Report for statistics personal final assignment

Description of the strategy:

To finish this assignment I coded a Python program that implements some statistical tests like z-test, t-test and ANOVA, the program in question is in a different file in the folder.

Question 1:

The first question gives us a dataset regarding the US crime rate of two different moments distant 10 years from each other.

Part a:

Is the Crime Rate in the southern states higher than the other states?

To solve this question I used a z-test, the null hypothesis is that the southern states crime rate is less than or equal to the other states crime rate while the alternative hypothesis is that it's higher, I run the test with alpha equal to 0.01, the result is a p-value equal to 0.66, since the p-value is greater than the significance level (alpha) there is not enough evidence to reject the null hypothesis so the statement is false.

Part b:

Have the crime rates increased in 10 years?

To solve this question I used again a z-test, the null hypothesis is that the crime rates after 10 years are less than or equal to the one before, the alternative hypothesis is that they are higher, I run the test with alpha equal to 0.01, the result is a p-value equal to 0.46, since the p-value is again greater than the significance level I don't have enough evidence to reject the null hypothesis, hence the statement is false.

Part c:

Divide the education time into high education time (>13), median education time (>11 and <=13), and low education time (<=11). Are the Crime Rate different among these education groups? Assume the population variances of Crime Rate are the same.

To solve this question I used an ANOVA test, I divided the data set in three education group, the null hypothesis is that the crime rate among these education groups is the same, while the alternative

hypothesis is that they are different, performing the test with a significance level alpha of 0.01, the result of the ANOVA test is 0.66, since the result is greater than alpha I don’t have enough evidence to reject the null hypothesis, so the statement is false.

Part d:

Is there a relationship between high youth unemployment and southern states?

To solve this question I used a chi-squared test with a significance level of 0.01, the result of the test is 8.30^{-13}, looking at the chi squared table we can assert that the result of the test is smaller than the expected value, hence the statement is true.

Question 2:

This question gives us a dataset regarding the COVID-19 spreading during the pandemic period.

Part a:

In the global dataset, is the Fatality Ratio on 2022/07/07 lower than the Fatality Ratio on 2021/07/07?

To solve this question I used a z-test with a significance level of 0.01, the null hypothesis is that the the fatality ratio on 2022/07/07 is greater than or equal to the fatality ratio on 2021/07/07, the alternative hypothesis is that is lower, performing the test we find that the result is 0.25, running the calculation of the expected z value we find him being -2.33, since the result of the test is greater than the expected value we can conclude that there is enough evince to reject the null hypothesis, the statement is then true.

Luca Veronelli, 111550804