Problem Set 2

Applied Stats/Quant Methods 1

Due: October 15, 2023

Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in `R`, please include the code you used to get your answers. Please also include the `.R` file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.

- Your homework should be submitted electronically on GitHub.

- This problem set is due before 23:59 on Sunday October 15, 2023. No late assignments will be accepted.

Question 1: Political Science

The following table was created using the data from a study run in a major Latin American city.[1] As part of the experimental treatment in the study, one employee of the research team was chosen to make illegal left turns across traffic to draw the attention of the police officers on shift. Two employee drivers were upper class, two were lower class drivers, and the identity of the driver was randomly assigned per encounter. 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.

---

[1]Fried, Lagunes, and 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.

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

(a) Calculate the χ^2 test statistic by hand/manually (even better if you can do "by hand" in R).

Answer The steps of calculating χ^2 is as follows:

```
1  #a
2  grandtotal<-14+6+7+7+7+1
3  row_total_1<-14+6+7
4  row_total_2<-7+7+1
5  col_total_1<-14+7
6  col_total_2<-6+7
7  col_total_3<-7+1
8  fe11<-row_total_1*col_total_1/grandtotal
9  fe12<-row_total_1*col_total_2/grandtotal
10 fe13<-row_total_1*col_total_3/grandtotal
11 fe21<-row_total_2*col_total_1/grandtotal
12 fe22<-row_total_2*col_total_2/grandtotal
13 fe23<-row_total_2*col_total_3/grandtotal
14 (14-fe11)^2/fe11
15 chi2<-((14-fe11)^2)/fe11+((6-fe12)^2)/fe12+((7-fe13)^2)/fe13+((7-fe21)^2)
       /fe21+((7-fe22)^2)/fe22+((1-fe23)^2)/fe23
```

(b) Now calculate the p-value from the test statistic you just created (in R).[2] What do you conclude if $\alpha = 0.1$?

---

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

Answer The steps of calculating p_value is as follows:

```
1  #b
2  df<-(2-1)*(3-1)
3  p_value<-round(pchisq(chi2,df,lower.tail = FALSE),4)
4  #p_value (= 0.1502) is bigger than alpha (=0.1). So, we don't have enough
5  #evidence to reject the null hypothesis that Officer are not likely to
6  #solicit a bribe from drivers depending on their class.
```

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

Answer The steps of calculating standardized residuals is as follows:

```
1  #c
2  z11<-round((14-fe11)/sqrt((1-row_total_1/grandtotal)*(1-col_total_1/
      grandtotal)),4)
3  z12<-round((6-fe12)/sqrt((1-row_total_1/grandtotal)*(1-col_total_2/
      grandtotal)),4)
4  z13<-round((7-fe13)/sqrt((1-row_total_1/grandtotal)*(1-col_total_3/
      grandtotal)),4)
5  z21<-round((7-fe21)/sqrt((1-row_total_2/grandtotal)*(1-col_total_1/
      grandtotal)),4)
6  z22<-round((7-fe22)/sqrt((1-row_total_2/grandtotal)*(1-col_total_2/
      grandtotal)),4)
7  z23<-round((1-fe23)/sqrt((1-row_total_2/grandtotal)*(1-col_total_3/
      grandtotal)),4)
```

The table is as follows:

	Not Stopped	Bribe requested	Stopped/given warning
Upper class	1.1832	-4.7467	3.4539
Lower class	-0.8819	3.538	-2.5744

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

Answer According to the concpet of standardized residuals, the more the observation get closer to the expectation, the values of standardized residual gets closer to 0. In the pciture, we can see all standardized residuals are relatively small, which mean the observation is very close to the expecttion, so we cannot reject the null hypothsis that class of driver don't have correlation with whether officer were more or less likely to solicit a bribe.

Question 2: Economics

Chattopadhyay and Duflo were interested in whether women promote different policies than men.[3] 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 randomized 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 is called "GP"). Figure 1 below shows the names and descriptions of the variables in the dataset. The authors 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. You need to estimate the effect of the reservation policy on the number of new or repaired drinking water facilities in the villages.

Figure 1: Names and description of variables from Chattopadhyay and Duflo (2004).

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

---

[3]Chattopadhyay and Duflo. (2004). "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India. *Econometrica*. 72 (5), 1409-1443.

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

Answer   Null Hypothsis: the corelation coefficient between the reservation policy and the number of new repaired drinking water facilities is equal to 0.

Alternative Hypothsis: the corelation coefficient between the reservation policy and the number of new repaired drinking water facilities isn't equal to 0.

(b) Run a bivariate regression to test this hypothesis in `R` (include your code!).

(c) Interpret the coefficient estimate for reservation policy.