Problem Set 2

QinGuo 24338859

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

	Not Stopped	Bribe Requested	Stopped/Given Warning
Upper class	14	6	7
Lower class	7	7	1

Null Hypothesis: There is no relation between a driver's socioeconomic class and bribe solicitation, the variables are statistically independent. Alternative Hypothesis: A driver's socioeconomic class impacts whether they are solicited for a bribe, the variables are statistically dependent.

(a) Calculate the χ^2 test statistic by hand/manually (even better if you can do "by hand" in R).

The chi-square test statistic is 3.80. This was calculated by hand and through R using the Frequency table and code below.

Figure 1: Frequency Table

	Not Stopped	Bribe Requested	Stopped/Given Warning	Total
Upper Class	obs: 14 | exp: 13.5	obs: 6 | exp: 8.36	obs: 7 | exp: 5.14	27
Lower Class	obs: 7 | exp: 7.5	obs: 7 | exp: 4.64	obs: 1 | exp: 2.86	15
Total	21	13	8	42

```
1  # Setting up Matrix observed <= matrix (c (14 , 6 , 7 , 7 , 7 , 1) , nrow = 2 , byrow=TRUE)
2  rownames( observed ) <= c("Upper Class" , "Lower Class")
3  colnames ( observed ) <= c("Not Stopped" , "Bribe Requested" , "Stopped/Given
4      Warning")       row       total       <=
   rowSums( observed ) column total <=
5  colSums( observed )
6  overall <= sum( observed )
7
8  #expected frequency expected <= round (( outer (row ₋ total , column ₋ total ) /
9  overall ) ,2) expected
10 #overall table f u l l      ₋ table <= matrix ( paste ("obs : " , observed , " | exp : " , round( expected ,
11      2) )
12     , nrow = nrow( observed ) , ncol = ncol ( observed )) rownames( f
   u l ltable ) <= ₋ rownames( observed ) colnames ( f u l ltable )
13 <=    colnames ̄ ( observed )
14    ̄ f u l l table <= cbind ( f ̄ u l l table , row total = paste (row ₋ total ) ) f u l l table <=
15    ̄ rbind ( f u l l table , ̄ c(column total = paste (column ₋ total ) ,
16    Total = paste ( overall ) ) ) rownames( f u l l table ) [ nrow( f u l
   ltable ) ] <= "Total" colnames ( f u l l ̄ table ) [ ncol ( f u l
17 ltable ) ] <= "Total"                    ̄
18
19
20 #part a chi ₋ square stat <= round(sum(( observed = expected )ˆ2 / expected ) , 2)
```

(b) Now calculate the p-value from the test statistic you just created (in R).[1] What do you
conclude if $\alpha = 0.1$?

The p-value of 0.15 is greater than confidence level of 0.1, so we cannot reject the null
hypothesis,and there is no significant association between a driver's social class and
the police officer's behavior. Code:

```
1  chi _ pvalue  <=  pchisq  (  chi _ square _ stat ,  df = (nrow( observed )=1)*( ncol (
       observed )=1) , lower . tail = FALSE)
```

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

```
standardized      -              residuals <= round ((( observed = expected ) / ( sqrt ( expected *
     (1=(row   -   total / overall ) ) %*% t(1=(column _ total / overall ) ) ) ) ) ,3)
```

	Not Stopped	Bribe requested	Stopped/given warning
Upper class	0.322	-1.644	1.526
Lower class	-0.322	1.644	-1.525

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

The upper class drivers being stopped without a bribe has a relatively large negative
standardized residual, indicating fewer stops than expected.

The lower class drivers being asked for a bribe has a relatively small negative standardized
residual, indicating slightly fewer bribe requests than expected.hypothesis while standardized
residuals below -3 or above 3 provide convincing evidence of a cells true effect.

---

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

Question 2: Economics

Chattopadhyay and Duflo were interested in whether women promote different policies than men.[2] 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).

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-Hypothesis(H0): The reservation policy does not impact the number of new or repaired drinking water facilities in villages. Alternative Hypothesis(Ha): The number

[2] Chattopadhyay and Duflo. (2004). "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India. *Econometrica*. 72 (5), 1409-1443.

of new or repaired drinking water facilities in villages are impacted by the reservation policy.

Ho: b = 0 Ha: b/= 0

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

Figure 3: Bivariate Regression

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

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
```

```
1  bireg <= lm( water~reserved , data = women)
2  summary( bireg )
```

(c) Interpret the coefficient estimate for reservation policy.

The regression analysis yields a coefficient estimate of 9.252 with a p-value of 0.0197, indicating strong statistical evidence. Since the p-value (0.0197) is less than the conventional significance level of 0.05, we have sufficient reason to reject the null hypothesis, which states that reservation policies have no effect on the number of new or repaired drinking water facilities in villages. In simpler terms, at a 95% confidence level, we can conclude that villages with reserved positions for women leaders have approximately 9 more new or repaired drinking water facilities compared to those without such a policy in place.