

Question 1

The following is a regression where the outcome is attitudes towards Hillary Clinton as measured by a standard feeling thermometer ratings (the variable ranges from 0 to 100 where 100 indicates high levels of support for Clinton). Researchers use two explanatory variables in their regression. First, they include a standard 7-point party identification measure that ranges from 1 (Strong Republican) to 7 (Strong Democrat). Second, they include a dummy variable (0 or 1) indicating whether the respondent lives in the South (the 13 states of the former Confederacy). The regression includes $N=100$ observations.

- Write a paragraph interpreting these results both substantively and statistically. For example, discuss both the statistical significance and size of the effect of each independent variable.
- The researchers claim that the regression indicates that partisanship *causes* changes in attitudes towards Clinton. That is, what assumptions would we need to make to interpret these results as causal? And do these assumptions seem reasonable in this situation? Give your answer in a paragraph.
- Calculate the F-statistic for this regression and interpret. You *do not* need to provide a precise p-value, but you *do* need to provide some indication that you understand what the F-statistic means (i.e., what it tells us about the model).

Outcome variable is thermometer rating of Clinton

Intercept	41.00 (4.24)
Party ID	6.09 (2.49)
South	-13.53 (7.06)
R-squared: 0.24	

Stand. errors are in parentheses.

$$F = \frac{0.24/2}{(1-.24)/[100-(2+1)]} = \frac{0.12}{(0.76)/97}$$

$$= \frac{0.12}{.0078} = 15.38$$

Question 2

Using the setup in problem 1 (but now using a new sample of $N=100$), we now fit a regression including an interaction between our two main explanatory variables.

- Explain both statistically and substantively the meaning of each of the parameters in the model.
- Write out the prediction equation for people from the south and another prediction equation for those not from the south. Explain why they are different and provide a substantive interpretation. This means that you need to tell me what this interaction model tells us about the political world.

Outcome variable is thermometer rating of Clinton

Intercept	36.00 (5.44)
Party ID	8.09 (2.49)
South	-10.53 (7.06)
South \times Party ID	-3.57 (0.88)
R-squared: 0.35	

Stand. errors are in parentheses.

$$\hat{Y}_i = b_0 + b_1 X_1 + \dots$$

$$\hat{Y}_i = 36.00 + 8.09 X_1 + (-10.53) X_2 + (-3.57) X_3$$

~~$$\hat{Y}_i = 36.00 + 8.09 X_1 + 2.49$$~~

~~$$\hat{Y}_i = 38.49 + 8.09 X_1$$~~

$$\hat{Y}_i = 36.00 + 8.09 X_1 + \epsilon_i$$

Question 4

Congressional Quarterly tracks the percentage of the time members of Congress vote with their party on partisan roll call votes (i.e., those in which a majority of Republicans oppose a majority of Democrats). The measure can take values from 0 to 100. If we take a simple random sample of 11 Democratic incumbents from the group described above and their mean party unity score is 93.87 with a sample standard deviation of 9.50, what is the 90% confidence interval for our estimate of mean party unity among this group as a whole? Please show your work!

$$\begin{aligned}\bar{X} &\pm Z \frac{SD}{\sqrt{n}} \\ 93.87 &\pm 2 \frac{9.50}{\sqrt{11}} \\ 93.87 &\pm 1.372 \left(\frac{9.50}{\sqrt{11}} \right) \\ 93.87 &\pm 1.372 \left(\frac{9.50}{3.32} \right) \\ 93.87 &\pm 1.372 (2.86) \\ 93.87 &\pm 3.92 \\ [89.92, 97.79]\end{aligned}$$

Question 5

A group of researchers are examining attitudes about the Affordable Care Act. They asked the following question to 781 respondents, "From what you've heard or read, do you approve or disapprove of the health care law that was enacted last year?"

Unfortunately, the researchers have had computer trouble, and they have only been able to retrieve the information presented in the following table. It contains partial information for each cell, including some observed counts, some expected frequencies (in parentheses), and some column and row totals.

$$\frac{(446 \times 421)}{781} = 240.42$$

	Response		Total
	Yes	No	
Democrats	221 (205.58)	225 (240.42)	446
Republicans	139 (154.42)	196 (180.58)	335
Total	360	421	781

4 columns

5 rows

$$(\text{Row total} \times \text{Column total}) / \text{Grand total} = \text{Expected Total}$$

- Use the information listed to complete the table. Be sure to calculate both the observed and expected frequencies for each cell.
- Calculate the cell component for the χ^2 statistic for the lower-right cell of the table (i.e., Republicans who responded No).
- The χ^2 statistic for this table is (approximately) 5.02. Specify and conduct a hypothesis test using this number.
- In one sentence, what does this table tell us about the relationship between party and attitudes on the Affordable Care Act?

