

# Problem Set 2

## Applied Stats/Quant Methods 1

Due: October 16, 2022

### 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 16, 2022. No late assignments will be accepted.
- Total available points for this homework is 80.

### Question 1 (40 points): Political Science

The following table was created using the data from a study run in a major Latin American city.<sup>1</sup> 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.

---

<sup>1</sup>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" in R).

```
foUpper1 ← 14
foUpper2 ← 6
foUpper3 ← 7
foLower1 ← 7
foLower2 ← 7
foLower3 ← 1
Column1 ← sum(foUpper1, foLower1)
Column2 ← sum(foUpper2, foLower2)
Column3 ← sum(foUpper3, foLower3)
Row1 ← sum(foUpper1, foUpper2, foUpper3),
Row2 ← sum(foLower1, foLower2, foLower3)
Grand_Total := sum(Row1, Row2)
f1e ← (Row1 / Grand_Total) * Column1
f2e ← (Row1 / Grand_Total) * Column2
f3e ← (Row1 / Grand_Total) * Column3
f4e ← (Row2 / Grand_Total) * Column1
f5e ← (Row2 / Grand_Total) * Column2
f6e ← (Row2 / Grand_Total) * Column3

x2_test_statistic = (((foUpper1-f1e)^2)/f1e)+(((foUpper2-f2e)^2)/f2e)+(((foUpper3-f3e)^2)/f3e)+(((foLower1-f4e)^2)/f4e)+(((foLower2-f5e)^2)/f5e)+(((foLower3-f6e)^2)/f6e))
```

$x^2\_test\_statistic = 3.79$

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

```
P_Value <- pchisq(x2_test_statistic, df = 2, lower.tail = F)
```

```
P_Value <- as.numeric(format(round(P_Value, 2), nsmall = 2))
```

```
P_Value = 0.15
```

Because the P-value  $> 0.1$ , we can't reject the null hypothesis that social class and being asked for a bribe are statistically independent.

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

	Not Stopped	Bribe requested	Stopped/given warning
Upper class	0.32	-1.64	1.52
Lower class	-0.32	1.64	-1.52

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

Because the standardised residuals are either below and above 2 and -2 respectively, we can interpret that the variables are likely to be independent. Put otherwise, all residual values are less than the expected frequency.

---

<sup>2</sup>Remember frequency should be  $> 5$  for all cells, but let's calculate the p-value here anyway.

## Question 2 (40 points): Economics

Chattopadhyay and Duflo were interested in whether women promote different policies than men.<sup>3</sup> 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).

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

H0: Female politicians support new or repaired drinking water facilities in the villages at the same rate as male politicians.

Ha: Female politicians support new or repaired drinking water facilities in the villages more or less than males politicians.

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

```
Regression_analysis_water_and_reserved <- lm(formula = water ~ reserved, data=Indian_data)
```

```
Regression_analysis_water_and_reserved =  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 = \hat{y} = 14.738 + 9.252(reservedseats)$ 
```

---

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

(c) Interpret the coefficient estimate for reservation policy.

The regression equation of  $14.738 + 9.252(\text{reserved seats})$  indicates that for each additional district with reserved seats, there are 9.252 new or repaired drinking water facilities, indicating the reserve policy has positive impacts on drinking water facilities.

This tentatively, although somewhat indirectly, supports Chattopadhyay and Duflo's hypothesis that female politicians are more likely to support policies female voters want.