

Problem Set 2

Applied Stats/Quant Methods 1

Jianxiong Wu—23354731

Due: October 14, 2024

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 Monday October 14, 2024. 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.

¹Fried, 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 χ^2 test statistic by hand/manually (even better if you can do "by hand" in R).

```

1 resulting_data_table <- matrix(c(14, 6, 7, 7, 7, 1), nrow = 2, byrow =
  TRUE)
2 resulting_data_table
3 row_total <- rowSums(resulting_data_table)
4 col_total <- colSums(resulting_data_table)
5 grand_total <- sum(resulting_data_table)
6
7 row_total
8 col_total
9 grand_total
10
11 fe <- (row_total/grand_total) %*% t(col_total)
12 fe
13
14 chi_square <- sum((resulting_data_table - fe)^2/fe)
15 chi_square

```

Results: 3.791168

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

```

1 p_values <- pchisq(chi_square, df = (2-1)*(3-1), lower.tail = FALSE)
2 p_values

```

Results:

The p-value is 0.1502306, greater than 0.1, so there is not enough evidence to reject the null hypothesis, and we cannot conclude from these data that officers were more or less likely to solicit a bribe from drivers depending on their class.

²Remember 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	0.3220306	-1.641957	1.523026
Lower class	-0.3220306	1.641957	-1.523026

```

1 standardized_residuals <- ((resulting_data_table - fe) /
2                             (sqrt(fe * (1-(row_total/grand_total)) %*%
3                             (1-t(col_total/grand_total))))))
head(standardized_residuals)

```

Results:

```

      [,1]      [,2]      [,3]
[1,] 0.3220306 -1.641957 1.523026
[2,] -0.3220306 1.641957 -1.523026

```

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

Results:

Since the absolute value of the standardised residuals in each cell does not exceed 2, indicating that the observed wisdom is not significantly different from the expected value, there is no significant effect of the driver's class on whether the police officer solicits bribes from the driver. So the results can be interpreted as: there is no evidence of whether officers were more or less likely to solicit a bribe from drivers depending on their class.

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

³Chattopadhyay 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.

Results:

Ho(null hypothesis):The reservation policy has no effect on the number of new or repaired drinking water facilities in the villages.

H1(alternative hypothesis):The reservation policy has effect on the number of new or repaired drinking water facilities in the villages.

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

```
1 women_data <- read.csv("https://raw.githubusercontent.com/kosukeimai/qss/master/PREDICTION/women.csv")
2 model <- lm(water ~ reserved, data = women_data)
3 summary(model)
```

Results:

Call:

```
lm(formula = water ~ reserved, data = women_data)
```

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 ***
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

Since the p-value = 0.0197 is less than 0.05, the null hypothesis is rejected, so the reservation policy has effect on the number of new or repaired drinking water facilities in the villages.

(c) Interpret the coefficient estimate for reservation policy.

Results:

The coefficient estimate for the reservation policy is 9.252, indicating that the reservation policy is positively correlated with the number of new or repaired drinking water facilities in the villages. As a result, villages with the reservation policies had an average of 9.252 more new or repaired drinking water facilities than villages without reservation policies.