Problem Set 2

Applied Stats/Quant Methods 1

Due: October 23, 2025

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 Thursday October 23, 2025. 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`).

The null hypothesis is that class and willingness to solicit a bribe is independent. The alternative hypothesis is that class and willingness to solicit a bribe is dependent.

```r
class_bribe <- as.table(matrix(c(14,6,7,7,7,1), nrow = 2, byrow = TRUE,
                               dimnames = list(class = c("Upper class", "Lower class"),
                                               bribe = c("Not Stopped","Bribe requested", "Stopped/given warning"))))
print(class_bribe)

sum_class_rows <- rowSums(class_bribe)
print(sum_class_rows)
#Upper class Lower class
#27          15
sum_bribe_columns <- colSums(class_bribe)
print(sum_bribe_columns)
#Not Stopped       Bribe requested Stopped/given warning
#21                    13                    8
total_sum <- sum(class_bribe)
#[1] 42

expected_up_not <- (27/42)*21
expected_low_not <- (15/42)*21
expected_up_bribe <- (27/42)*13
expected_low_bribe <- (15/42)*13
expected_up_stop <- (27/42)*8
expected_low_stop <- (15/42)*8


expected_values_class_bribe <- matrix(c(expected_up_not, expected_up_bribe, expected_up_stop, expected_low_not, expected_low_bribe, expected_low_stop), nrow = 2, byrow = TRUE,
                                      dimnames = list(class = c("Upper class", "Lower class"),
                                                      bribe = c("Not Stopped","Bribe requested", "Stopped/given warning")))
print(expected_values_class_bribe)

chi_sq_class_bribe <- sum(((class_bribe - expected_values_class_bribe)^2)/expected_values_class_bribe)
#[1] 3.791168
```

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

```
pchisq(chi_sq_class_bribe, df = 2, lower.tail = FALSE)
#[1] 0.1502306
```

Since the p value from the chi squared test is larger than the alpha of 0.1, then we fail to reject the null hypothesis that class and soliciting bribes are independent. There is insufficent evidence to state the alternative, that class and soliciting bribes are dependent.

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

	Not Stopped	Bribe requested	Stopped/given warning
Upper class	0.3220306	-1.6419565	1.5230259
Lower class	-0.3220306	1.6419565	-1.5230259

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

Since the standardized residuals above are all less than -/+ 2, they suggest that the observed values do not differ substantially from the expected values, therefore, supporting the idea that class and soliciting bribes are independent.

[2]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.[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 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 is that there is no association between reservation policy and number of new or repaired drinking water facilities in the village. The alternative hypothesis is that there is an association between reservation policy and number of new or repaired drinking water facilities in the village.

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

```
summary(lm(women_data$water~women_data$reserved))
#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 ***
#  women_data$reserved     9.252       3.948    2.344    0.0197 *
  ---
#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 coefficient estimate for reservation policy.

When a village has a policy reserving seats for women, the number of new or repaired drinking water facilities are expected to increase by 9.252 units.