Project Report

APPLIED STATISTICS

Achyut Sridhar Kulkarni

About Dataset:

The purpose of this statistical analysis report is to delve into the intricacies of the provided dataset, which captures a detailed record of online sales transactions. The dataset encompasses a rich array of information, ranging from customer demographics and tenure to specific transaction details like product category, quantity, and coupon usage. Through rigorous statistical analysis, our goal is to unearth meaningful patterns, trends, and insights that can inform strategic decision-making for the online sales platform.

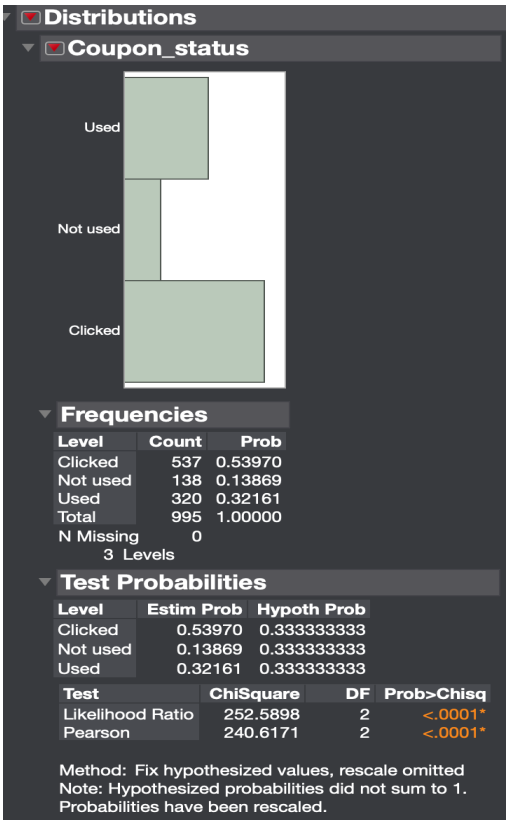
1. Introduction:

In the dataset, for the coupon's status column, it is expected that there are equal distributions of coupons that are used, unused and clicked (33.3% each). Does this data suggest that the percentage of coupon distribution was not correct? (α = 0.05)

Method used: Chi Square Test

$$H_0: \pi_{used} = \pi_{unused} = \pi_{clicked} = 0.333$$
$$H_1: at\ least\ one\ \pi_i \neq 0.333$$

Results:



$$Test\ Statistics: X_0^2 = 240.6171$$
$$p - value: < 0.0001$$

As p-value is less than α, **we reject the null hypothesis**.

Conclusion:

There is sufficient evidence to claim that at least one proportion is different from 0.333.

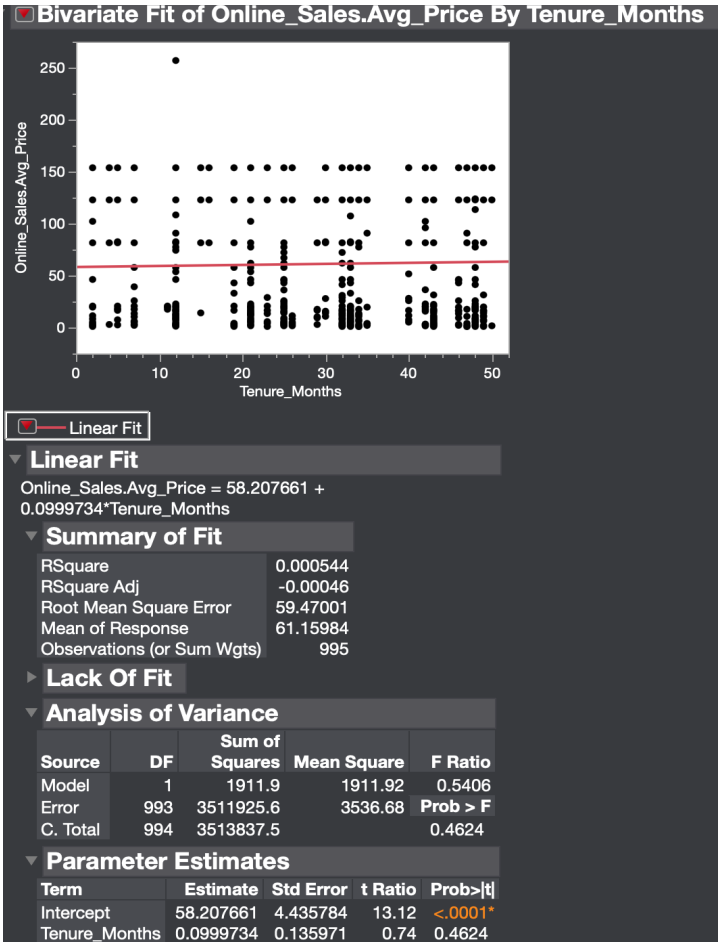
2. Introduction:

