# Problem Set 2

Applied Stats/Quant Methods 1

Due: October 15, 2023

## 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 15, 2023. 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 $\chi^2$ test statistic by hand/manually (even better if you can do "by hand" in R).

**Step 1: Define the Hypothesis**

**H0:** There is no link between class and bribery attempt.

**H1:** There is a link between class and bribery attempt.

**Step 2:** Calculate the Expected Values

| | Not Stopped | Bribe requested | Stopped/given warning |
|---|---|---|---|
| Upper class | 13.5 | 8.357142857 | 5.142857143 |
| Lower class | 7.5 | 4.642857143 | 2.857142857 |

**Step 3: Calculate the Test Statistic $\chi^2$**

$\chi^2 = 3.791168$

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

Critical Value = 4.60517

Fail to reject the null hypothesis: There is no significant link between class and bribery attempt

[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 | 0.136082763 | -0.815374248 | 0.818923025 |
| Lower class | -0.182574186 | 1.093939348 | -1.098700531 |

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

For $\alpha = 0.1$, the critical stats is 4.605. Our calculated Chi2 test value is 3.791168.

Since Chi2 value is lower than the critical stats, we fail reject null hypothesis where There is no link between class and bribery attempt.

Reference: Simplilearn - Chi-Square Test

Reference: LibreTexts - Critical Values of Chi-Square Table

# 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).


[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.

H0: There is no link between reservation policy and repaired drinking water facilities.
H1: There is a link between reservation policy and repaired drinking water facilities.

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

```
data <- read.csv(url("https://raw.githubusercontent.com/kosukeimai/qss/master/PREDI

model <- lm(formula = data$reserved ~ data$water)
y <- data$reserved
x <- data$water
model <- lm(y ~ x, data = data)
summary(model)

Call:
lm(formula = y ~ x, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-0.5856 -0.3247 -0.3083  0.6602  0.6971

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.3028613  0.0296247  10.223   <2e-16 ***
x           0.0018240  0.0007782   2.344   0.0197 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4696 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

coefficient_x <- coef(model)["x"]
p_value_x <- summary(model)$coefficients["x", "Pr(>|t|)"]

alpha <- 0.05
```

```
if (p_value_x < alpha) {
  cat("Reject the null hypothesis: There is no link between reservation policy and
} else {
  cat("Fail to reject the null hypothesis: There is no link between reservation pol
}
```

**Reject the null hypothesis:** There is no link between reservation policy and repaired drinking water facilities.

(c) Interpret the coefficient estimate for reservation policy.