assignment_3_written

March 19, 2025

1 Assignment 3 Part 1

1.1 Q1

A researcher wants to investigate whether the type of persuasive technique used influences people's decision to agree with a donation to the church. The study involves 200 participants, divided based on whether they agreed or disagreed after being exposed to one of three persuasive techniques: Emotional Appeal, Logical Argument, or Social Proof. The observed frequencies are as follows

Decision Outcome	Emotional Appeal	Logical Argument	Total
Agree	30	25	55
Disagree	20	30	50
Neutral	10	25	35
Total	60	80	140

1. Calculate the probability of people agreeing to donate $P(donate)$ and the probability of people being exposed to emotional appeal $P(emotion)$.

2. Assuming the people's decision outcome is independent of the type of persuasive technique they are exposed to. What is the probability of people being exposed to emotional appeal **and** agreeing to donate $P(donate \cap emotion)$?

3. The research would like to conduct a chi-squared test to test the dependency between persuasive techniques and people's decision to donation. State the null hypothesis and the alternative hypothesis.

4. Calculate the expected frequency of people agreeing to donate after being exposed to emotional appeal, assuming the null hypothesis is true.

5. Calculate the χ^2* statistic using the formula

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

6. What is the sampling distribution for this χ^2*? In other words, what is the degree of freedom for this statistic?

7. Write down the Python code to calculate the p value. Recall that the Python code to calculate the cummulative probability for a χ^2 distribution is

```
from scipy import stats
1-stats.chi2.cdf(x, df=degree_of_freedom)
```

1.2 Q2

A researcher is studying whether there is a relationship between the amount of time voters spend watching political debates (in hours) X and their trust level in political candidates (measured on a scale of 1–10, where 10 indicates the highest trust) Y. He studies the research question asking whether there exists a correlation between the time spent watching political debates and voter trust in political candidates. The collected data is shown below:

Voter	Time Spent Watching Debates (X)	Trust Level (Y)
1	2	5
2	4	6
3	3	7
4	5	8
5	6	9

1. Compute the sample mean and sample variance for both X and Y. $\bar{X}$, $\bar{Y}$, S_x^2, and S_y^2
2. Compute the sample covariance $\widehat{Cov(X,Y)} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$
3. Compute the correlation coefficient between X and Y. $r_{x,y} = \frac{\widehat{Cov(X,Y)}}{S_x S_y}$
4. Draw the scatterplot between X and Y.
5. Describe the relationship between amount of time voters spend watching political debates and their trust level in political candidates.

1.3 Assignment 3 part 2

In this part of the assignment, you need to download and open the jupyter notebook `assignment_3_coding.ipynb` to follow the instructions and complete the assignment.

You also need to download the data file `smoking.csv` and place it at the same file folder of your jupyter notebook file.

After you have done it, submit both your written assignment and code assignment.