or another function/package that gives an answer without requiring you to understand anything (and that we did not teach); compute the standard error as we taught you.

In this assignment, you will show your mastery of concepts and skills we covered in this course using data from a paper that will soon be published in the *American Political Science Review*.

The paper is "Election-Denying Republican Candidates Underperformed in the 2022 Midterms", by Janet Malzahn and Andrew Hall (link).

The complete replication package (Stata-formatted dataset, Stata code, documentation) is available on the APSR's Dataverse here.

A pruned version of their dataset in CSV format (`malzahn_hall_selected.csv`) is available on the course github under https://github.com/UChicago-pol-methods/IntroQSS-F24/tree/main/data. The rows of this dataset are Republican candidates for the U.S. House of Representatives, the U.S. Senate, and three offices of state government (Attorney General, Governor, Secretary of State) in the U.S. general election in November, 2022.

The columns of that dataset, as documented by the authors in their replication package, are:

- `state`: Abbreviation of state where election took place.
- `office`: The office for which the election was held {AG = Attorney General, SOS= Secretary of State, GOV = Governor, SEN = Senator, H = House}.
- `pres_voteshare_d`: The [Democratic] presidential voteshare in 2020 for the district/state.

- name: The name of the candidate.
- `inc`: Indicator variable for incumbency status (1 if the candidate is an incumbent, 0 otherwise).
- `deny`: Composite indicator variable for election denial (1 if a denier, 0 otherwise).
- `voteshare_g`: Voteshare [of the Republican candidate] in the [2022] general election.

```
set.seed(60637) # to make solutions comparable
```

```
# load dataset
```

Question 1: Sample mean

1a. Compute the average vote share among Republican candidates who denied the 2020 election result.

Answer:

```
# your code here
```

1b. Plot a histogram of the vote share among Republican candidates who denied the 2020 election result. Use the `breaks` argument to make sure that the histogram shows plenty of detail (with each bin no more than .05 wide). Would you say this is approximately a normal distribution?

Answer:

The dataset you are working with contains all Republicans who were candidates for federal or statewide office in the 2022 midterm, i.e. this is not a sample. But we will suppose for the rest of this assignment that this data *was* the result of iid sampling from a "super-population" of Republican candidates in 2022. Then we would be interested in how our estimates might vary across "samples" (and we can report standard errors, confidence intervals, p-values, etc).

1c. Using the formula $\mathrm{V}[\overline{X}_{(n)}] = \frac{\mathrm{V}[X]}{n}$, estimate a standard error for your estimate of the (population) mean vote share among Republican candidates who denied the 2020 election result.

Answer:

1d. Now use the bootstrap instead (with 10,000 resamples) to estimate a standard error for your estimate of the (population) mean vote share among Republican candidates who denied the 2020 election result.

Answer:

1e. Use a histogram to plot the distribution of your 10,000 bootstrap sample means, again making sure that the bins are narrow enough to show the distribution in detail. Does this distribution look normal? Explain its shape.

Answer:

1f. Using your bootstrap estimate of the standard error of the sample mean, compute a 95% confidence interval for the (population) mean vote share among Republican candidates who denied the 2020 election result. Report the confidence interval and explain what it means. Make sure to state what assumptions you are making in constructing this confidence interval.

Answer:

Question 2: difference in means

2a. Malzahn and Hall's analysis focuses on the subset of candidates who faced an opponent. Create a version of the data that excludes candidates who were unopposed, i.e. those with vote shares of 1 in the 2022 election. Use this version for all analysis for the rest of the problem set.

Answer:

```
# your code here
```

2b. Compute the difference between the average vote share among Republican candidates who denied the 2020 election result and those who did not. Did candidates who denied the 2020 election result perform worse in 2022 than those who did not?

Answer:

2c. Again supposing this is a sample from a super-population of candidates, report a standard error for this difference in means. Use the variance rule as in question 1b on Problem set 7, and assume the two sample means have zero covariance (as is reasonable given iid sampling). Interpret your result, i.e. state what it means in a sentence.

Answer:

2d. Compute the same standard error using the (naive) bootstrap (10000 re-samples) and report it.

Answer:

2e. Using your estimate of the difference in means, and using your bootstrap estimate of the standard error, compute a two-sided p-value, where the null hypothesis is that the population

difference in means is 0. Explain in words what this number means, and state any assumptions you are making.

Answer:

Question 3: Election denial and district/state partisanship

3a. Make a scatterplot showing each candidate's 2022 voteshare on the vertical axis and the Democratic presidential vote share in 2020 for the candidate's district/state on the horizontal axis. You should notice one unusual observation. By looking at your dataset and doing a bit of internet research, determine whether that observation is an error or not.

