

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.

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

```
library(tidyverse)
library(modelsummary)
```

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.

```
# load dataset
dd <- read_csv("~/projects/IntroQSS-F24/data/malzahn_hall_selected.csv")
```

Question 1: Sample mean

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

Answer:

```
(denier_mean <- mean(dd$voteshare_g[dd$deny == 1]))
```

```
[1] 0.5684162
```

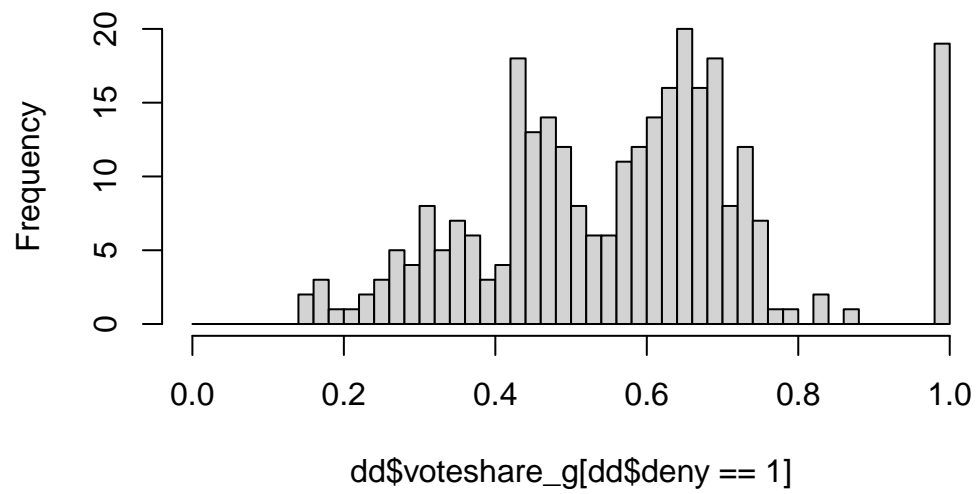
1b. Plot a histogram of the vote share among Republican candidates who denied the 2020 election result. Use the **breaks** argument to make sure that the histogram shows plenty of detail (with each bin no more than .05 wide). Would you say this is approximately a normal distribution?

Answer:

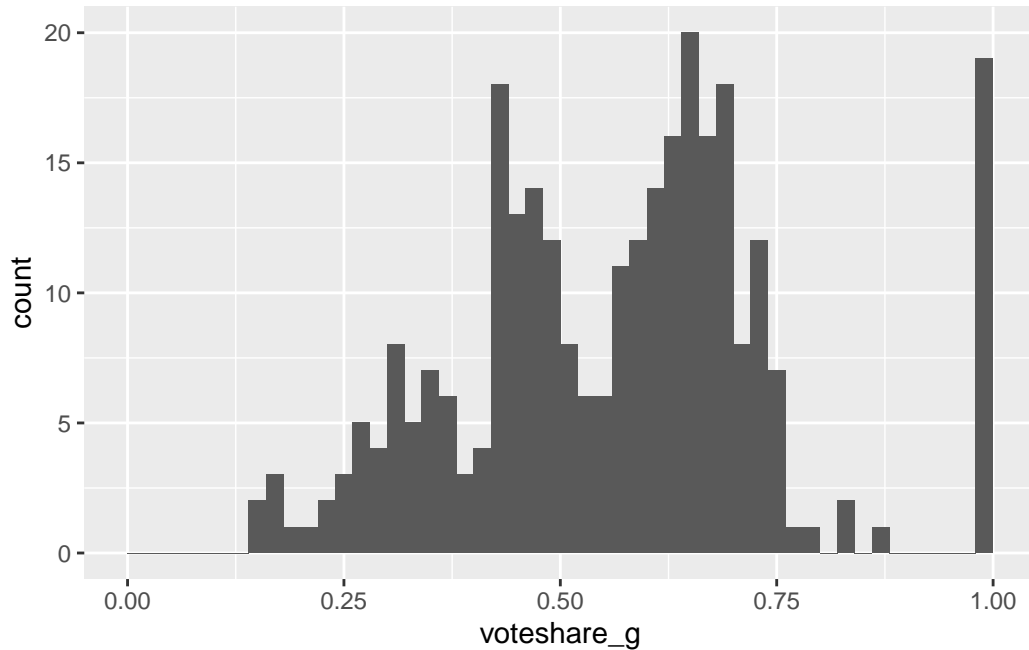
Here is the histogram:

```
# base R plot
hist(dd$voteshare_g[dd$deny == 1],
      breaks = seq(0, 1, length = 51)) # bins width .02
```

