

Final assignment: general comments and sample solutions

Due December 5, 2023, at 9pm

Andy Eggers

General comments

The final exams for this class showed a lot of learning. I know that many students started with very little background, so it was pleasing to see you showing facility with foundations of social science statistics.

That said, it was frustrating to see many mistakes and misunderstandings. I will draw on what I observed on the final to try to improve the class next year, but I hope you can also learn from these mistakes.

Moksha and I entered comments in the PDF of your assignments on Canvas. Please review these to try to learn what you missed.

Here are three points that came up several times:

- When asked to indicate the assumptions they were making, many students listed assumptions that were not necessary for the analysis they were doing. The most common one was people saying that they were assuming a variable in the dataset was normally distributed. **At no point in this course did we ever assume that about a variable we analyzed.** Our normal approximation based confidence intervals and p-values do require that *sampling distributions* for statistics were normally distributed, and this is true asymptotically for the sample mean given mild regularity conditions under the CLT and more generally for plug-in estimators. The whole point of the CLT is that the sampling distribution of the sample mean is normal even if the variable is not normally distributed. (This is why it was very frustrating to see students say, “I assume the CLT holds. I also assume the variable is normal.”) If this does not make sense to you, please review it as it is very important. Similarly, many students said that they assumed normal residuals/errors and linearity when they ran a regression. Again, **we never assumed this.** In fact, a major point of the course and the book we used was to recognize that in many cases such assumptions are not necessary – e.g. if we view OLS as

approximating the BLP that approximates the CEF, then we don't need to assume that the CEF is linear, and we don't need to assume normality of errors for inference if we have enough data (because OLS is a plug-in estimator and plug-in estimators have normal sampling distributions under mild regularity conditions). This was a very common and very frustrating error to see because this was intended to be an important theme of the course.

- Many students provided poor interpretations of p -values. It was not enough to just say “the p -value is low so I reject the null hypothesis”. Many students said something like “This is the probability of observing the data I have if the null is true.” This is close, but it is (i) not specific enough about what aspect of the data is being tested (e.g. the mean of a variable or the slope coefficient), and (ii) failing to recognize that it's the probability of obtaining a statistic as far from the null *or further*. You'll see below that the language I use to describe the p -value is, “This p -value approximates the probability that, assuming [state null hypothesis], we would obtain a random iid sample in which the [test statistic] deviates from [null value] by at least as much as the observed test statistic does.” In future I will give students more practice at expressing the p -value this way.
- Many students used the `t.test()` function when asked for a confidence interval and p -value. We never used this function in the class, so I thought it was clear that we were asking you to construct confidence intervals and p -values from the standard error and/or bootstrap as in all of the assignments and slides. We only took off partial credit for this because we didn't specifically say not to use a function like this. In future I will be more explicit. Note that in many cases when people used this function, they used it in ill-conceived ways, e.g. to produce a two-sided p -value for the null hypothesis that the mean of a variable that ranges from 0 to 100 is 0 (it's not possible for the population mean to be below zero, so a two-sided test is doesn't make sense, and anyway a single observation with a value above 0 would disprove the null).
- There continues to be confusion about the role of dividing by n in computing a standard error. The only time we ever did this was for the sample mean, but students used the same formula when using the bootstrap, which is wrong. The standard error of an estimator is the standard deviation of the estimator in repeated sampling. The bootstrap samples of your estimator is a simulation of the distribution of the estimator in repeated sampling, so you just take the standard deviation of it, and there is no need to divide by n or \sqrt{n} .

The appearance of things we didn't teach made me think that students were not understanding our presentation of the core material and relying instead on outside sources (ChatGPT, Google, other classes they've taken), which is disappointing and tells me we need to do better.

Sample solutions

The analysis below is to show you how I would implement the tasks on the final. I hope it helps clarify things.

I will use some components of the tidyverse below, so I load that library:

```
library(tidyverse)
options(scipen = 9999) # discourage scientific notation
```

1. Data description

What do the rows of the dataset represent (people, municipalities, a country-year, etc)? What are some of the important variables you will be using, and (if not obvious) how were they measured? Is this a sample from a population (if so what is that population?), or is this the whole population? (If your dataset is not a sample from a population, you will assume that is for the purpose of inference below, as is standard in social science research.)

