Danielle Senechal

MAT-374: Data Analytics

Hom­­ework 3

March 31st, 2020

Task 1:

2. In the training set of the data, 76.0612% of people have incomes less than or equal to $50k. 23.9388% have incomes more than $50k. In the testing data, 76.07522% of people have incomes less than or equal to $50k, while 23.92478 have incomes higher than $50k.

3. The partition is validated. This was done with a Chi-Squared test. This is because the p-value is 0.9834. Since the p-value is above 0.05, this means there is no statistical significance between the two groups, and there is not enough evidence to reject the null hypothesis, which states the two groups have proportions that are equal.

4. The partition is validated. This was done with a Kruskal-Wallis test. The p-value is 0.952, which is well above 0.05. Because the p-value is above 0.05, there is not enough evidence to reject the null hypothesis which states the groups have equal proportions.

Task 2:

1. In this dataset, 93.493151% of people have central air, and 6.506849% of people do not have central air.

3. About 20% of people do not have central air, while about 80% of people have central air.

A picture containing drawing

Description automatically generated