Histogram of dd\$voteshare_g[dd\$deny == 1]



```
# ggplot
dd |>
  filter(deny == 1) |>
  ggplot(aes(x = voteshare_g)) +
  geom_histogram(breaks = seq(0, 1, length = 51))
```



No, this is not a normal distribution. Notably, there is a big spike at 1 for the uncontested races.

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

Answer:

```
sampling_variance <- var(dd$voteshare_g[dd$deny == 1])/sum(dd$deny == 1)
(vx_n_se <- sqrt(sampling_variance))
```

```
[1] 0.01093539
```

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:

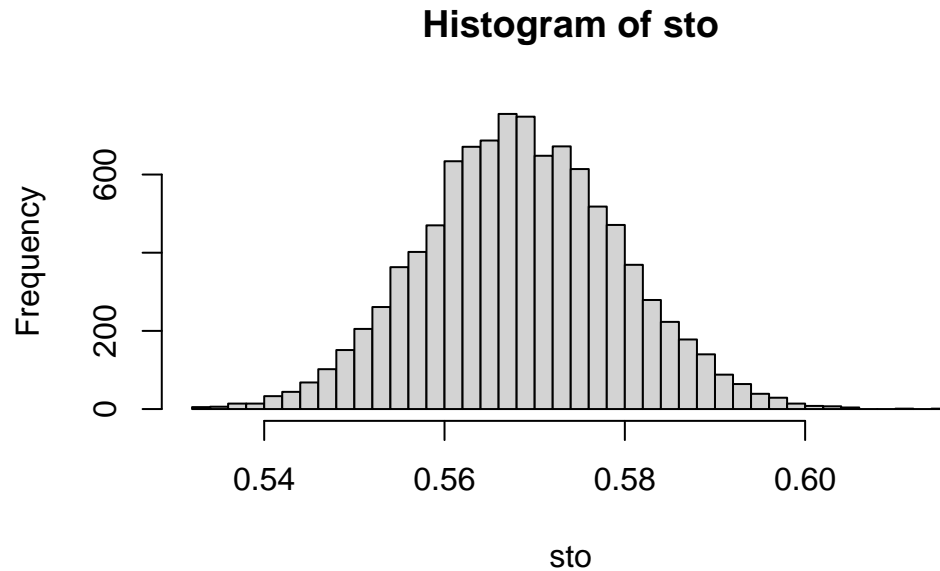
```
# bootstrap sampling distribution
m <- 10000
sto <- rep(NA, m)
for(i in 1:m){
  sto[i] <- mean(sample(dd$voteshare_g[dd$deny == 1],
                        size = sum(dd$deny == 1),
                        replace = T))
}
(bootstrap_se <- sd(sto))
```

```
[1] 0.0108977
```