Question 6

You are interested in studying the effect of political knowledge on partisan identification. Party ID is measured on a 7-point scale where 1 indicates a "Strong Republican" and a 7 indicates a "Strong Democrat." The randomly assigned treatment group for your study completed a short class on basic *civics* and the control group completed a short course on *art* appreciation. The data you get back is as follows:

	Civics Class	Art Class
Mean	3.8	3.5
SD	2.4	2.2
N	288	242

- Provide a point and interval estimate (use $\alpha = 0.95$) for the difference in Party ID for the treatment and control groups.
- Test the theory that the civics class changed Party ID.
- Is it OK to treat this estimate as causal? Why or why not?

$$df = 528$$

$$(3.8 - 3.5) \pm \frac{1.5(.975)}{2042} \sqrt{\frac{2.4^2}{288} + \frac{2.2^2}{242}}$$

$$= (0.3) \pm .71 \sqrt{.0083 + .0091}$$

$$= 0.3 \pm .71 \sqrt{.0174}$$

$$= 0.3 \pm .71 (.13)$$

$$= 0.3 \pm 0.092$$

QPM Final 2019

1. Clinton Attitudes

- a. The result suggests an average “temperature” of 41 for Clinton. The explanatory variables say that this number is positively affected by party identification. It also suggests this number heavily negatively affected by geography (whether or not someone is from the south). This says to me that if all people were voting along party lines, there are more democrats in the survey than republicans, and southerners are overwhelmingly republican.
- b. We would have to assume the observed participants are honest in their assessments. We would also need to assume that the subjective nature of the thermometer is an accurate enough representation and that the sample is random and sufficiently large (with $n=100$ this is no longer an assumption)
- c. $F = (.024/2) / ((1-.24)/(100-3)) = 15.38$ This number tells me there is a relatively large difference between the expected and observed outcomes. It also tells me that in a hypothesis test where the null is there is no relation between the variables and the outcome, we can reject that hypothesis.

2. Continuance from Problem 1

- a. The general “temperature” towards Clinton is now 36. There is a stronger positive causation between party identification and attitude towards Clinton. There is a strong negative causation for those from the south and this still results in a negative correlation when the two interact, suggesting that southern democrats are not fond of Clinton.

- b. See scanned scratch work for equation. The difference in the equations come from the error and whether or not it is present. Without the interaction of southern states we don't have the cancelling out of the error we would naturally see in the prediction equation. The two models tell us there is more uniformity across specific geographies than party identifications.

3. Causal Effect

- a. The causal effect that that when mailers were received, there is a drop in the overall rating of Clinton.
- b. We have to make the assumption that people were honest in their survey of their change of heart both before and after the received mailers. This can be violated by the very nature of political opinion. People afraid to give their opinion in the first place may be hesitant to share a change in opinion after new information is presented. So in this example, someone who changes their opinion about Clinton because of the mailer may be hesitant to admit they were favorable beforehand.

4. Democratic Party Unity

- a. $93.87 \pm t\text{-score} * sd/\sqrt{n} = 93.87 \pm 1.372(9.50/\sqrt{11}) = 93.87 \pm 3.92$
- b. We can be certain to 90% confidence that party unity in the democratic party will fall between [89.92 , 97.79]

5. Affordable Care Act Incomplete Table

- a.

	Responses		
	Yes	No	Total
Democrats	221	225	446

	(205.58)	(446*421)/781=240.42	
Republicans	360-221=139 (154.42)	335-139=196 (180.58)	335
Total	360	781-360=421	781

b. 196

c. Hypthothesis test of Independence

- i. Null: Party affiliation and response to the ACA are independent
- ii. Alternate: They are not independent
- iii. $DF = (2-1)(2-1)=1$
- iv. With a chi-squared of 5.02 we get a p-value of 0.025 which is underneath our 0.05 threshold. We can reject the null hypothesis


d. This table tells us that there is at least some correlation between party and attitude towards the ACA.

6. Civics Class


- a. See the scratch work for point and interval estimate of the difference
- b. The p-value is insignificantly small given the z score and degrees of freedom
- c. You can think of them as causal. With a null hypothesis that civics class does not affect party ID being rejected with the given numbers, we can be relatively confident that there is a correlation.

7. Question Seven

- a. A sample distribution is the distribution of a values for a specific sample of a population. The population distribution is the same thing, but with every single



element without a population accounted for. A sampling distribution is the probability distribution of statistics for the sample.

- 
- b. The GDP of a country is not as simple as the spending of all subjects within the entity. Ignoring that, there is the problem of inflation and deflation within a country's spending habits. German currency was so worthless at one point that it was counted by weight as opposed to denomination. The best way to fix this is to set up your regression over a relatively short period of time to negate inflation or to correct for inflation when you're setting up your linear model by making spending proportional to one's income. What proportion of income was spent during x amount of time?



8. Definitions

- a. P-value: The level of marginal significance within a statistical hypothesis, representing the probability of the occurrence of a given event. Basically, if a p-value is too low (generally under .05) then there is weak evidence in favor of the null hypothesis and we can reject it. If the p-value is high (over .05), then there is weak evidence to reject the null hypothesis.
- b. Outlier: It is an element of a dataset that distinctly stands out; something outside the 1st and 4th quartiles in a box and whisker plot.
- c. Instrument: An instrument variable is what is used to calculate (estimate) casual relationships

- d. Standard Error: The standard deviation of multiple sample statistics. When the sample size is sufficiently large it can be described as $SE = sd/\sqrt{n}$ where “n” is equal to the sample size.
- e. Autocorrelation: The degree of similarity between a given time series and a lagged version of itself.