Problem Set 2

Applied Stats/Quant Methods 1

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.

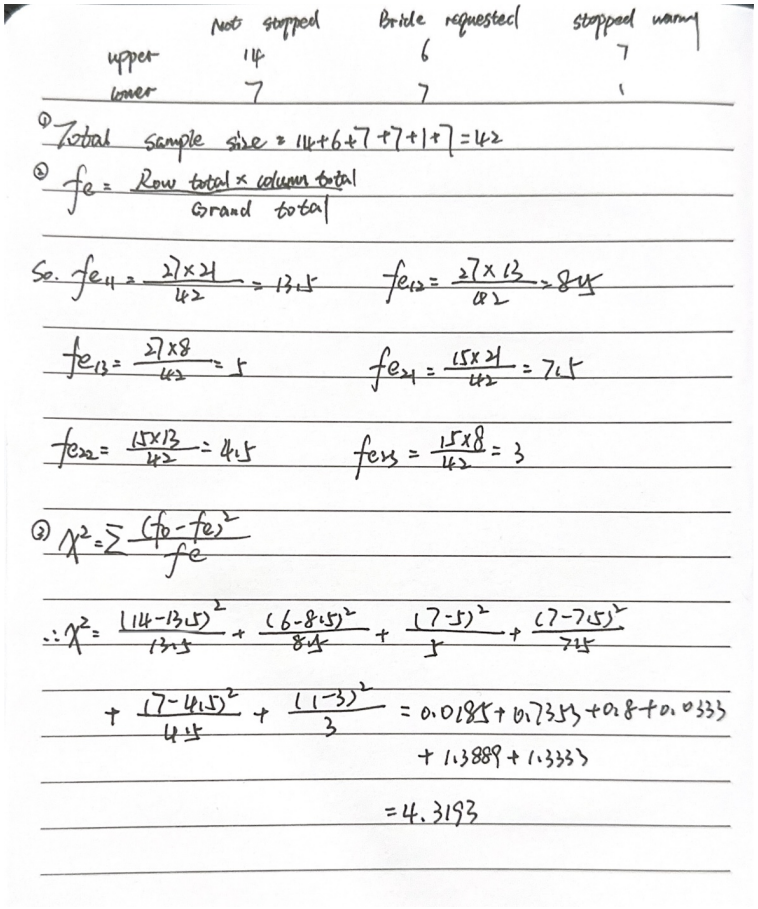
Question 1: Political Science

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).



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

```r
pchisq(4.3193,df=2,lower.tail=FALSE) #answer:0.1153655
```

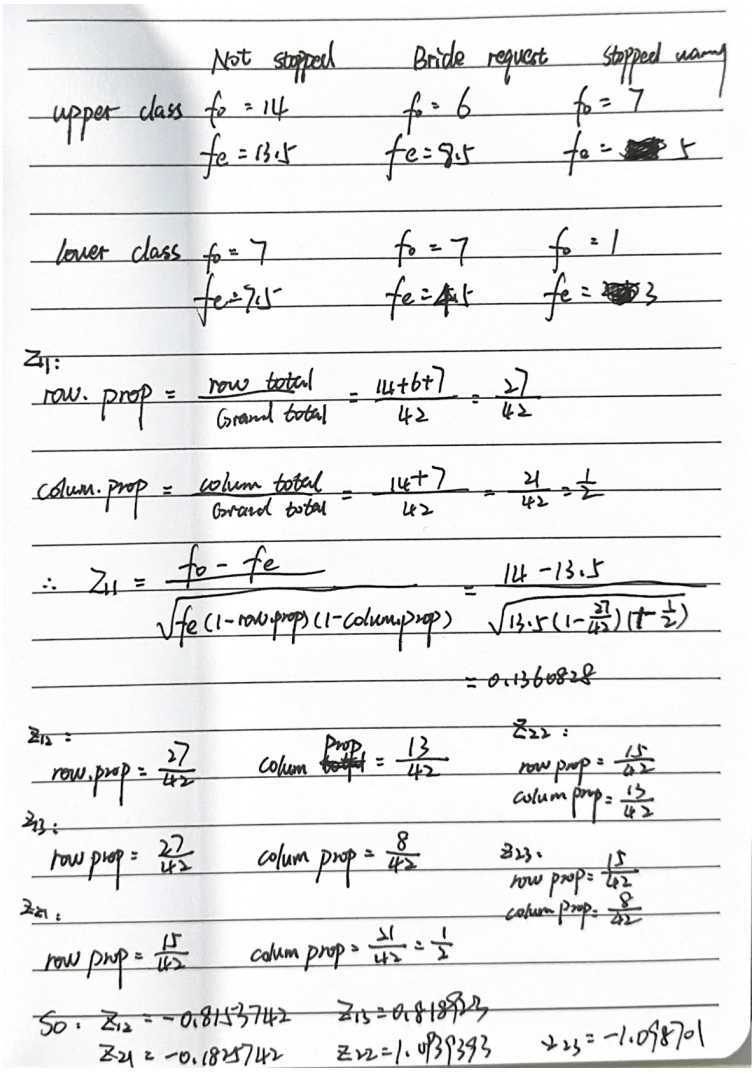
df=(row-1)*(column-1)=(2-1)*(3-1)=2

ANSWER:
According to the obtained results, p is 0.1153655,
which is greater than a=0.1. Therefore, we cannot reject the null hypothesis,
which means that there is not enough evidence to
suggest that police officers are more or less likely to solicit bribes
based on the driver's class.

[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	0.1360828	-0.8153742	0.818923
Lower class	-0.1825742	1.0939393	-1.098701



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

ANSWER:
1. Not stopped:
(1)The standardized residual of the upper class is 0.1360828, close to 0,
indicating that this result is consistent with expectations and has
no significant deviation.
(2)The standardized residual of the lower class is -0.1825742,
which is also a value close to 0, indicating that this result is also consistent
with expectations and has no significant deviation.
2. bribe requested:
(1)The standardized residual of the upper class is -0.8153742,
which is a negative value but with an absolute value less than 2,
and is generally not considered an outlier.
(2)The standardized residual of the lower class is 1.0939393, which is positive
3. stopped warning:
(1)The standardized residual of the upper class is 0.818923, which is positive
and has an absolute value greater than 1 but less than 2.
This may indicate some bias, but does not necessarily mean it is an outlier.
(2)The standardized residual of the lower class is -1.098701,
which is a negative value with an absolute value greater than 1 but less than 2
and may also indicate some bias.
Conclusion:The absolute value of the standardized residuals does not exceed 2,
therefore, there are no obvious outliers,
and the assumptions of the model are satisfied.

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.

ANSWER:
Null hypothesis: the reservation policy have no impact on
the number of new or repaired drinking water facilities
in the villages.
Alternative hypothesis: the reservation policy have impact 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!).

```
RJC <- read.csv("https://raw.githubusercontent.com/kosukeimai/qss/master/
    PREDICTION/women.csv")
model <- lm(water ~ reserved, data = RJC)
summary(model)
```

(1)The results obtained by running in R:

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

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
```

(2)Answer:
The p-value of the reservation policy is 0.0197,
which is less than 0.05. Therefore, we can reject the null hypothesis that
the reservation policy have no impact on the number of new or repaired drinking
water facilities in the villages.

(c) Interpret the coefficient estimate for reservation policy.

the coefficient estimate for reservation policy is 9.252,
indicating that for every additional unit of reservation policy,
the number of the number of new or repaired drinking water facilities
in the villages will increase by 9.252 units.

This also indicates that the reservation policy have impact on
the number of new or repaired drinking water facilities
in the villages.