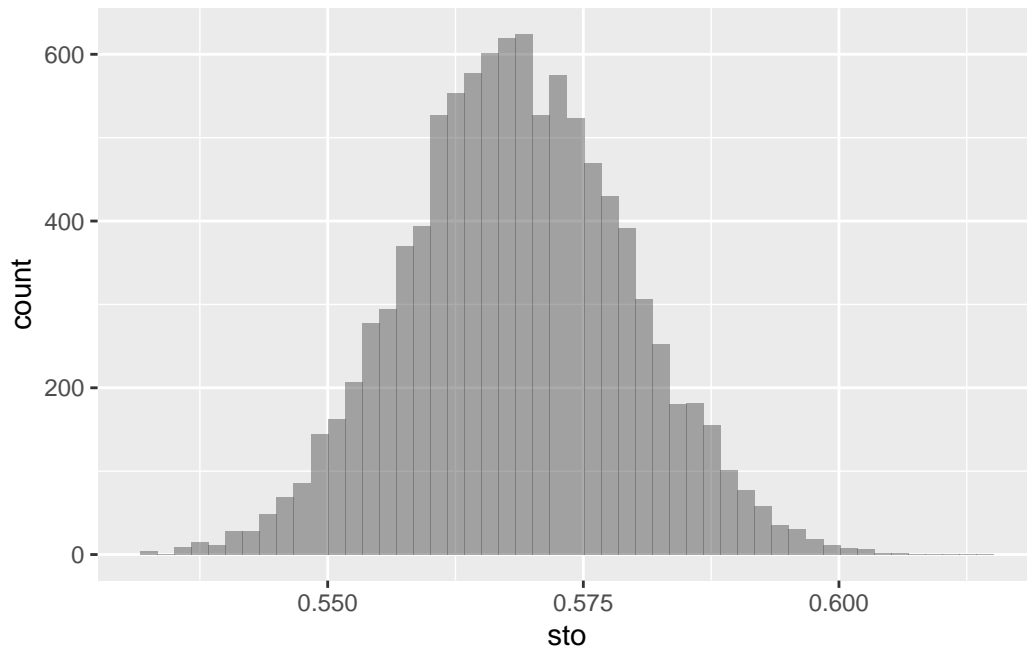
1e. Use a histogram to plot the distribution of your 10,000 bootstrap sample means, again making sure that the bins are narrow enough to show the distribution in detail. Does this distribution look normal? Explain its shape.

Answer:

```
hist(sto, breaks = 50)
```



```
tibble(sto) |>  
  ggplot(aes(x = sto)) +  
  geom_histogram(bins = 50, alpha = .5)
```



Yes it does look approximately normal. The normal shape is a consequence of the central limit theorem for the sample mean (which says that the sampling distribution of the sample mean is asymptotically normal, given mild regularity conditions). Here the sample size is large enough (given the underlying distribution) for the sampling distribution to appear approximately normal.

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:

```
(boot_ci <- denier_mean + c(-1, 1)*1.96*bootstrap_se)
```

```
[1] 0.5470567 0.5897757
```

My 95% confidence interval runs from about 0.547 to about 0.59. Assuming that the mild regularity conditions required by the CLT are met, about 95% of the confidence intervals that I construct in this way will contain the relevant estimand (in this case, the population mean vote share among Republican candidates who denied the 2020 election result). If I am not a strict frequentist, and if I don't have any reason to think that this is an unusual sample, then I can say that there is a 95% chance that this particular estimand (the population mean) is

contained in this confidence interval, or that I am 95% confident that this particular estimand is contained in this confidence interval.

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:

```
# base R version
ddd <- dd[dd$voteshare_g < 1,]
# tidy version
dd |>
  filter(voteshare_g < 1) -> ddd
```

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:

```
(diff_means <- mean(ddd$voteshare_g[ddd$deny == 1]) -
  mean(ddd$voteshare_g[ddd$deny == 0]))
```

```
[1] 0.07789705
```

Candidates who denied the 2020 election result obtained a vote share about 7.8 percentage points higher than those who did not. This does not mean that they did better on average than they would have if they had accepted the election result. It could be that election deniers ran in places where Republicans generally tend to do better. (We'll see that this is the case.)

