Applied Stats/Quant Methods 1: Problem Set 2

Jamie McClenaghan

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
Upper_class_total <- 14 + 6 + 7
Lower_class_total <- 7 + 7 + 1
Grand_total <- Upper_class_total + Lower_class_total

Not_Stopped_total <- 14 + 7
Bribe_requested_total <- 6 + 7
Stopped_givenwarning_total <- 7 + 1

oneone_o <- 14
oneone_e <- (Upper_class_total / Grand_total) * Not_Stopped_total

onetwo_o <- 6
onetwo_e <- (Upper_class_total / Grand_total) * Bribe_requested_total

onethree_o <- 7
onethree_e <- (Upper_class_total / Grand_total) * Stopped_givenwarning_total

twoone_o <- 7
twoone_e <- (Lower_class_total / Grand_total) * Not_Stopped_total

twotwo_o <- 7
twotwo_e <- (Lower_class_total / Grand_total) * Bribe_requested_total

threethree_o <- 1
threethree_e <- (Lower_class_total / Grand_total) * Stopped_givenwarning_total

X2 <- ((oneone_o - oneone_e)^2 / oneone_e) + ((onetwo_o - onetwo_e)^2 / onetwo_e) + ((onethree_o - onethree_e)^2 / onethree_e) + ((twoone_o - twoone_e)^2 / twoone_e) + ((twotwo_o - twotwo_e)^2 / twotwo_e) + ((threethree_o - threethree_e)^2 / threethree_e) #implementing R code to execute formula for X2
```

I have calculated the χ^2 as 3.791168 utilizing the above code.

(b) Now calculate the p-value from the test statistic you just created (in R).[2] What do you

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

conclude if $\alpha = 0.1$?

```
df = (2 - 1) * (3 - 2) #calculating degrees of freedom

pchisq(X2, df = 2, lower.tail=FALSE) #calculating p-value from test
    statistic
```

I have calculated the p-value as 0.1502306.

(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	0.8189230
Lower class	-0.1825742	1.0939393	-1.0987005

```
chi2 <- chisq.test(final) #chi square test with aid of R function
chi2$residuals #calculating residuals
```

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

The standardized residuals help us to identify any outliers in the results as they give us a value which we can attribute to the difference between expected and observed results in a standardized form after taking into account their raw values.

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.

The *null hypothesis* states that the reservation policy has no effect on the number of new or repaired drinking water facilities in the villages.

The *alternative (two-tailed) hypothesis* states that the reservation policy has an 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!).

```r
reg <- lm(water~reserved,data = data) #bivariate regression
```

(c) Interpret the coefficient estimate for reservation policy:
When looking at the regression table, we can see that, on average, for a unit increase in the reserved variable (given that it is binary) there is an increase of 9.25 units for the water variable. Furthermore, this is statistically significant as $p < 0.05$ and we are in a position to reject the null hypothesis.

	Model 1
(Intercept)	14.74^{***}
	(2.29)
reserved	9.25^{*}
	(3.95)
R^2	0.02
Adj. R^2	0.01
Num. obs.	322

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 1: Statistical models

.