# Problem Set 2

Applied Stats/Quant Methods 1

Charis Rogers (19333698)

## 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.

| | 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`).

```
#expected freq.
fe1 <- (27/42)*21
fe2 <- (15/42)*21
fe3 <- (27/42)*13
fe4 <- (15/42)*13
fe5 <- (27/42)*8
fe6 <- (15/42)*8
```

[1]Fried, Lagunes, and Venkataramani (2010). "Corruption and Inequality at the Crossroad: A Multi-method Study of Bribery and Discrimination in Latin America. *Latin American Research Review.* 45 (1): 76-97.

```r
chisquare <- sum(((14-fe1)^2)/fe1,
    ((7-fe2)^2)/fe2,
    ((6-fe3)^2)/fe3,
    ((7-fe4)^2)/fe4,
    ((7-fe5)^2)/fe5,
    ((1-fe6)^2)/fe6
)
chisquare # 3.791168
```

The $\chi^2$ test statistic for this data is 3.79

(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
p_value <- pchisq(3.791168, df=2, lower.tail=F)
p_value #0.1502306
```

The p-value is 0.15 which is $> 0.1$. This means we fail to reject the null hypothesis which states that the two variables are independent at $\alpha = 0.1$. This supports the view that class and the officers response are not independent.

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

```r
stan_res <- c((14-fe1)/sqrt(fe1*((1-27/42)*(1-21/42))),
              (7-fe2)/sqrt(fe2*((1-15/42)*(1-21/42))) ,
              (6-fe3)/sqrt(fe3*((1-27/42)*(1-13/42))) ,
              (7-fe4)/sqrt(fe4*((1-15/42)*(1-13/42))) ,
              (7-fe5)/sqrt(fe5*((1-27/42)*(1-8/42))) ,
              (1-fe6)/sqrt(fe6*((1-15/42)*(1-8/42)))
)
stan_res #0.3220306 -0.3220306 -1.6419565 1.6419565 1.5230259
  -1.5230259
```

| | 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?

Standardized residuals help us analyse how far away each value is from the 'expected' observation, showing where the deviation from independence takes place. None of the values in the table above are statistically significant at $\alpha = 0.1$. This supports the decision to not reject the null hypothesis.

[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 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.

Null hypothesis: There is no relationship between the reservation policy and the number of new or repaired water facilities in the village.

Alternative hypothesis: There is a relationship between the reservation policy and the number of new or repaired water facilities in the village.

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

```
lm <- lm(formula = q2$water~ q2$reserved)
summary (lm)
#Coefficients:
#               Estimate Std. Error t value Pr(>|t|)
#(Intercept)    14.738      2.286   6.446 4.22e-10 ***
#q2$reserved     9.252      3.948   2.344   0.0197 *
```

(c) Interpret the coefficient estimate for reservation policy.

According to the data above, there was a difference in the number of new or repaired drinking water facilities comparing villages which reserved for women leaders and those that did not since the reservation policy started. In areas where there was no GP reserved for women there was about 15 new or repaired facilities. In areas where there was a GP reserved, there were about 9 more new or repaired drinking water facilities, compared to the areas without.

This difference observed is a statistically meaningful result as the p-value of 0.02 is $< 0.05$. This means we can reject the null hypothesis that there is no relationship between the reservation policy and the number of new or repaired water facilities in the village