# Problem Set 2

Applied Stats/Quant Methods 1

Due: October 15, 2023///WeiTang/23362496

## 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 Multi-method 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 $\chi^2$ test statistic by hand/manually (even better if you can do "by hand" inR).

```r
#####(1a)
rm(list=ls())
#build the data matrix
driver_class<-c("Upper class","Lower class")
reaction<-c("Not stopped","Bribe requested","Stopped/given warning")
main_table<-matrix(nrow=length(driver_class),ncol=3)
rownames(main_table)<-driver_class
colnames(main_table)<-reaction
main_table[,]<-c(14,7,6,7,7,1)

#chi^2_test by hand
prop.table(main_table)
#function fe() is built to calculate the estimate value
fe<-function(table,x,y){
  num<-table[x,y]
  row_sum<-sum(table[x,])
  col_sum<-sum(table[,y])
  grand_sum<-sum(table[,])
return(row_sum*col_sum/grand_sum)
}
fe(main_table,1,2)
#calculate the chi-square value
chi_square<-0
for(i in 1:length(reaction)){
  for(j in 1:length(driver_class)){
    chi_square<-chi_square+(((main_table[j,i]-fe(main_table,j,i))^2)/fe(main_table,j,i))
  }
}
print(chi_square)
```

```
My Answer:
console result
> print(chi_square)
[1] 3.791168

I get my answer of chi-square: 3.79
```

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

```
#####(1b)
#calculate p-value
p_value<-pchisq(chi_square,df<-(length(driver_class)-1)*(length(reaction)
    -1),lower.tail=FALSE)
print(p_value)
#check_my_answer
chisq.test(main_table)
```

```
My Answer:
console result
> print(p_value)
[1] 0.1502306

I get my answer of p-value: 0.15
Because that p-value is bigger than alpha, we fail to reject the null
hypothesis that officers were not more or less likely to solicit a bribe
from drivers depending on their class.
```

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

| | Not Stopped | Bribe requested | Stopped/given warning |
|---|---|---|---|
| Upper class | 0.3220306 | -1.641957 | 1.523026 |
| Lower class | -0.322030 | 1.641957 | -1.523026 |

```
#####(1c)
#build a blank matrix for standardized residual
std_residual_table<-matrix(nrow=length(driver_class),ncol=3)
rownames(std_residual_table)<-driver_class
colnames(std_residual_table)<-reaction
#for loop to fill in the corresponding standardized residuals
for(i in 1:length(driver_class)){
  for(j in 1:length(reaction)){
    std_residual_table[i,j]<-(main_table[i,j]-fe(main_table,i,j))/sqrt(fe
    (main_table,i,j)*(1-sum(main_table[i,])/sum(main_table[,]))*(1-sum(
    main_table[,j])/sum(main_table[,])))
    #for every element in the matrix, I use the formula to calculate the
    standardized residual.
```

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

```
  }
}
print(std_residual_table)
#check my answer
(chisq.test(main_table))$stdres
```

```
My Answer:

console result
> std_residual_table
            Not stopped   Bribe requested    Stopped/given warning
Upper class   0.3220306       -1.641957            1.523026
Lower class  -0.3220306        1.641957           -1.523026

This table shows the corresponding standardized residuals
of the elements in the table
```

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

```
My Answer:

(1.)the absolute values of standardized residual values of
Not Stopped are obviously lower than those of Bribe Requested
and Stopped/Given Warning, so the model fits Not Stopped
observations better.

(2.)the row cells of standardized residuals of the Upper class
and Lower class are symmetric, maybe all column sums are 0.

(3.)the absolute values of all standardized residual values
are lower than 2, so the model fits the data well in general.
```

## 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.

`My Answer:`

H0: reservation policy puts no effect on the number of new or repaired drinking water facilities.($\beta = 0$)

Ha: reservation policy puts effect on the number of new or repaired drinking water facilities.($\beta \neq 0$)

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

```
#####(2b)
#import the dataset and build the data frame
data_set <- read.csv("https://raw.githubusercontent.com/kosukeimai/qss/
    master/PREDICTION/women.csv", header=T)
str(data_set)
#run bivariate regression
temp_model<-lm(water~reserved,data_set)
summary(temp_model)$coefficients
```

`My Answer:`

```
console result
> summary(temp_model)$coefficients
             Estimate Std. Error  t value     Pr(>|t|)
(Intercept) 14.738318   2.286300 6.446363 4.216474e-10
reserved     9.252423   3.947746 2.343723 1.970398e-02
```

`According to the result, we get the estimate formula :`

$$water = 14.738 + 9.252 * reserved$$

`Conclusion:`
`Because that we calculated that estimated coefficient (slope of the estimated function) is 9.25 and corresponding p-value is 0.0197 which is much less than α=0.05, we find evidence that we should reject the null hypothesis, and we find that the evidence supports the alternative hypothesis that reservation policy puts effect on the number of new or repaired drinking water facilities.`

(c) Interpret the coefficient estimate for reservation policy.

`My Answer:`

`We get the expected value of water both when reserved is 1 or 0.`

```
(1.)When reserved = 0:
expected water = 14.738+9.252*0 = 14.738

This means that when reserved = 0 (GP was not reserved for women leader),
the average value of the numbers of new or repaired drinking-water
facilities in the village since the reserved policy started is 14.738.

(2.)When reserved = 1:
expected water = 14.738+9.252*1 = 23.990

This means that when reserved = 1 (GP was reserved for women leader),
the average value of the numbers of new or repaired drinking-water
facilities in the village since the reserved policy started is 23.990.
```