Roisin Gallagher - Problem set 2

October 14, 2023

1 Question 1 Political Science

6 **(a)** The formula for the chi-squared test statistic is:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Where: - O is the observed frequency. - E is the expected frequency.

The expected frequency for each cell in a contingency table is calculated was calculated by hand in R

Then, I plugged in expected frequencies using the formula of chi-squared:

$$\begin{aligned}\chi^2 = &\frac{(14-13.5)^2}{13.5} + \frac{(6-8.375)^2}{8.375} + \frac{(7-5.142)^2}{5.142} \\ &+ \frac{(7-7.5)^2}{7.5} + \frac{(7-4.642)^2}{4.642} + \frac{(1-2.857)^2}{2.857}\end{aligned}$$

Calculating this:

$$\chi^2 = 0.0185 + 0.6735 + 0.6713 + 0.0333 + 1.1977 + 1.207$$

$$\chi^2 = 3.8013$$

The calculated chi-squared test statistic is $\chi^2 = 3.8013$.

(b) In order to calculate the p-value from the chi-squared test statistic in R, I can use the function 'pchisq' given that the degrees of freedom for our question is the number of rows - 1, number of columns - 1 = 1 * 2 = 2 and the chi-squared from the previous section is 3.8013. Now, I can plug in these values in the function.

p_value <- 1 - pchisq(3.8013, df = 2)

and is equal to 0.1494. I can compare this with 0.1 and since it is bigger than 0.1, I can conclude that there is no association between the two variables and haven't sufficient evidence to reject the null hypothesis

(c) The standardised residual tells us where the deviation from independence takes place In order to calculate the standardised residual, the formula is: $r_{ij} = \frac{O_{ij}-E_{ij}}{\sqrt{E_{ij}}}$

where

$$O_{ij}: \text{Observed frequency for the } i^{th} \text{ row and } j^{th} \text{ column}$$
$$E_{ij}: \text{Expected frequency for the } i^{th} \text{ row and } j^{th} \text{ column}$$

(d) Standardized residuals indicate the difference between observed and expected counts. Positive residuals show higher-than-expected observations; negative show lower-than-expected.

This data aids police, policy decisions and further research.

2 Question 2 Economics

(a) Null Hypothesis: The reservation policy has no effect on the number of new or repaired drinking water facilities in the village.

(b) To run a bivariate regression model in R I can use function lm() as below

model <- lm(water   reserved, data = women_data)

(c) If I increase the reservation policy by 1, then increase in water by 0.304