Each row in the dataset describes a single US individual based in a swing state (competitive state) before the 2016 election who visited a vote swap website (trumptraders.org) and asked to be matched with a Clinton supporter in a non-swing state for a **vote swap**. In a vote swap, the swing-state voter agrees to vote for Clinton and the non-swing-state voter agrees to vote for the swing-state voter's preferred third party candidate. For example, a voter in Wisconsin who wants to vote for Jill Stein is matched with a Hillary Clinton supporter in Massachusetts, and the Wisconsin voter votes for Clinton while the Massachusetts voter votes for Stein.

Important variables include:

- **vote_choice**: respondent's vote choice
- **swap_treatment**: 1 if assigned to a non-swing-state voter to discuss a swap, 0 if not
- **matched**: 1 if actually exchanged message with a non-swing-state voter for a swap
- demographic variables including **white**, **female**,
- **swap_open**: respondent's test response when asked "In a few words or sentences, please tell us what attracted you to try to swap votes. Did you have any concerns?"

This is a random sample ($n = 4,500$) from around 45,000 swing-state voters who signed up to swap votes at trumptraders.org. Everyone in the sample was asked to take a post-election survey. Only 218 of the 4500 subjects filled out the survey. Survey respondents may not be a random sub-sample. (This is something we can investigate below.)

The data is from an experiment conducted by Alexander Coppock. The paper from the experiment ("A Field Experimental Test of Vote Swapping", July 13 2018) is available on the internet but has not been published. I am collaborating with Coppock to explore publishing the paper.

2. Potential research questions

What research questions might this data be useful for answering?

The main research question of interest is whether participating in a vote swap affects the vote choice of the participant. Define the outcome Y_i as 1 if swing-state voter i votes for Clinton and 0 otherwise; the treatment D_i is 1 if the voter is paired with a swap partner and 0 otherwise. The potential outcomes $Y_i(1)$ and $Y_i(0)$ indicate the outcome if the voter is or is not paired with a swap partner. The estimand is $E[Y_i(1) - Y_i(0)]$, the ATE of pairing a voter who enrolls at trumptraders.org with a swap partner.

If the swap works as intended, the swing state voter who signs up for a swap would vote for a third party candidate if no swap is agreed upon but would vote for Clinton otherwise. For such a voter, $Y_i(1) = 1$ and $Y_i(0) = 0$.

Logically, there are three other possibilities.

- The swing-state voter may vote third-party whether or not they are assigned to a swap partner, i.e. $Y_i(1) = Y_i(0) = 0$. This could be because the voter never intended to vote for Clinton and engaged in the swap to trick their partner into supporting their preferred third-party candidate, or because the voter intended to vote Clinton if a swap was agreed but then changed their mind. (Swaps are not binding.)
- The swing-state voter may vote for Clinton whether or not they are assigned to a swap partner, i.e. $Y_i(1) = Y_i(0) = 1$. This could be because the voter always intended to vote for Clinton and engaged in the swap to trick their partner into supporting their preferred third-party candidate, or because the voter intended to vote third-party in the absence of a swap but then changed their mind.
- The swing-state vote may vote for Clinton only if a swap is *not* agreed. This is perverse behavior.

3. Dataset loading, variable creation and other data manipulation

The tasks below will probably require you to create new variables and exclude missing observations; you may also want to rename some variables for your convenience. Put your data loading and data manipulation code here so we can see the skills you have accumulated in this area.

```
dat <- readRDS("../trumptraders_for_AE/tt_cleaned.rds")
```

The data is quite clean, but I can demonstrate some data cleaning skills anyway:

```

dat$dist_channel <- dat$DistributionChannel # renaming a variable
# does the respondent mention Trump when explaining why they were interested in a swap?
dat$mentioned_trump <- as.integer(stringr::str_detect(dat$swap_open, "Trump"))
dat$vote_choice_chr <- dplyr::recode(dat$vote_choice,
                                     "1" = "Trump",
                                     "2" = "Clinton",
                                     "3" = "McMullin",
                                     "4" = "Stein",
                                     "5" = "Johnson",
                                     "6" = "Other",
                                     "99" = "Did not vote")

# I think this was reverse coded as disapproval
dat$pres_approval <- 10 - as.integer(dat$pres_approve)
# indicator for self-identified democrats
dat$dem <- as.integer(dat$pid_7 %in% c(1,2))

```

4. Inference for a mean

Estimate the mean of 3 important variables in the dataset. For each one, report a 95% confidence interval in a table. For one of the variables, state a null hypothesis that could be considered interesting to test and report a p-value for that null hypothesis. Explain in words what your p-value means. In all cases explain what assumptions you are making.

