Problem Set 2

Vismante Dringelyte (173413781

Due: October 15, 2023

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 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).

First, I create a data frame from the table above:

```
PolSciTab <- matrix(c(14, 6, 7, 7, 7, 1), ncol = 3, byrow = TRUE)
colnames(PolSciTab) <- c('Not stopped', 'Bribe requested', 'Stopped/given
    warning')
rownames(PolSciTab) <- c('Upper class', 'Lower class')
PolSciTab <- as.table(PolSciTab)
PolSciTab
```

I calculated the χ^2 test statistic "by hand" in R. First, I calculated the row, column and grand totals.

```
# Calculating row and column totals
row_tot <- rowSums(PolSciTab)
col_tot <- colSums(PolSciTab)

# calculating grand total
grand_tot <- sum(PolSciTab)
```

Which you can see in this table:

	Not Stopped	Bribe requested	Stopped/given warning	Total
Upper class	14	6	7	27
Lower class	7	7	1	15
Total	21	13	8	42

Then, the expected values:

```
row_tot
col_tot

exp_val <- matrix(0, nrow = nrow(PolSciTab), ncol = ncol(PolSciTab))

for ( i in 1:nrow(PolSciTab)){
  for (j in 1:ncol(PolSciTab)) {
    exp_val[i, j] <- (rowSums(PolSciTab)[i] * colSums(PolSciTab)[j]) /
    grand_tot
  }
}
```

	Not Stopped	Bribe requested	Stopped/given warning
Upper class	13.5	8.4	5.1
Lower class	7.5	4.6	2.9

I tabulated these as well.

Finally, calculating the χ^2 test statistic.

```
chi_sq <- sum((PolSciTab - ex_val)^2 / ex_val)
chi_sq
```

This returns a value of 3.79.

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

```
deg_f <- (nrow(PolSciTab) -1) * (ncol(PolSciTab) -1)
```

Which returned a value of 2.

Then, used this to calculate the P-value.

```
p_val <- pchisq(chi_sq, df = 2, lower.tail = FALSE)
```

This returned a value of 0.1502

And I checked my results using `chisq.test()`.

```
chi_square <- chisq.test(PolSciTab)
```

My results, both for the χ^2 test statistic and the P-value, seem to match.

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

	Not Stopped	Bribe requested	Stopped/given warning
Upper class	0.322	-1.642	1.523
Lower class	-0.322	1.642	-1.523

I got these results by calculating standardised residuals by hand, using the `for` function.

I first turned the data into a matrix.

```
std_res <- matrix(0, nrow = nrow(PolSciTab), ncol = ncol(PolSciTab))
```

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

I then calculated row and column proportions.

```
row_proportions <- rowSums(PolSciTab) / sum(PolSciTab)
column_proportions <- colSums(PolSciTab) / sum(PolSciTab)
```

And, finally, used these objects in my function to find the standardised residuals.

```
for (i in 1:nrow(PolSciTab)) {
  for (j in 1:ncol(PolSciTab)) {
    observed_value <- PolSciTab[i, j]
    expected_value <- exp_val[i, j]
    row_proportion_i <- row_proportions[i]
    column_proportion_j <- column_proportions[j]

    std_res[i, j] <- (observed_value - expected_value) / sqrt(expected_
    value * (1 - row_proportion_i) * (1 - column_proportion_j))
  }
}
```

This returned the values shown in the table above.

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

Agresti and Finlay claim that "values below -3 and above +3 ... are very convincing evidence of a true effect in that cell". However, none of these values do, so we have further evidence that these variables are independent. Being rich or poor did not have a significant correlation to being asked for a bribe in this sample.

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.

Null hypothesis: The reservation policy does not affect the number of new or repaired drinking water facilities in the villages.

Alternative hypothesis: The reservation policy affects 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!).

First, I subset the variables "reserved" and "water" from the data set.

```
EconRW <- Econ[c("reserved", "water")]
```

Then, I calculated the slope (β) and the intercept (α) by hand.

```
beta <- sum((EconRW$water - mean(EconRW$water)) * (EconRW$reserved - mean
    (
  EconRW$reserved)))/
  sum((EconRW$reserved - mean(EconRW$reserved))^2)
beta
alpha <- mean(EconRW$water) - beta*mean(EconRW$reserved)
alpha
```

This returned the values $\beta = 9.25$ and $\alpha = 14.74$

I checked my results:

```
reg1 <- lm(EconRW$water~EconRW$reserved, data = EconRW)
```

They seemed to be correct.

Calculating the standard deviation:

```
sd_est <- sqrt( sum(resid(reg1)^2) / (dim(EconRW)[1]-2))
sd_est

# or, simply

sigma(reg1)
```

This gave me the value 33.45.

The standard errors for β and α

```
beta_se <- sd_est / sqrt(sum((EconRW$reserved - mean(EconRW$reserved))^2)
    )
```

```
alpha_se <- sd_est * sqrt((1/dim(EconRW)[1]) + (mean(EconRW$reserved)^2/
    sum(
  (EconRW$reserved-mean(EconRW$reserved))^2)))
```

The standard error was 3.95 for β and 2.29 for α

Calculating the Test statistic and P-value:

```
2 * pt((beta-0)/beta_se, dim(EconRW) [1]-2, lower.tail =F)
2 * pt((alpha-0)/alpha_se, dim(EconRW) [1]-2, lower.tail =F)
```

```
[1] 0.01970398
[1] 4.216474e-10
```

I checked my results:

```
model <- summary(lm(EconRW$water~EconRW$reserved, data = EconRW))
```

And they appear to match.

```
Call:
lm(formula = EconRW$water ~ EconRW$reserved, data = EconRW)

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 ***
EconRW$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
```

(c) Interpret the coefficient estimate for reservation policy.

My interpretation if these results is that we can reject the null hypothesis, as the P-value is lower than 0.05. This suggests that villages which have reserved seats for female politicians are more likely to have new or repaired water facilities.