Problem Set 5

Quantitative Political Methodology (L32 PS 363)

Instructions

1. Print out and write your WUSTL ID at the top of **each** page, and complete each exercise in the space allotted. You may attach extra pages if the space provided is not sufficient, but please indicate that you have done so below.

   Pages attached:_____

2. Please **show your work** if possible. You may lose points by simply writing in the answer. If the problem requires you to use `R`, please include the code you used to get your answers. If you are not sure if work needs to be shown for a particular problem, please ask a TA or post a question on Facebook.

3. The various pages of your homework should be **stapled together** (no paper clips please). If pages are lost because of a lack of a staple, no credit will be granted for that portion of the homework.

4. This problem set is **due at the beginning of class on Wednesday, November 8, 2017**. No late assignments will be accepted.

5. Total available points for this homework: 40.

Question 1 (Total: 10 points)

The following table was created using the data from a study run in a major Latin American city.[1] As part of the study, confederate made illegal left turns across traffic to draw the attention of the police officers. Two of the confederates were upper class drivers and two were lower class drivers. The researchers were interested in whether officers were more or less likely to solicit a bribe from drivers depending on their class (officers use phrases like, "We can solve this the easy way" to draw a bribe). The table below shows the resulting data.

	Not Stopped	Bribe requested	Stopped/given warning
Upper class	14	6	7
Lower class	7	7	1

(a) (3 points) Calculate the χ^2 test statistic by hand.

(b) (2 points) Now calculate the p-value (in R).[2] What do you conclude if $\alpha = .1$?

[1] Fried, Brian J, Paul Lagunes, and Atheendar Venkataramani. 2010. "Corruption and Inequality at the Crossroad: A Multimethod Study of Bribery and Discrimination in Latin America. *Latin American Research Review*. 45 (1): 76-97.

[2] Remember frequency should be > 5 for all cells, but let's calculate the pvalue here anyway.

(c) (3 points) Calculate the standardized residuals for each cell and put them in the table below.

	Not Stopped	Bribe requested	Stopped/given warning
Upper class			
Lower class			

(d) (2 points) How might the standardized residuals help you interpret the results?

Question 2 (Total: 12 points)

For the following question, we use data from Michael J. LaCour and Donald P. Green, "When contact changes minds: An experiment on transmission of support for gay equality" (2015). In this study, it was said that registered voters in Southern California were asked online to rate their opinion on the issue of same-sex marriage from 1 to 5, with 5 being the highest rate of approval. The voters were then said to be surveyed on their views 7 times (or in 7 waves). Between each wave, the same voters were said to receive visits in person by canvassers who conducted discussions with them regarding either same-sex marriage, recycling (an unrelated issue), or made no contact. The canvassers also indicated to the respondents their own sexual preference (gay or straight) during the discussion. You can obtain the data from `https://raw.githubusercontent.com/kosukeimai/qss/master/CAUSALITY/gay.csv`.[3]

(a) (3 points) In this exercise we will be comparing those respondents who received the treatment "No Contact", those who received the treatment "Same-Sex Marriage Script by Straight Canvasser", and those who received the treatment "Same-Sex Marriage Script by Gay Canvasser". We are only interested in the results after these treatments has been administered, so we will only be looking at wave 7. Create three new dataframes, one each for the respondents receiving these three treatments, **but make sure to subset the data so it only includes wave 7**. What is the mean value of `ssm` (the respondents' support for same-sex marriage on a scale from 1 to 5) for each of these three treatments?

(b) (3 points) We want to find if the two scripted treatments increased support for same-sex marriage compared to the "No Contact Treatment". State your null and alternative hypotheses. Will you need one- or two-tailed tests?

[3] As Imai notes, this study was later retracted due to allegations of fabricated data. Keep this in mind while analyzing the data. Another study (by those who discovered the initial study was likely fabricated) did find a reduction in antitransgender prejudice using a similar research design (Broockman and Kalla 2016).

