Applied Stats/Quants Methods 1

Olívia Freitas

October, 2023

Instrutions

- *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 le 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 trac to draw the attention of the police ocers 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 ocers were more or less likely to solicit a bribe from drivers depending on their class (ocers use phrases like, can solve this the easy way" to draw a bribe). The table below shows the resulting data.

a) Calculate the X2 test statistic by hand/manually (even better if you can do "by hand" in R).

```
> # We will first create a data frame with our data
> data <- data.frame(
+   Class = c("Upper class", "Lower class"),
+   `Not Stopped` = c(14, 7),
+   `Bribe requested` = c(6, 7),
+   `Stopped/given warning` = c(7, 1)
+ )
> # Now we will calculate row and column totals
> row_totals <- rowSums(data[, -1])
> col_totals <- colSums(data[, -1])
>
> # We will calculate then the grand total
> grand_total <- sum(row_totals)
>
> # We will also calculate the expected frequencies for each cell
> expected_values <- outer(row_totals, col_totals) / grand_total
>
> # Now it is time to clculate the χ² value for each cell
> chi_squared <- sum((data[, -1] - expected_values)^2 / expected_values)
>
> # We will determine the degrees of freedom
> df <- (nrow(data) - 1) * (ncol(data) - 1)
>
> # Finally, we will [rint the Chi-squared statistic
> cat("Chi-squared statistic:", chi_squared, "\n")
Chi-squared statistic: 3.791168
>
```

b) Now calculate the p-value from the test statistic you just created (in R).2 What do you conclude if $= 0{:}1$?).

```
# Let's calculate the p-value and run the chi-squared test
p_value <- 1 - pchisq(chi_squared, df)

# Then print the p-value
cat("Chi-squared p-value:", p_value, "\n")

## With our p-value 0.2849151 result we can not reject the null hypothesis because
##it suggests that there is no significant relationship between the social class of
##drives and the solicitation of a bribe
```

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

```
# We need now to calculate the expected frequencies
expected_values <- outer(row_totals, col_totals) / grand_total


standardized_residuals <- (data[, -1] - expected_values) / sqrt(expected_values)

# We need to print the standardized residuals table
standardized_residuals_table <- cbind(data[, 1], standardized_residuals)
colnames(standardized_residuals_table) <- c("Class", "Not Stopped", "Bribe requested",
"Stopped/given warning")

standardized_residuals_table
```

```
> standardized_residuals_table
        Class Not Stopped Bribe requested Stopped/given warning
1 Upper class   0.1360828      -0.8153742              0.818923
2 Lower class  -0.1825742       1.0939393             -1.098701
```

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

Standardized residuals detect significant deviations from expected values, helping identify cells that lead to associations and their practical significance.

Question 2: Economics

Chattopadhyay and Du o were interested in whether women promote dierent policies than men.3 Answering this question with observational data is pretty dicult due to potential confounding problems (e.g. the districts that choose female politicians are likely to system- atically dier in other aspects too). Hence, they exploit a randomized policy experiment in India, where since the mid-1990s, 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 eect 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 Duo (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

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

Null: The reservation policy for women has no effect on the number of new or repaired drinking water facilities in the villages.

H1: The reservation policy for women has an effect on the number of new or repaired drinking water facilities in the villages.

Now I will load the data into R

data ¡- read.csv(”https://raw.githubusercontent.com/kosukeimai/qss/master/PREDICTION/women.csv”)

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

```
# Now we will run a bivariate regression
model <- lm(water ~ reserved, data = data)

# We will now print the summary of the regression
summary(model)
```

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

c) Interpret the coecient estimate for reservation policy.

The positive coefficient of ”9.252” indicates that villages with policies reserving positions for women tend to have more new or repaired drinking water facilities compared to villages without this policy. The data demonstrate a statistically significant positive relationship between the policy of reserving positions for women as village council heads and the number of new or repaired drinking water facilities in the villages.