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:

Denote by Y_1 the RV representing vote share among election denying Republicans and \bar{Y}_1 the sample mean for this RV, with Y_0 and \bar{Y}_0 defined analogously. Then by the variance rule,

$$V[\bar{Y}_1 - \bar{Y}_0] = V[\bar{Y}_1] + V[\bar{Y}_0] - 2\text{Cov}[\bar{Y}_1, \bar{Y}_0],$$

and assuming zero covariance we have

$$V[\bar{Y}_1 - \bar{Y}_0] = \frac{V[Y_1]}{n_1} + \frac{V[Y_0]}{n_0}$$

where n_1 is the size of the denier sample and n_0 is the size of the non-denier sample.

Then our estimated standard error is this:

```
(se_var_rule = sqrt(
  var(ddd$voteshare_g[ddd$deny == 1])/sum(ddd$deny == 1) +
  var(ddd$voteshare_g[ddd$deny == 0])/sum(ddd$deny == 0)))
```

```
[1] 0.01317
```

This is an estimate of the standard deviation of the sampling distribution of our estimator of the difference in mean vote share between deniers and non-deniers.

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

Answer:

```
m <- 10000
sto <- rep(NA, m)
for(i in 1:m){
  ddd_resamp <- ddd[sample(1:nrow(ddd), replace = T), ]
  sto[i] <- mean(ddd_resamp$voteshare_g[ddd_resamp$deny == 1]) -
    mean(ddd_resamp$voteshare_g[ddd_resamp$deny == 0])
}
(se_boot <- sd(sto))
```

```
[1] 0.01298471
```


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:

Assuming that the sampling distribution of the difference in means is approximately normal (which is an asymptotically valid assumption given “mild regularity conditions”, and works well in practice for the difference in means given a large enough sample size), and assuming that the bootstrap procedure gives me a good estimate of the standard deviation of that sampling distribution under the null hypothesis (which is again valid asymptotically given “mild regularity conditions”, and works well in practice for the difference in means given a large enough sample), the p-value is

```
format(2*(1 - pnorm(diff_means/se_boot)), scientific = F)
```

```
[1] "0.000000001983665"
```

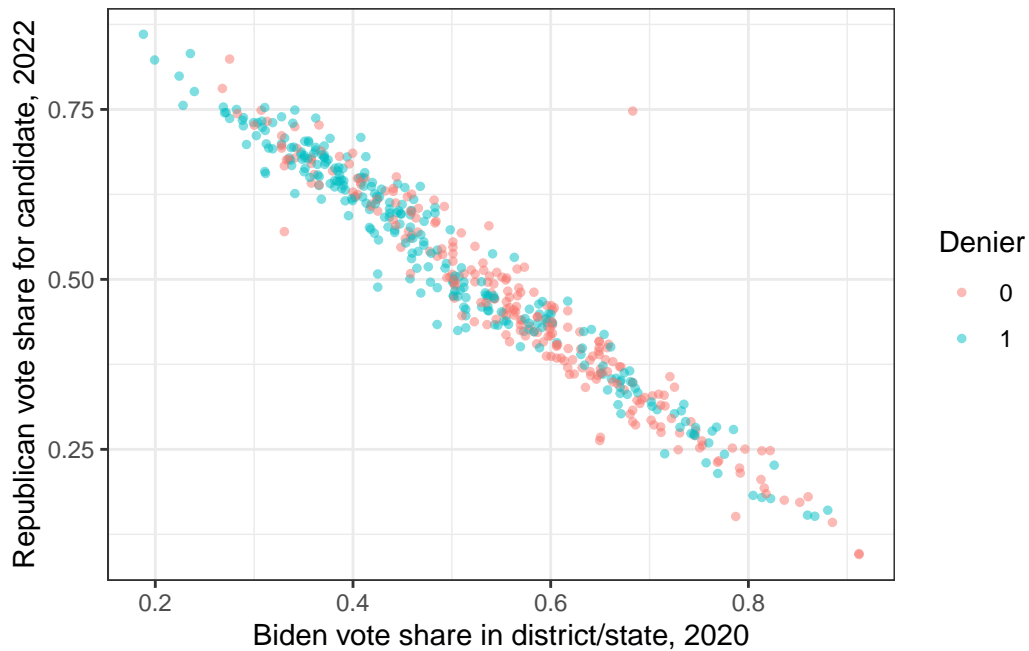
This number is an estimate of the probability of obtaining a difference in means at least as far from zero as the one I obtained if the true population difference in means were in fact 0.

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:

```
ddd |>
  ggplot(aes(x = pres_voteshare_d, y = voteshare_g, col = factor(deny))) +
  geom_point(size = 1, alpha = .5) +
  theme_bw() +
  labs(x = "Biden vote share in district/state, 2020",
       y = "Republican vote share for candidate, 2022",
       col = "Denier")
```



The point at about (.68, .75) (i.e. a case where 68% supported Biden in 2020 and 75% supported the Republican in 2022) represents Vermont Governor Phil Scott, who is a very popular Republican governor in a strongly Democratic state. So it's not an error; it is an unusual case.

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:

```
deny_reg <- lm(deny ~ pres_voteshare_d, data = ddd)
summary(deny_reg)
```

Call:

```
lm(formula = deny ~ pres_voteshare_d, data = ddd)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.7774	-0.4732	0.2529	0.4163	0.8137

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.0359	0.0798	12.982	< 2e-16 ***
pres_voteshare_d	-0.9651	0.1469	-6.572	1.23e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.48 on 507 degrees of freedom

Multiple R-squared: 0.07851, Adjusted R-squared: 0.07669

F-statistic: 43.2 on 1 and 507 DF, p-value: 1.232e-10

The coefficient on `pres_voteshare_d` indicates that for a .1 increase in Biden's vote share in a district/state 2020, the predicted probability that the 2022 Republican candidate in that district/state is an election denier goes down by about .097. This is a strong relationship. The intercept indicates that, in a hypothetical district/state where Biden received no votes in 2020, the predicted probability of the Republican candidate denying the election is 1.0359; in a hypothetical district/state where Trump received no votes in 2020, the predicted probability of the Republican candidate denying the election is 0.07.¹

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: Compared to non-deniers, election deniers in 2022 tended to be running in districts that were more Republican in 2020. Therefore, when we compare the average 2022 results of deniers and non-deniers (as we did in previous questions above), we conflate the effect of election denial with pre-existing differences in the districts where deniers and non-deniers were running. This comparison of means could thus produce misleading estimates of the effect of election denying i.e. the degree to which voters punish election deniers.

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:

¹It is obviously not possible for a probability to be above 1, which is one reason why some people insist on using logit or probit when the dependent variable is binary, but there are no districts/states where Biden's vote share was zero, so in this case and all other cases I've ever seen the linear probability model produces sensible predictions where there is non-negligible data.

- 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:

```
reg_1 <- estimatr::lm_robust(votesshare_g ~ deny,
                             data = ddd)
reg_2 <- estimatr::lm_robust(votesshare_g ~ deny + pres_votesshare_d,
                             data = ddd)
reg_3 <- estimatr::lm_robust(votesshare_g ~ deny + pres_votesshare_d,
                             data = ddd |> filter(office == "H"))
reg_4 <- estimatr::lm_robust(votesshare_g ~ deny + pres_votesshare_d,
                             data = ddd |> filter(office != "H"))
reg_5 <- estimatr::lm_robust(votesshare_g ~ deny + pres_votesshare_d,
                             data = ddd |> filter(office != "H" & name != "SCOTT, PHIL"))

modelsummary(list("All" = reg_1,
                  "All" = reg_2,
                  "House" = reg_3,
                  "Statewide" = reg_4,
                  "Statewide,\nno VT Gov" = reg_5),
              stars = T,
              coef_map = c("(Intercept)" = "Intercept",
                           "deny" = "Election denier",
                           "pres_votesshare_d" = "Biden vote share (2020)"),
              gof_map = c("r.squared", "nobs"))
```

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:

	All	All	House	Statewide	Statewide, no VT Gov
Intercept	0.460*** (0.009)	1.046*** (0.006)	1.046*** (0.006)	1.021*** (0.030)	1.043*** (0.021)
Election denier	0.078*** (0.013)	−0.006+ (0.003)	0.000 (0.003)	−0.029** (0.009)	−0.023** (0.007)
Biden vote share (2020)		−1.034*** (0.010)	−1.037*** (0.009)	−0.971*** (0.067)	−1.028*** (0.038)
R2	0.064	0.943	0.968	0.790	0.889
Num.Obs.	509	509	393	116	115

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

In the first column, the coefficient on **deny** reproduces the difference in means from above (estimate and standard error): on average, election-deniers got a vote share that is .078 higher than non-deniers.

The the second column, we add the 2020 election result to the regression. Conditional on the 2020 election result, there is only a small average difference in vote share between deniers and non-deniers: the predicted vote share of deniers is .006 lower. This highlights the importance of adjusting for the fact (documented above) that deniers tended to run in more strongly Republican areas, so their stronger average performance is partly due to that.

In the third column, we use the same regression but focus on the House races only. Adjusting for 2020 results, deniers and non-deniers receive about the same predicted vote share.

In the fourth column, we use the same regression but focus on statewide races only. Adjusting for 2020 results, deniers receive a predicted vote share about .029 lower than non-deniers.

In the final column, we again focus on statewide races but exclude the Phil Scott case. The estimated difference between deniers and non-deniers (adjusting for 2020 results) is now smaller – approximately .023. Eliminating that one observation, we thus find a punishment that is about 20% smaller. The basic conclusion is not very different from what Malzahn and Hall report: there was a small but detectable electoral penalty for election deniers.

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.

	All statewide	No VT Gov
Intercept	1.003*** (0.045)	1.040*** (0.026)
Election denier	0.023 (0.056)	-0.014 (0.043)
Biden vote share (2020)	-0.936*** (0.097)	-1.021*** (0.049)
Interaction	-0.105 (0.115)	-0.019 (0.079)
R2	0.792	0.889
Num.Obs.	116	115

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

5a. Focusing on statewide races (i.e. non-House races), run a regression that assesses this conjecture using an interaction. 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:

```
intx_reg <- estimatr::lm_robust(votesshare_g ~ deny*pres_votesshare_d,
                              data = ddd |> filter(office != "H"))
intx_reg_2 <- estimatr::lm_robust(votesshare_g ~ deny*pres_votesshare_d,
                                 data = ddd |> filter(office != "H" &
                                                       name != "SCOTT, PHIL"))
modelsummary(list("All statewide" = intx_reg,
                  "No VT Gov" = intx_reg_2),
              stars = T,
              coef_map = c("(Intercept)" = "Intercept",
                           "deny" = "Election denier",
                           "pres_votesshare_d" = "Biden vote share (2020)",
                           "deny:pres_votesshare_d" = "Interaction"),
              gof_map = c("r.squared", "nobs"))
```

5b. For each of the two regression models in the previous question, 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:

This table reports the estimated difference between the vote share of election deniers and non-deniers at three levels of 2020 Biden support, for each of the models:

2020 Biden support	Model 1	Model 2
.3	.023 - .105 x .3 = -.0085	-.014 - .019 x .3 = -.0197
.5	.023 - .105 x .5 = -.0295	-.014 - .019 x .5 = -.0235
.7	.023 - .105 x .7 = -.0505	-.014 - .019 x .7 = -.0273

So the estimates from the first model (which includes Phil Scott) appear to support the conjecture that the difference in vote share between deniers and non-deniers is bigger in more Democratic districts/states. The second model does too, but the change across levels of Democratic support is smaller.

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:

If the estimated BLP is written

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \text{Deny}_i + \hat{\beta}_2 \text{PresVoteShareD}_i + \hat{\beta}_3 \text{Deny}_i \times \text{PresVoteShareD}_i,$$

then 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 is $\hat{\beta}_1 + .5\hat{\beta}_3$.

By the variance rule, the variance of $\hat{\beta}_1 + .5\hat{\beta}_3$ is $V[\hat{\beta}_1] + .5^2 V[\hat{\beta}_3] + 2 \times .5 \times \text{Cov}[\hat{\beta}_1, \hat{\beta}_3]$.

We can estimate this variance using the regressions's variance-covariance matrix:

```
vc <- vcov(intx_reg)
(var_5c <- vc["deny", "deny"] +
  .5^2*vc["deny:pres_voteshare_d", "deny:pres_voteshare_d"] +
  2*.5*vc["deny", "deny:pres_voteshare_d"])
```

```
[1] 8.612877e-05
```

The standard error is the square root of this, i.e. 0.0093.

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:

Here is the regression equation:

```
intx_reg_2 <- estimatr::lm_robust(votesshare_g ~ deny*I(office != "H") +
                                pres_votesshare_d,
                                data = ddd)
```

And here is the regression table:

```
modelsummary(intx_reg_2,
              stars = T,
              coef_map = c("(Intercept)" = "Intercept",
                           "deny" = "Denier",
                           'I(office != "H")TRUE' = "Statewide race",
                           'deny:I(office != "H")TRUE' = "Denier X Statewide",
                           "pres_votesshare_d" = "Biden 2020 vote share"),
              gof_map = c("r.squared", "nobs"))
```

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

Answer:

The coefficient on `deny` (.001) indicates the predicted difference in vote share for deniers and non-deniers in House races, conditional on 2020 results in the district/state. The interaction term (-.030) indicates how much this difference differs between statewide races and House races; thus there is a difference of about -.029 in statewide races and basically no difference in House races. The coefficient on “Statewide race” (.009) indicates the predicted difference in vote share for Republicans running in statewide races vs House races, conditional on the 2020 results in the district/state.

	(1)
Intercept	1.041*** (0.007)
Denier	0.001 (0.003)
Statewide race	0.009 (0.009)
Denier X Statewide	−0.030** (0.011)
Biden 2020 vote share	−1.028*** (0.012)
R2	0.945
Num.Obs.	509
+ p <0.1, * p <0.05, ** p <0.01, *** p <0.001	

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:

The (two-sided) p-value on the interaction term is:

```
summary(intx_reg_2)$coefficients['deny:I(office != "H")TRUE', "Pr(>|t|)"]
```

```
[1] 0.004131017
```

This is an estimate of the probability of obtaining an interaction term at least as far from zero as the one we obtained if the true population interaction term were zero. It relies on the normality of the sampling distribution of this regression coefficient, which relies on the central limit theorem for plug-in estimators.

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).²

Answer:

```
# get the standard errors
se_3 <- summary(reg_3)$coefficients["deny", "Std. Error"]
se_4 <- summary(reg_4)$coefficients["deny", "Std. Error"]
se_combined <- sqrt(se_3^2 + se_4^2)
coef_diff <- abs(coef(reg_3)["deny"] - coef(reg_4)["deny"])
t_stat <- coef_diff/se_combined
(diff_p <- 2*(1 - pnorm(t_stat)))
```

```
deny
0.00314691
```

The p-value is 0.0031, which is similar but not the same. To get something closer to this p-value via regression, you can re-run the interaction regression but allow the coefficient on `pres_voteshare_d` to vary between House elections and other elections:

```
intx_reg_3 <- estimatr::lm_robust(voteshare_g ~ deny*I(office != "H") +
                                pres_voteshare_d*I(office != "H"),
                                data = ddd)
summary(intx_reg_3)$coefficients['deny:I(office != "H")TRUE', "Pr(>|t|)"]
```

```
[1] 0.003294092
```

²Your 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.