Problem Set 2

Eleni Karagianni

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.

[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
# First, I create the table in R.
table <- matrix(c(14, 6, 7, 7, 7, 1), nrow = 2, byrow = TRUE)

# Formula: (Row total/Grand total) X Column total
# I calculate the different components of this formula as follows"
row_total_upper <- sum(table[1, ])  # 14 + 6 + 7 = 27
row_total_lower <- sum(table[2, ])  # 7 + 7 + 1 = 15

column_total_notstopped <- sum(table[, 1])  # 14 + 7 = 21
column_total_bribe <- sum(table[, 2])  # 6 + 7 = 13
column_total_warning <- sum(table[, 3])  # 7 + 1 = 8

# Grand total (sum of all frequencies)
grand_total <- sum(table)  # 14 + 6 + 7 + 7 + 7 + 1 = 42

# And now I calculate the expected frequencies ($f_e$)
fe_notstopped_upper <- (row_total_upper / grand_total) * column_total_
    notstopped
fe_notstopped_lower <- (row_total_lower / grand_total) * column_total_
    notstopped

fe_bribe_upper <- (row_total_upper / grand_total) * column_total_bribe
fe_bribe_lower <- (row_total_lower / grand_total) * column_total_bribe

fe_warning_upper <- (row_total_upper / grand_total) * column_total_
    warning
fe_warning_lower <- (row_total_lower / grand_total) * column_total_
    warning

# And the chi-square 'manually'
chi2 <- ((((14 - fe_notstopped_upper) ^ 2) / fe_notstopped_upper) + (((7
    - fe_notstopped_lower) ^ 2) / fe_notstopped_lower)
+ (((6 - fe_bribe_upper) ^ 2) / fe_bribe_upper) + (((7 - fe_bribe_lower)
    ^ 2) / fe_bribe_lower) +
  (((7 - fe_warning_upper) ^ 2) / fe_warning_upper) + (((1 - fe_warning_
    lower) ^ 2) / fe_warning_lower))

# ...and alternatively:
chi2_alt <- sum(((table - matrix(c(fe_notstopped_upper, fe_notstopped_
    lower,
                                   fe_bribe_upper, fe_bribe_lower,
```

```
                                        fe_warning_upper, fe_warning_lower),
                                      nrow = 2, byrow = TRUE))^2) /
              matrix(c(fe_notstopped_upper, fe_notstopped_lower,
                       fe_bribe_upper, fe_bribe_lower,
                       fe_warning_upper, fe_warning_lower), nrow = 2,
    byrow = TRUE))

chi2 # 3.791168
chi2_alt # same
```

The χ^2 is 3.791168.

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

```
# Degrees of freedom = (number of rows - 1) * (number of columns - 1)
df <- (2 - 1) * (3 - 1)
p_value <- pchisq(chi2, df = df, lower.tail = F)
p_value
```

The p-value is 0.1502306 which is smaller than 0.10 (the α level of confidence requested), therefore we do not have enough evidence to reject the null hypothesis.

[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.

	Not Stopped	Bribe requested	Stopped/given warning
Upper class	0.1360828	-0.8153742	0.818923
Lower class	-0.1825742	1.0939393	-1.098701

```
chi2_test <- chisq.test(table)
chi2_test$residuals
```

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

The standardized residuals measure the strength of the difference between observed and expected values in our table. We can see how much each value of the original table contributes to the χ^2 value. The closer the value to 0, the less the difference between the expected and the observed value. Since the standardized residuals are not larger than +/-2 or +/-3, no extreme values are present in the table. Therefore, there is no strong evidence of a relationship between class (upper, lower) and the likelihood of being stopped, being requested a bribe, or given a warning.

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.
H0 (null hypothesis): The reservation policy had no effect on the number of new or repaired drinking water facilities.
Ha (alternative hypothesis): The researvation policy had an effect on the number of new or repaired drinking water facilities.

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

```
biv_reg <- lm(water ~ reserved, data = data)
summary(biv_reg)
```

(c) Interpret the coefficient estimate for reservation policy.

Table 1: Coefficient Table

	Dependent variable:
	New/Repaired Working facilities
Reserved Status	9.252**
	(3.948)
Constant	14.738***
	(2.286)
Observations	322
R^2	0.017
Adjusted R^2	0.014
Residual Std. Error	33.446 (df = 320)
F Statistic	5.493** (df = 1; 320)
Note:	*p<0.1; **p<0.05; ***p<0.01

The reservation policy led, on average, to an increase of approximately 9 new or repaired drinking water facilities in reserved villages compared to non-reserved ones. The result is statistically significant at the 5% confidence level ($p < 0.05$). Therefore, we can reject the null hypothesis and claim that there is a positive impact of the women politicians on the improvement of drinking water facilities.