Answer:

```
# your code here
```

3b. Using regression, fit the coefficients to predict (the probability of) a candidate being an election denier as a linear function of the state/district's support for Biden in 2020, i.e. the following linear predictor:

$$\text{Deny}_i = \beta_0 + \beta_1 \text{PresVoteShareD}_i$$

Interpret your results.

Answer:

3c. Based on this regression result, why might it be important to consider `pres_voteshare_d` when we attempt to determine whether voters in 2022 punished candidates who denied the 2020 election result?

Answer:

Question 4: Regression analysis

4a. Present a regression table (using `modelsummary::modelsummary()` or similar) that shows the results of the following analysis, one in each column:

- Regress 2022 vote share on an indicator for being a denier
- Regress 2022 vote share on an indicator for being a denier and the 2020 Democratic presidential vote share

- Regress 2022 vote share on an indicator for being a denier and the 2020 Democratic presidential vote share, only for House elections
- Regress 2022 vote share on an indicator for being a denier and the 2020 Democratic presidential vote share, only for statewide elections (i.e. all but House elections)
- Regress 2022 vote share on an indicator for being a denier and the 2020 Democratic presidential vote share, only for statewide elections (i.e. all but House elections) but excluding the unusual observation you identified above

Use `estimatr::lm_robust()` instead of `lm()`, so that your standard errors are robust to heteroskedasticity.

Answer:

4b. Interpret the coefficient on `deny` in each of the regressions. Highlight and attempt to explain any important differences. (Do not yet discuss statistical significance or other inferential considerations.)

Answer:

Question 5: Does electoral punishment vary with partisanship?

If voters who support Biden are especially intolerant of election-denying Republicans, you might expect that the predicted difference in 2022 vote share between election deniers and non-deniers would be larger in districts/states that supported Biden more strongly in 2020.

To assess this, we will consider a linear prediction of the form

$$Y_i = \beta_0 + \beta_1 \text{Deny}_i + \beta_2 \text{PresVoteShareD}_i + \beta_3 \text{Deny}_i \times \text{PresVoteShareD}_i.$$

5a. Focusing on statewide races (i.e. non-House races), run a regression to fit the coefficients for this linear prediction. Do this both with and without the unusual observation noted above. Report the resulting coefficients for both models in a regression table using e.g. `modelsummary()`.

Answer:

```
# your code here
```

5b. For each of these two regression models, state the predicted difference in vote share between election deniers and non-deniers at three levels of 2020 Democratic vote share: .3, .5 and .7. Interpret the results.

Answer:

5c. Focusing on the regression model that includes the unusual observation, compute the standard error for the predicted difference in vote share between a denier and a non-denier in a district/state where Biden received a vote share of .5 in 2020. (Hint: Write this difference in terms of estimated regression coefficients; use the variance rule to compute the variance in this expression; extract the necessary estimates from the regression's variance covariance matrix to compute the standard error.)

Answer:

Question 6: Comparing House races and statewide races

The soon-to-be-published paper claims that election deniers were punished more in statewide races than in House races, which they speculate is because House races are more partisan. Above you estimated separate regressions and confirmed that the predicted difference in 2022 vote share between deniers and non-deniers (conditional on 2020 results) was bigger in statewide races than in House races. Here we will examine this difference in other ways.

6a. Use `estimatr::lm_robust()` to regress 2022 vote share on an indicator for being a denier and the 2020 Democratic presidential vote share, as well as an indicator for running in a statewide race (i.e. `office != 'H"`) and an interaction between this indicator and the indicator for being a denier. Show the regression coefficients in a regression table using `modelsummary()`.

Answer:

6b. Interpret the estimated coefficients on `deny` and the interaction term in this regression.

Answer:

6c. Extract the two-sided p-value on the interaction term from the regression output (where the null hypothesis is that the interaction term is zero), report it, and interpret it.

Answer:

6d. Here is a different way to compute a p-value for the statewide-vs-House difference in the election-denying "punishment". In question 3 you computed the predicted difference in vote share for deniers and non-deniers in separate regressions for statewide races and House races. Use your regression table from question 3 plus the variance rule to compute a standard error for the difference in the coefficients across the two regressions, and use that to compute a two-sided p-value for the difference in those coefficients (with the null hypothesis being that there is no difference).[1]

[1] Your answer should be similar to the p-value in question 5c, but don't expect it to be identical. When you ran the regression with the interaction, you assumed that the relationship between 2020 results and 2022 results is the same in House races and other races; in the separate regressions, you did not assume this.

Answer: