Problem Set 2

Han Li

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`).

```r
class_bribery_matrix <- matrix(
  c(14,6,7,7,7,1),
  nrow = 2,
  ncol = 3,
  byrow = TRUE
)
#Observerd Matrix
colnames(class_bribery_matrix) <- c("NotStop", "Bribery", "Stopped")
rownames(class_bribery_matrix) <- c("UpperClass", "LowerClass")
Total<- cbind(class_bribery_matrix, RowTotal=rowSums(class_bribery_matrix))
Total<- rbind(Total,ColTotal= colSums(Total))

#Exp Matrix
Exp_UC<-Total['ColTotal',]*Total[1,'RowTotal']/Total['ColTotal','RowTotal']
Exp_LC <-Total['ColTotal',]*Total[2,'RowTotal']/Total['ColTotal','RowTotal']
Exp_Matrix <- rbind(Exp_UC,Exp_LC)[,c(1,2,3)]
#chi squared
chi2<-sum((class_bribery_matrix-Exp_Matrix)^2/Exp_Matrix)
sanity_check <- chisq.test(class_bribery_matrix)
```

```
> chi2
[1] 3.791168
> sanity_check$statistic
X-squared  3.791168
```

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

```
#pvalue (note we know there are 2 rows and 3 columns hence df =2)
pvalue<- pchisq(chi2,df=(2-1)*(3-1),lower.tail = FALSE)
```

```
> pvalue[
1] 0.1502306>
sanity_check$p.value
[1] 0.1502306
```

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

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

```r
#std. residual
ColP <-Total["ColTotal",]/Total["ColTotal","RowTotal"]
RowP <-Total[,"RowTotal"]/Total["ColTotal","RowTotal"]
SE_UC <- sqrt(Exp_Matrix["Exp_UC",]*(1-RowP[1])*(1-ColP[c(1,2,3)]))
SE_LC <- sqrt(Exp_Matrix["Exp_LC",]*(1-RowP[2])*(1-ColP[c(1,2,3)]))
SE_Matrix <- rbind(SE_UC,SE_LC)
z_matrix <- (class_bribery_matrix-Exp_Matrix)/ SE_Matrix
```

```
> z_matrix
                      NotStop   Bribery   Stopped
UpperClass 0.3220306 -1.641957  1.523026
LowerClass -0.3220306  1.641957 -1.523026
> sanity_check$stdres
                         NotStop   Bribery   Stopped
UpperClass  0.3220306 -1.641957  1.523026
LowerClass -0.3220306  1.641957 -1.523026
```

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

(d) How might the standardized residuals help you interpret the results?
The residuals shows us how exactily each cell contribute to the results and gives a

standardardized way to interpret the difference between each observed and expected value. In this case as each distance between the observed number and expected number are less than 2, we dont have engough evidence to reject the the null hypothesis that the bribery and stopping are independent from class of the driver.

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.
Null Hypothesis (H0): There is no association between reservation policy and the number of new or repaired drink water facilities in the village. Alternative Hypothesis(H1) : There is an associate between the reservation policy and the number of new or repaired water facilities in the village.

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

```
#q2
model<- lm(policy$water~policy$reserved)
summary(model)
```

```
Call:lm(formula = policy$water ~ policy$reserved)
Residuals:    Min      1Q  Median      3Q     Max
 -23.991 -14.738  -7.865   2.262 316.009
 Coefficients:                Estimate Std. Error t value Pr(>|t|)
  (Intercept)       14.738      2.286   6.446 4.22e-10 ***
  policy$reserved    9.252      3.948   2.344   0.0197 *
   ---Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
   Residual standard error: 33.45 on 320 degrees of freedom
   Multiple R-squared:  0.01688,
   Adjusted R-squared:  0.0138
   F-statistic: 5.493 on 1 and 320 DF,
    p-value: 0.0197
```

(c) Interpret the coefficient estimate for reservation policy.

The coeefcients are 14. 738(intercept) and 9.252 (reservation policy). It suggests on average without reservation policy there are about 14.738 new/ repaired water facilities per village. When there is reservation policy for female leader, there is on average an increment of about 9.252 new/repaired water facilities per village. The p-value is 0.0197 for the reservation policy,which is below alpha of 0.05. This suggest we have evidence in support of the alternative hypothesis that there is a positive association between the reservation policy and the number of new/repaired water facilities in the village.