I will report the proportion white, the proportion women, and approval of the president.

I am making these confidence intervals under the assumption that the sampling distribution of the sample mean is approximately normal. (Otherwise, the value 1.96 has no special relationship to the quantiles of the sampling distribution.) This is true asymptotically by the CLT.

```

dat |>
  select(white, female, pres_approval) |>
  pivot_longer(cols = everything()) |>
  group_by(name) |>
  summarize(Mean = mean(value, na.rm = T),
            SD = sd(value, na.rm = T),
            n = sum(!is.na(value))) |>
  mutate(`Std. error` = SD/sqrt(n),
         `Conf low` = Mean - 1.96*`Std. error`,

```

Variable	Mean	n	Std. error	Conf low	Conf high
female	0.532	218	0.034	0.466	0.598
pres_approval	0.713	216	0.104	0.508	0.918
white	0.858	218	0.024	0.811	0.904

```

      `Conf high` = Mean + 1.96*`Std. error`) |>
select(-SD) |>
dplyr::rename(Variable = name) -> ci_table

ci_table |>
  kableExtra::kbl(digits = 3, booktabs = T) |>
  kableExtra::kable_styling(full_width = F)

```

A possibly interesting null hypothesis is that the proportion of women in the sample is .5. We will compute a two-tailed p -value for this null hypothesis (again assuming normality of the sampling distribution for the sample mean, which is true asymptotically under the CLT).

```

this_std_error <- ci_table$`Std. error`[ci_table$Variable == "female"]
t_stat <- abs(mean(dat$female, na.rm = T) - .5)/this_std_error
2*(1 - pnorm(t_stat))

```

```
[1] 0.34314
```

This p -value approximates the probability that, assuming the true population proportion of women is .5, we would obtain a random iid sample in which the proportion of women deviates from .5 by at least as much as the observed proportion does.

5. Difference between means

Define two means that could be interesting to compare. (It could be two variables in the dataset, or a single variable for two groups of observations defined by another variable, e.g. average growth rate of democracies and non-democracies). Estimate the difference in the two means, report a confidence interval for that difference, and report a p -value. (Do not use regression for this question.) Explain in words what your p -value means. Explain what assumptions you are making.

The most important difference in means to estimate is the difference in the proportion voting for Clinton between those assigned matches (treated) and those not assigned matches (control).

Note that, by the Variance Rule,

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

Given iid sampling, the covariance term is zero. Using the formula for the variance of the sample mean (which also assumed iid sampling), we then have

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

where n_1 and n_0 are the number of treated and control units.

```
y1 <- dat$voted_HRC[dat$swap_treatment == 1]
y0 <- dat$voted_HRC[dat$swap_treatment == 0]
diff_in_means <- mean(y1, na.rm = T) - mean(y0, na.rm = T)
var_diff_in_means <- var(y1, na.rm = T)/sum(!is.na(y1)) +
  var(y0, na.rm = T)/sum(!is.na(y0))
conf_int <- diff_in_means + c(-1,1)*1.96*sqrt(var_diff_in_means)
p_val <- 2*(1 - pnorm(abs(diff_in_means)/sqrt(var_diff_in_means)))
```

I estimate that the difference in means is 0.353, i.e. those given a partner for a vote swap were 35.3 percentage points more likely to vote for Clinton.

The normal approximation-based confidence interval for this difference in means lies between 0.173 and 0.532.

The two-sided, normal approximation-based p -value for this difference (given a null hypothesis of no difference) is 0.00012. This approximates the probability that, if there were no difference in means in the population, we would obtain a sample by iid random sampling in which the difference in means is as large or larger than the one we obtained.

For the confidence interval and p -value, I need the sampling distribution to be normal, which is true asymptotically under iid random sampling.

6. Scatterplot

Choose a dependent variable (Y) and an independent variable (X). Present a scatterplot with Y on the vertical axis and X on the horizontal axis. Both X or Y should be numerical variables, not categorical variables such as religion or race. It is helpful if either X or Y is a continuous

variable (i.e. one that takes on many values) rather than a discrete variable; if not, you may want to use `jitter()` to avoid plotting many points on top of each other. Comment on any apparent relationship between X and Y and why it might arise.

A major concern with this study is that, because the response rate is so low ($218/4500 \approx 4.8\%$), treated respondents may differ from control respondents (even if treatment was randomly assigned). This could happen e.g. if among treated subjects, those who followed through with their swap were more likely to reply, while a random sample of control subjects reply.

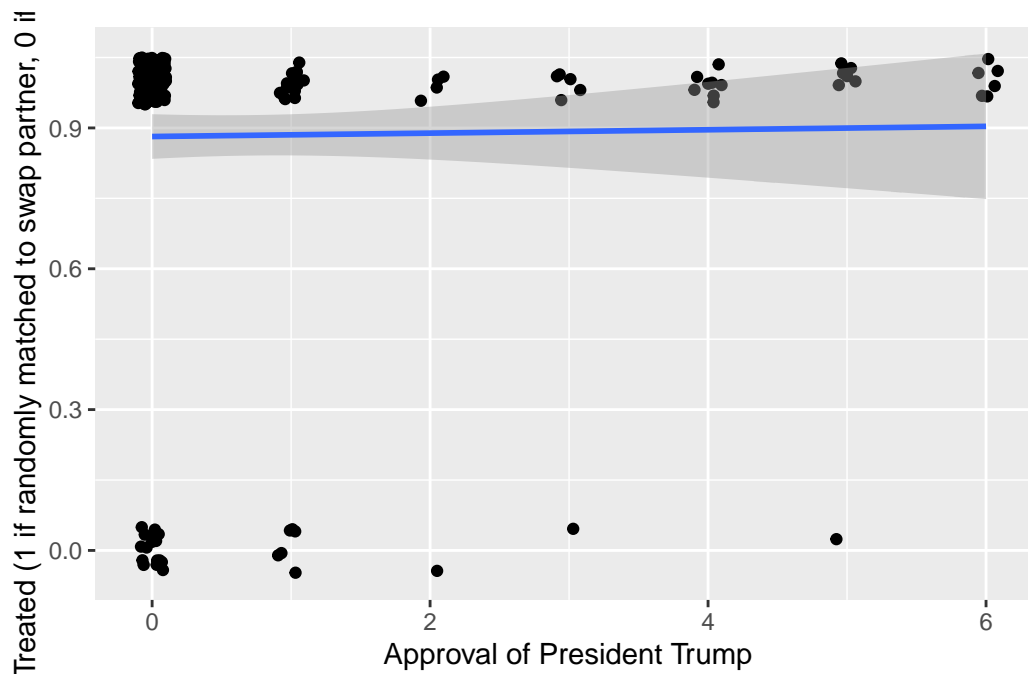
To assess this, we want to look for differences in respondent characteristics between those assigned to treatment and control. We should be concerned if these seem to be related. This scatterplot does so by looking at treatment status (Y) as a function of the respondent's approval for President Trump. (There might be relationship because treated subjects who voted for Clinton are likely more opposed to Trump, so if those respondents were more likely to reply we might see a higher probability of treatment among those with lower approval of Trump.)

The regression line on the scatterplot suggests that the probability of treatment is not strongly related to the respondent's approval of Trump, which is reassuring evidence that selection into response did not differ between the treatment and control group.

```
dat |>
  ggplot(aes(x = pres_approval, y = swap_treatment)) +
  geom_jitter(width = .1, height = .05) +
  geom_smooth(method = lm, formula = y ~ x) +
  labs(x = "Approval of President Trump",
       y = "Treated (1 if randomly matched to swap partner, 0 if not)")
```

Warning: Removed 4284 rows containing non-finite values (``stat_smooth()``).

Warning: Removed 4284 rows containing missing values (``geom_point()``).



7. Regression with one predictor

Regress Y on X to estimate the BLP of Y as a function of X . Interpret the slope coefficient. In a table, report a 95% confidence interval and p -value for the slope coefficient based on three approaches:

- classical standard errors as computed by `lm()`
- robust (Huber-White) standard errors as computed by `estimatr::lm_robust()`
- bootstrap standard errors (naïve bootstrap) State any assumptions you are making.

We will go back to the main research question: the dependent variable Y is whether the respondent voted for Clinton, and the independent variable X is whether the respondent was assigned a swap partner.

```
# classical
lm_reg <- lm(voted_HRC ~ swap_treatment, data = dat)
# robust
lmr_reg <- estimatr::lm_robust(voted_HRC ~ swap_treatment, data = dat)
# bootstrap
```

Approach to inference	Std. error	Conf. low	Conf. high	p
Classical	0.1018	0.1530	0.5521	0.0005
Robust	0.0915	0.1732	0.5319	0.0001
Bootstrap	0.0911	0.1740	0.5311	0.0001

```

m <- 1000
coefs <- rep(NA, length = m)
small_dat <- dat |> filter(!is.na(voted_HRC)) |> select(voted_HRC, swap_treatment)
for(i in 1:m){
  this_reg <- lm(voted_HRC ~ swap_treatment,
                 data = small_dat[sample(1:nrow(small_dat), replace = T),])
  coefs[i] <- coef(this_reg)[2]
}
tibble(`Approach to inference` = c("Classical", "Robust", "Bootstrap"),
       `Std. error` = c(sqrt(diag(vcov(lm_reg)))[2],
                        sqrt(diag(vcov(lmr_reg)))[2],
                        sd(coefs))) |>
mutate(`Conf. low` = coef(lm_reg)[2] - 1.96*`Std. error`,
       `Conf. high` = coef(lm_reg)[2] + 1.96*`Std. error`,
       p = 2*(1 - pnorm(abs(coef(lm_reg)[2]/`Std. error`)))) |>
kableExtra::kbl(digits = 4, booktabs = T) |>
kableExtra::kable_styling(full_width = F)

```

In all cases I am using a normal approximation to generate the confidence intervals and p -values. This means I need the sampling distribution of the regression coefficient to be approximately normal, which it is asymptotically given iid sampling and the “weak regularity conditions” (roughly, the statistics changes by a small amount if the CDF changes by a small amount). The classical standard errors (and therefore CI and p -value) also rely on the assumption of homoskedasticity.

For completeness (and because it takes 20 seconds to adapt the code!) I’ll also do this with the same X and Y as in my scatterplot.

```

# classical
lm_reg <- lm(swap_treatment ~ pres_approval, data = dat)
# robust
lmr_reg <- estimatr::lm_robust(swap_treatment ~ pres_approval, data = dat)
# bootstrap
m <- 1000
coefs <- rep(NA, length = m)
small_dat <- dat |> filter(!is.na(voted_HRC)) |> select(pres_approval, swap_treatment)

```

Approach to inference	Std. error	Conf. low	Conf. high	p
Classical	0.0143	-0.0244	0.0316	0.8008
Robust	0.0114	-0.0187	0.0259	0.7510
Bootstrap	0.0117	-0.0193	0.0265	0.7581

```
for(i in 1:m){
  this_reg <- lm(swap_treatment ~ pres_approval,
                 data = small_dat[sample(1:nrow(small_dat), replace = T),])
  coefs[i] <- coef(this_reg)[2]
}
tibble(`Approach to inference` = c("Classical", "Robust", "Bootstrap"),
       `Std. error` = c(sqrt(diag(vcov(lm_reg)))[2],
                        sqrt(diag(vcov(lmr_reg)))[2],
                        sd(coefs))) |>
mutate(`Conf. low` = coef(lm_reg)[2] - 1.96*`Std. error`,
       `Conf. high` = coef(lm_reg)[2] + 1.96*`Std. error`,
       p = 2*(1 - pnorm(abs(coef(lm_reg)[2]/`Std. error`)))) |>
kableExtra::kbl(digits = 4, booktabs = T) |>
kableExtra::kable_styling(full_width = F)
```

The same assumptions apply.

8. Confidence interval for a prediction

Based on the model above, what is your prediction for Y at the highest value of X in your dataset? Report a 95% confidence interval for that prediction, stating and justifying any assumptions. (Hint: you'll need the variance rule and/or the bootstrap.)

Because it's more interesting for this question, I'll use the question of how presidential approval relates to probability of treatment. Our task then is to get an estimate and confidence interval for the probability of treatment for someone whose presidential approval score is 6.

Here is that model:

```
pred_mod <- estimatr::lm_robust(swap_treatment ~ pres_approval, data = dat)
```

(I use the robust version to make it coincide with the bootstrap.)

First I get the prediction itself:

```
(prediction <- coef(pred_mod)[1] + 6*coef(pred_mod)[2])
```

```
(Intercept)  
0.9033108
```

Now I'll get the confidence interval two different ways.

First I use the variance rule. My goal is to compute $V[\hat{\beta}_0 + 6\hat{\beta}_1]$. By the variance rule, this is $V[\hat{\beta}_0] + 6^2V[\hat{\beta}_1] + 12\text{Cov}[\hat{\beta}_0, \hat{\beta}_1]$. Here I compute this:

```
the_vcov <- vcov(pred_mod)  
(variance_of_prediction <- the_vcov[1,1] + 36*the_vcov[2,2] + 12*the_vcov[1,2])
```

```
[1] 0.00394718
```

```
(confint_variance_rule <- prediction + 1.96*c(-1,1)*sqrt(variance_of_prediction))
```

```
[1] 0.7801707 1.0264509
```

Now I take a bootstrap approach instead.

```
m <- 1000  
preds <- rep(NA, length = m)  
for(i in 1:m){  
  this_reg <- lm(swap_treatment ~ pres_approval,  
                 data = small_dat[sample(1:nrow(small_dat), replace = T),])  
  preds[i] <- coef(this_reg)[1] + 6*coef(this_reg)[2]  
}  
(variance_of_predictions <- var(preds))
```

```
[1] 0.004070267
```

```
(confint_bootstrap <- prediction + 1.96*c(-1,1)*sqrt(variance_of_predictions))
```

```
[1] 0.7782654 1.0283561
```

My two confidence intervals agree closely, as they should, even though they use totally different procedures to get the standard errors.

In both cases I am using the normal approximation approach, which requires the sampling distribution of the prediction to be normal, which (given iid sampling and weak regularity conditions) holds asymptotically.

Note that another approach using the bootstrap is to obtain the .025 and .975 quantiles of the bootstrap distribution, like so:

```
quantile(preds, c(.025, .975))
```

```
      2.5%      97.5%  
0.7560708 1.0095633
```

In this case I am not assuming the normality of the sampling distribution, but I am assuming that the shape of the sampling distribution under the null is the same as the bootstrap distribution. The fact that this confidence interval agrees closely with the normal approximation based ones reflects the near-normality of the bootstrap sampling distribution.

9. Regression with more than one predictor

Again regress Y on X , but now include additional predictors other than X . Explain why you might want to include these additional predictors in the model. Fit at least two models with different sets of predictors or transformations of predictors. Report the results in a single regression table (using `modelsummary::modelsummary()` or a similar approach). Compare the coefficient on X across models.

For the main research question, we may want to include additional predictors to adjust for any differences between treatment and control that arose due to differential selection into survey response (i.e. differences between the kind of treated subjects who fill out the survey and the kind of control subjects who fill out the survey). One function of regression is to adjust for other differences between subjects when we seek to analyze the effect of a treatment variable, as you'll learn in causal inference.

In column 1 we present the bivariate relationship. In column 2 we add indicators for male and white. In column 3 we instead add an indicator for self-identified Democrat and third-degree polynomial of presidential approval; in column 4 we include both sets of covariates. The coefficient on "Treated" is slightly larger in the second two models. The fact that this coefficient doesn't change much across models is consistent with random assignment of treatment. The

coefficient on a variable should not change when you add other predictors that are independent of that variable.

```
mod1 <- estimatr::lm_robust(voted_HRC ~ swap_treatment, data = dat)
mod2 <- estimatr::lm_robust(voted_HRC ~ swap_treatment + white + female, data = dat)
mod3 <- estimatr::lm_robust(voted_HRC ~ swap_treatment +
  poly(pres_approval, 3) + dem,
  data = dat |> filter(!is.na(pres_approval)))
mod4 <- estimatr::lm_robust(voted_HRC ~ swap_treatment + white + female +
  poly(pres_approval, 3) + dem,
  data = dat |> filter(!is.na(pres_approval)))
modelsummary::modelsummary(list(mod1, mod2, mod3, mod4),
  gof_map = c("nobs", "r.squared"),
  stars = T,
  coef_map = c("(Intercept)" = "Intercept",
    "swap_treatment" = "Treated",
    "white" = "White",
    "female" = "Female",
    "dem" = "Democrat",
    "poly(pres_approval, 3)1" = "Pres approval",
    "poly(pres_approval, 3)2" = "Pres approval^2",
    "poly(pres_approval, 3)3" = "Pres approval^3"))
```

10. Interactions

Again regress Y on X , but this time include an interaction between X and a numerical predictor W . Tell us what W is and explain why the relationship between X and Y could plausibly depend on W . Report a confidence interval and p -value for the interaction term, stating any assumptions you use. Explain in words what the estimated interaction term means.

We want to know whether the effect of treatment depends on which third-party candidate the voter favored (Jill Stein or Gary Johnson). This is recorded in the **candidate** variable. It may be that voters who supported Jill Stein (the Green candidate) were more favorable to Clinton than voters who supported Gary Johnson (the Libertarian candidate). Also, supporters of Hillary Clinton in other states may have been more favorable to Stein (though it's possible that the website found matches who were okay with voting for the swing state voter's preferred third-party candidate). Both of these factors could make the effect of assigning a vote swap match different for these groups of swing state voters.

	(1)	(2)	(3)	(4)
Intercept	0.231** (0.084)	0.293* (0.126)	0.190* (0.085)	0.274* (0.125)
Treated	0.353*** (0.092)	0.351*** (0.090)	0.379*** (0.092)	0.374*** (0.090)
White		-0.040 (0.093)		-0.057 (0.089)
Female		-0.051 (0.067)		-0.063 (0.063)
Democrat			0.094 (0.076)	0.104 (0.077)
Pres approval			-2.538*** (0.358)	-2.559*** (0.348)
Pres approval ²			0.496 (0.459)	0.470 (0.450)
Pres approval ³			0.133 (0.388)	0.093 (0.386)
Num.Obs.	218	218	216	216
R2	0.053	0.056	0.201	0.206

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

	(1)	(2)
Intercept	0.231** (0.084)	0.176+ (0.095)
Treated	0.353*** (0.092)	0.315** (0.106)
Preferred candidate: Jill Stein		0.157 (0.192)
Interaction		0.070 (0.204)
Num.Obs.	218	218
R2	0.053	0.099
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001		

```

intx_mod <- estimatr::lm_robust(voted_HRC ~ swap_treatment*candidate,
                               data = dat |> filter(candidate != "evan mcmullin"))
stuff_to_extract <- summary(intx_mod)$coef["swap_treatment:candidatejill stein",
                                           c("Pr(>|t|)",
                                             "CI Lower",
                                             "CI Upper")]
modelsummary::modelsummary(list(mod1, intx_mod),
                             gof_map = c("nobs", "r.squared"),
                             stars = T,
                             coef_map = c("(Intercept)" = "Intercept",
                                           "swap_treatment" = "Treated",
                                           "candidatejill stein" =
                                             "Preferred candidate: Jill Stein",
                                           "swap_treatment:candidatejill stein" =
                                             "Interaction"))

```

The regression indicates that the effect is stronger for voters who prefer Jill Stein (the estimated effect is .385 for them compared to .315 for Gary Johnson supporters), but the difference is not significant. The confidence interval on the interaction goes from -0.333 to 0.472. The p -value on the test of the null hypothesis of no interaction is 0.733. This is an estimate of the probability that, if the true interaction were zero, we would obtain (through iid sampling) a sample with an interaction term as large or larger than the one we obtained.

The confidence interval and p -value rely on the normality of the sampling distribution for the interaction coefficient, which holds asymptotically given weak regularity conditions given that OLS is a plug-in estimator.