(c) (3 points) Use the `t.test()` function to conduct a hypothesis test for $\alpha = 0.05$ that the "Same-Sex Marriage Script by Straight Canvasser" has a positive effect on support. Report your p-value as well as an interpretation of this test.

(d) (3 points) Use the `t.test()` function to conduct a hypothesis test for $\alpha = 0.05$ that the "Same-Sex Marriage Script by Gay Canvasser" has a positive effect on support. Report your p-value as well as an interpretation of this test.

Question 3 (Total: 8 points)

Chattopadhyay and Duflo were interested in studying the causal effect of having female politicians on policy outcomes.[4] Do women promote different policies than men? Answering this question with observational data is pretty difficult due to potential confounding problems (e.g. the districts that choose female politicians are likely to systematically differ in other aspects too). Hence, they exploit a randomised policy experiment in India, where since the mid-1990s, $\frac{1}{3}$ of village council heads have been **randomly** reserved for women. A subset of the data from West Bengal can be found at the following link: `https://raw.githubusercontent.com/kosukeimai/qss/master/PREDICTION/women.csv`

Each observation in the data set represents a village and there are two villages associated with one GP (i.e. a level of government called GP). The table below shows the names and descriptions of the variables in the "women.csv" dataset. They hypothesize that female politicians are more likely to support policies female voters want. Researchers found that more women complain about the quality of drinking water than men. Estimate the effect of the reservation policy on the number of new or repaired drinking water facilities in the villages.

Name	Description
`GP`	An identifier for the Gram Panchayat (GP)
`village`	identifier for each village
`reserved`	binary variable indicating whether the GP was reserved for women leaders or not
`female`	binary variable indicating whether the GP had a female leader or not
`irrigation`	variable measuring the number of new or repaired irrigation facilities in the village since the reserve policy started
`water`	variable measuring the number of new or repaired drinking-water facilities in the village since the reserve policy started

[4] Raghabendra Chattopadhyay and Esther Duflo. (2004). "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India. Econometrica, Vol. 72, No. 5, pp. 1409-1443.

(a) (2 points) State a null and alternative (two-tailed) hypothesis.

(b) (3 points) Run a bivariate regression to test this hypothesis in `R`. (Include your code)

(c) (3 points) Interpret the coefficient estimate for reservation policy.

Question 4 (Total: 10 points)

Researchers are interested in learning the effect of all of those yard signs on voting preferences.[5] Working with a campaign in Fairfax County, Virginia, 131 precincts were randomly divided into a treatment and control group. In 30 precincts, signs were posted around the precinct that read, "For Sale: Terry McAuliffe. Don't Sellout Virgina on November 5."

Below is the result of a regression with two variables and a constant. The dependent variable is the proportion of the vote that went to McAuliff's opponent Ken Cuccinelli. The first variable indicates whether a precinct was randomly assigned to have the sign against McAuliffe posted. The second variable indicates a precinct that was adjacent to a precinct in the treatment group (since people in those precincts might be exposed to the signs).

Impact of lawn signs on vote share

Precinct assigned lawn signs (n=30)	0.042
	(0.016)
Precint adjacent to lawn signs (n=76)	0.042
	(0.013)
Constant	0.302
	(0.011)
N	131
R^2	0.094

(a) (3 points) Use the results to determine whether having these yard signs in a precinct affect vote share (e.g., conduct a hypothesis test with $\alpha = .05$).

[5] Donald P. Green, Jonathan S. Krasno, Alexander Coppock, Benjamin D. Farrer, Brandon Lenoir, Joshua N. Zingher. 2016. "The effects of lawn signs on vote outcomes: Results from four randomized field experiments." Electoral Studies 41: 143-150.

(b) (3 points) Use the results to determine whether being next to precincts with these yard signs affect vote share (e.g., conduct a hypothesis test with $\alpha = .05$).

(c) (2 points) Interpret the coefficient for the constant term substantively.

(d) (2 points) Evaluate the model fit for this regression. What does this tell us about the importance of yard signs versus other factors?