

# 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).  $\chi^2 = 3.77$ .

R Code

Expected Values for each cell

```
(27 * 21) / 42
(27 * 13) / 42
(27 * 8) / 42
(15 * 21) / 42
(15 * 13) / 42
(15 * 8) / 42
```

T-Stats for each cell

```
(14 - 13.5)^2 / 13.5
(6 - 8.35)^2 / 8.35
(7 - 5.14)^2 / 5.14
(7 - 7.5)^2 / 7.5
(7 - 4.64)^2 / 4.64
(1 - 2.85)^2 / 2.85
```

Chi-Square

```
1.2 + 0.03 + 1.2 + 0.66 + 0.01 + 0.67
```

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

$P = 0.15$ . This does not indicate statistically reliable differences in bribe solicitation toward upper and lower class drivers.

R Code

Degrees of Freedom

```
(3 - 1) * (2 - 1)
```

P-Value

```
pchisq(3.77, df=2, lower.tail=F)
```

---

<sup>2</sup>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.

|             | Not Stopped | Bribe requested | Stopped/given warning |
|-------------|-------------|-----------------|-----------------------|
| Upper class | 0.02        | -0.13           | 0.13                  |
| Lower class | -0.02       | 0.18            | -0.18                 |

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

The standardized residuals give a standardized measure of the difference between the observed and predicted values, which is adjusted for sample size. Positive residuals indicate the cell frequency was higher than expected and negative indicates it was lower than expected.

As the difference between our observed and predicted values is not statistically significant here, our residuals are all very small. Were the difference between our observed and predicted value significant, the standardized residuals would give a useful indication of which cells were contributing most to that finding.

## 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 ?? 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.

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

Null Hypothesis: There is no relationship between village council heads being reserved for women and drinking water facilities being repaired or newly installed.

Alternative Hypothesis: There is a relationship between village council heads being reserved for women and drinking water facilities being repaired or newly installed.

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

R Code

```
str(women)
lm (women$water ~ women$reserved) -> lm1
summary(lm1)
lmplot (lm1)
```

---

<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.

At alpha 0.5, there is a statistically reliable ( $p = 0.0197$ ) and positive relationship between whether the council head is reserved for women and drinking water facilities being repaired or installed.

At the 95% confidence level, where the council head is reserved for women the number drinking water facilities that are repaired or newly installed increases by as few as 1.48 and as many 17.01. The mean estimate increase in repaired or installed facilities where the council heads are women is 9.25.

R Code

```
95\% Confidence Interval for Beta 1  
confint(lm1)
```