We want to build a model to relate average price of the online sales and the tenure in months for which they have been using the online site. (α = 0.05)

Method used: Regression

$$H_0: \beta_1 = 0$$
$$H_1: \beta_1 \neq 0$$

Results:



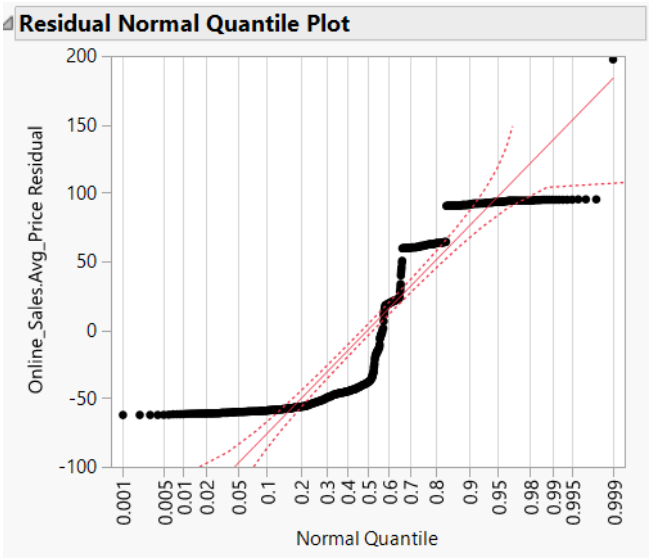
The fitted regression equation is:

$$Online\ Sales\ Average\ price = 58.207 * Tenure$$
$$Intercept: \beta_0 = 58.207$$
$$Slope: \beta_1 = 0.0999$$
$$Test\ Statistic: t_0 = 0.0999$$
$$p - value = 0.4624$$

As p-value is greater than α, we **fail to reject the null hypothesis.**

Conclusion:

There is no strong evidence to claim that the regression model for predicting the relationship between online sales and the tenure is significant.



We can conclude from the normal probability plot of the residuals that the residuals do not form a straight line. Therefore, the residuals appear to violate the assumption that they are normally distributed.

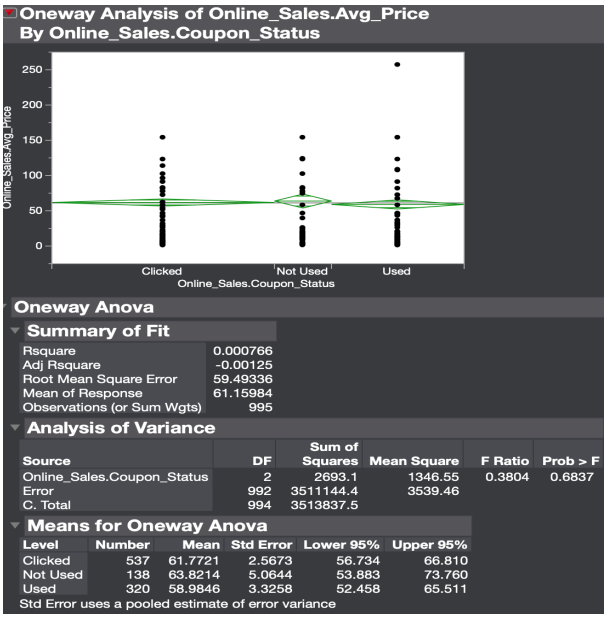
3. Introduction:

We are interested to find if there is relationship between average price and the coupon status for the transaction. The coupons are categorized into used, unused and clicked. Using α = 0.05, test the claim that the three sample coupons has means that are all equal.

