Problem Set 2

Ariana Antunes

Due: October 16, 2022

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 16, 2022. No late assignments will be accepted.
- Total available points for this homework is 80.

Question 1 (40 points): 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).

```
Upperclass <- c(14, 6, 7)
Lowerclass <- c(7, 7, 1)


sum((Upperclass - Lowerclass)**2/Upperclass)

#Manually chi.square value
((7 - 14)^2)/14 + ((7 - 6)^2)/6 + ((1 - 7)^2)/7
# > [1] 8.809524

n <- length(Lowerclass + Upperclass)
```

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

```
#> p-value : 1.88
```

p-value = 1.281584

[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	-0.5773503	1.1547005	-0.5773503
Lower class	0.4082483	-0.8164966	0.4082483

```
ls(chi)
residuals <- chi$residuals
chi$residuals

## (d)
```

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

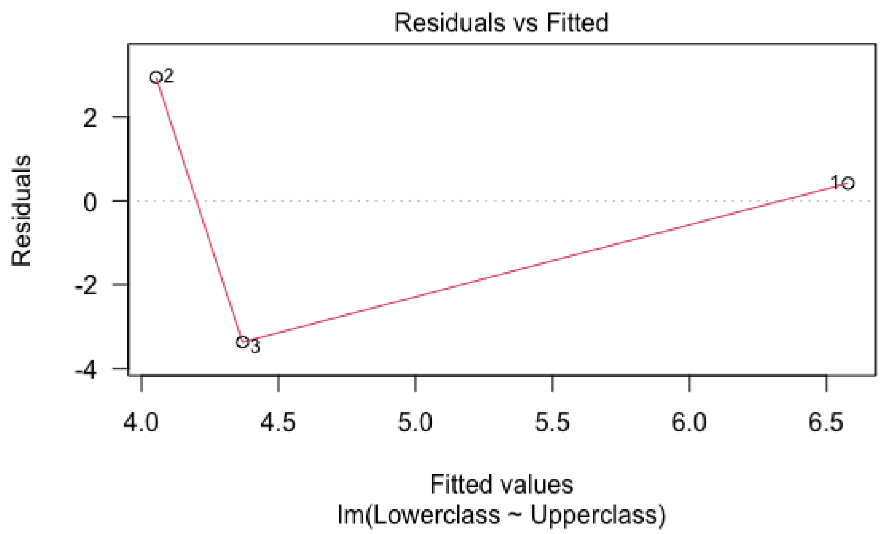
When compering the cells, the standardized residual makes it easy to see which cells are contributing the most to the value, and which are contributing the least. If the residual is less than -2, the cell's observed frequency is less than the expected frequency. Greater than 2 and the observed frequency is greater than the expected frequency.

```

lm <- lm(Lowerclass ~ Upperclass)
plot(lm, las = 1)

ggsave("Residuals plot")
```

Figure 1: Residuals plot



Question 2 (40 points): 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 2 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 2: Names and description of variables from Chattopadhyay and Duflo (2004).

[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.

For Null Hypothesis, Ho we will state that there is no relationship between the variables. For alternative hypothesis, Ha we assume there is a relationship between the variables policy changes for water and females.

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

```
pbar = 124/5745 #sample proportion
p0 = .5 #hypothesis value
n = 26082 #sample size

z = (pbar - p0)/sqrt(p0*(1 - p0)/n) #test statistic
z # [1] -154.5276

#critcal valye at .05 significant level

alpha = .05
z.half.alpha = qnorm(1 - alpha/2)
c(- z.half.alpha, z.half.alpha)

# [1] -1.959964  1.959964

t.test(women$reserved, women$water, mu = ,)

t.test(reserved ~ water,
        data = women,
        mu = 2,
        var.equal = FALSE,
        alternative = "two.sided",
        conf.level = .95)

t.test(women$water, conf.level = 0.5, alternative = "two.sided")

boxplot(women$reserved, women$water)

# p value
pval = 2* pnorm(-154.5, lower.tail=FALSE)  # upper tail
```

```
pval


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

(c) Interpret the coefficient estimate for reservation policy.

```
lm(women$female ~ women$water)
?coefficients

coef(rb)
# (Intercept)    reserved
#   14.738318    9.252423
```

With a high percentage of the reserved variable being 9.25, it is assumed that it has a positive affect on the water treament.