Problem Set 2

Applied Stats/Quant Methods 1

Due: October 16, 2022

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 16, 2022. No late assignments will be accepted.
- Total available points for this homework is 80.

Question 1 (40 points): 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 Multi-method 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).

Read in the data as a matrix.

```
observed  <- matrix( c (14, 6, 7, 7, 7, 1), nrow = 2, byrow = TRUE)
```

Calculate the expected values (as a matrix), then calculate the difference between each cell and its contribution to the chi-squared statistic. (expected = number in class * number of outcomes / total number; difference = observed - expected; contribution = difference^2 /expected)

Upper Class, Not Stopped

observed	14
expected	13.5 = (27 * 21 / 42)
difference	0.5 = (14 - 13.5)
chi sq contribution	0.0185 = $(0.5)^2$ / 13.5

observed = 14

```
ncols <- length(observed[1,])
nrows <- length(observed[,1])

# get totals
row_tots <- vector("double", nrows)
col_tots <- vector("double", ncols)

totals <- sum(observed)  # total number of observations

# calculate row and column totals, e.g, total for NotStopped, UpperClass,
    etc
for (i in 1:nrows) {row_tots[i] <- sum(observed[i, ])}
for (i in 1:ncols) {col_tots[i] <- sum(observed[, i])}

#get expected = row total * column total / total observations
expected <- observed
for (i in 1:nrows) {
  for (j in 1:ncols) {
```

```
    expected[i,j] <- row_tots[i] * col_tots[j] / totals
  }
}

# calculate difference between observed and expected
o_e <- observed
for (i in 1:nrows) {
  for (j in 1:ncols) {
    o_e[i,j] <- (observed[i,j] - expected[i,j])^2 / expected[i,j]
  }
}

#calculate chi-squared value & degrees of freedom
chi_sq_val <- sum(o_e)
df = (nrows-1) * (ncols-1)
```

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

```
p_value <- pchisq(chi_sq_val, df=df, lower.tail=FALSE)
alpha <- 0.1
```

```
The p-value is 15.02%, alpha is 10%
We cannot reject the null hypothesis that the two sets are from the
same population
1 observed cell(s) have less than 5 values
```

The observed and expected values are shown in Figure 1

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

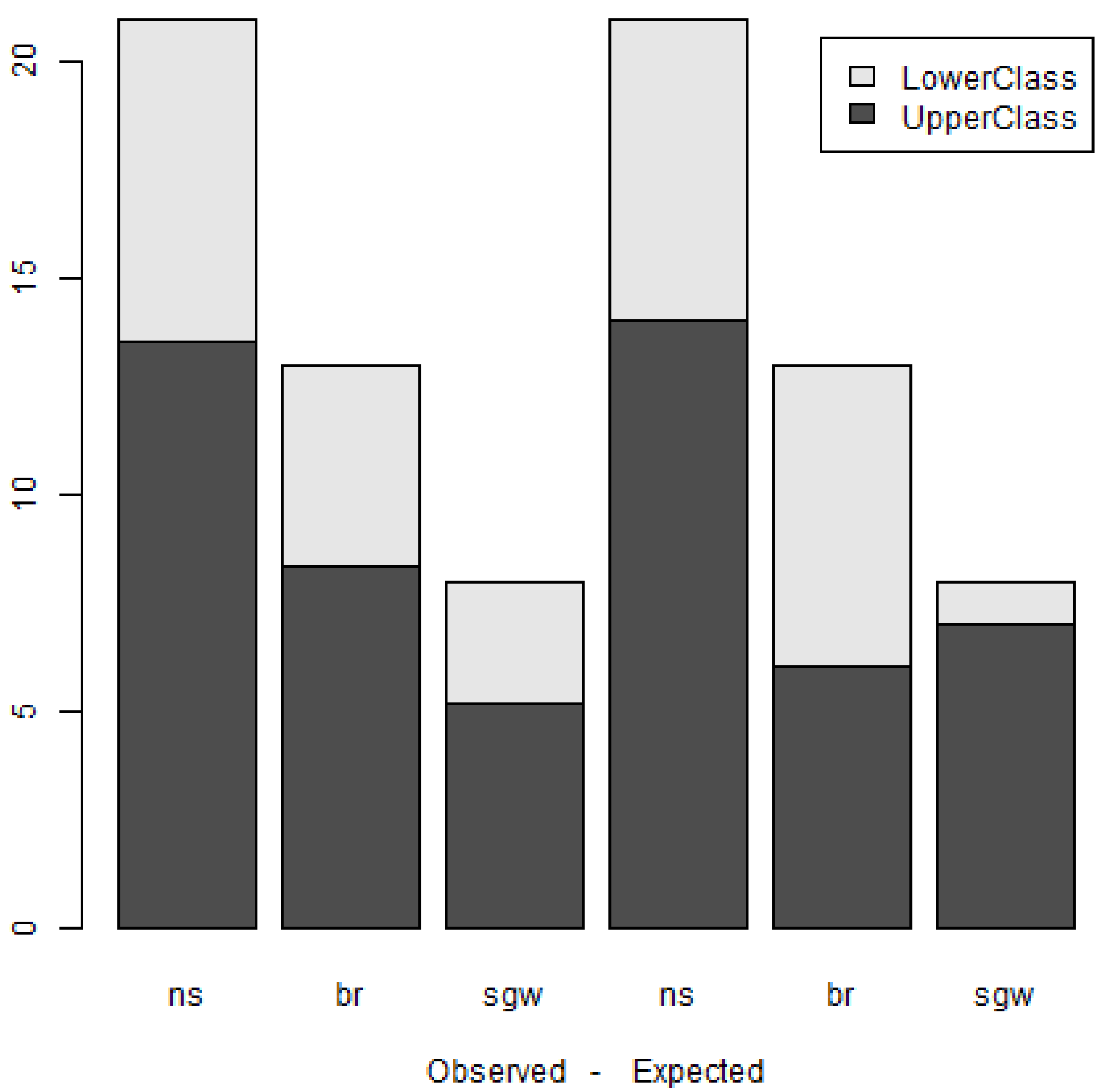

Figure 1: Observed vs Expected values for traffic stop. ns = Not Stopped; sgw = Stopped Given Warning; br = Bribe Requested

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

Table 1: Standardised Residuals

	NotStopped	BribeRequested	StoppedGivenWarning
UpperClass	0.32	-1.64	1.52
LowerClass	-0.32	1.64	-1.52

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

The biggest contribution to the residuals was from the ‘Bribe Requested’ variable - fewer upper class individuals were expected to hand over bribes. The difference between the two groups appears to be a combination of fewer upper class individuals being expected to hand over bribes and more of them being given a warning instead.

Question 2 (40 points): 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 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

[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 The reservation policy has no effect on the number of new or repaired drinking water facilities in the villages.

Alternate The reservation policy does have an effect 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!).

(c) Interpret the coefficient estimate for reservation policy.