Method used: ANOVA

$$H_0: \mu_{used} = \mu_{unused} = \mu_{clicked}$$
$$H_1: At\ least\ one\ of\ the\ means\ is\ different\ from\ the\ others$$

Results:



$$p - value = 0.6837$$

Since p-value is greater than α, we **fail to reject the null hypothesis**.

Conclusion:

We cannot claim that there is difference in average sales based on coupon status.

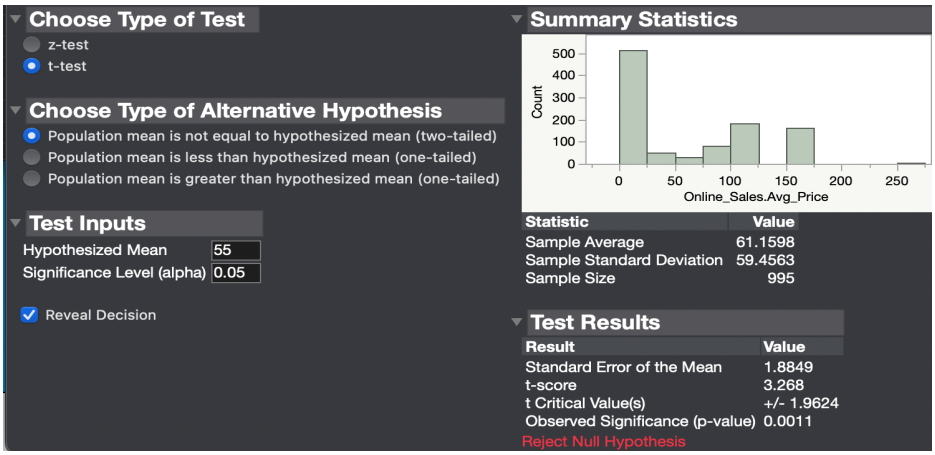
4. **Introduction:**

We are analysing a dataset where 995 transactions were selected at random and their average sales prices were measured. It is of interest if there is evidence that the mean average price is different than 55 which is the population mean. (α = 0.05)

Method used: One Sample t test

$$H_0: \mu = 55$$
$$H_1: \mu \neq 55$$

Results:

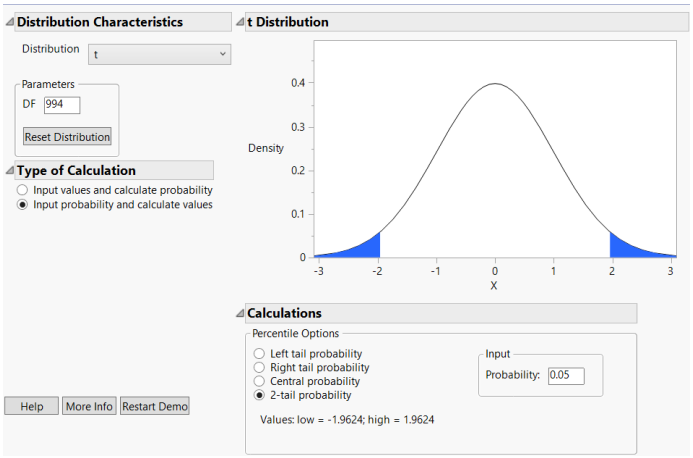


$$Test\ Statistic: t_0 = 3.268$$
$$p - value = 0.0011$$

As p-value is less than α, we **reject the null hypothesis**.

Conclusion:

At 5% level of significance, the true average price is different from 55.



Since the test statistic is greater than critical value, we reject the null hypothesis.
We have sufficient evidence to claim that the true average price is different than 55.