6PAM1037-0901-2022

Linear Modelling

Assignment 1

**Student Name: SrINIVAS Vegisetti**

**Student ID: 21080840**

Contents

1. GitHub Repository Link…………………………………………………………………………………………………. 1
2. Plotted Graphs for Data Analysis…………………………………………………………………………………… 2
3. Question1: Shapiro - Wilk Test………………………………………………………………………………………. 3
4. Question 2: Levene’s Test……………………………………………………………………………………………… 4
5. Question 3: One-Way Analysis………………………………………………………………………………………. 5
6. Question 4: ANCOVA Analysis………………………………………………………………………………………… 6
7. Question 5: ANCOVA Linear Regression line plot……………………………………………………………. 7

# GitHub Repository Link

Link: <https://github.com/Srinivas-herts/Herts-Assignments.git>

My GitHub Repository link contains my Linear Modelling Assignment 1, conducted Shapiro-Wilk test, Levene Test, boxplots, Probability Tests, One-way ANOVA, ANCOVA with co-variate analysis, and plotted Linear Regression lines. This document has detailed visual graphs and output code with a detailed analysis report.

## Plotting the graphs to analyse the data:



By seeing the graph, we can identify that p-value is above 0.05 (p > 0.05)



## **Question 1: Shapiro – Wilk Test**

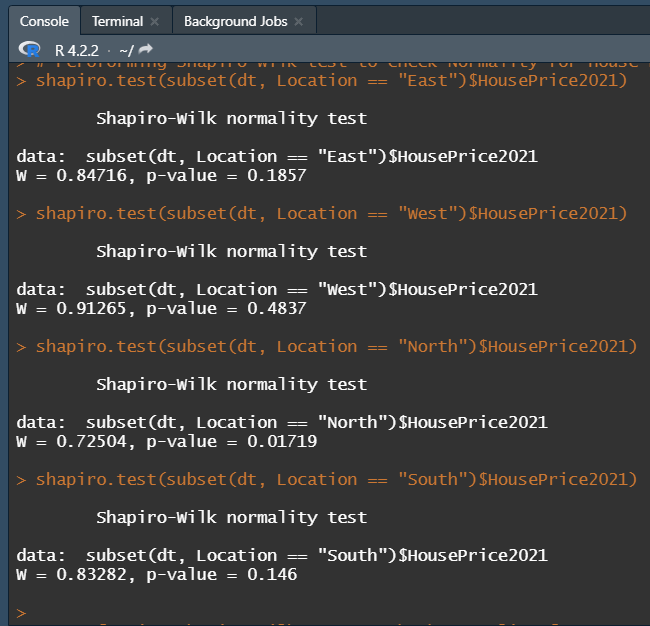
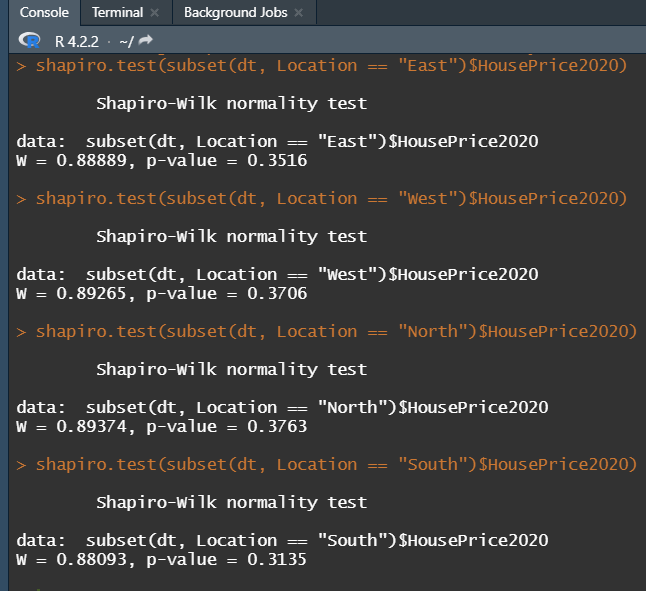
# Performing Shapiro wilk test to check Normality for House Prices in 2021 vs Location wise.

Q1 Comments:

The more significant the p-value, the more compelling the case is to reject the null hypothesis.

A p-value of 0.05 or less (usually 0.05) indicates statistical significance; a p-value of 0.05 or more (> 0.05) does not and suggests substantial evidence supporting the null hypothesis.

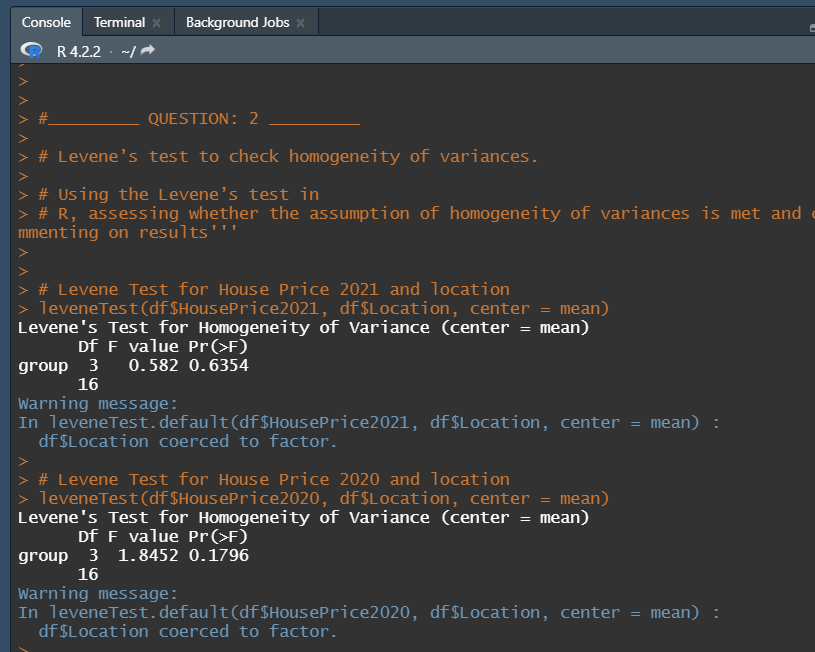
As we can see in below results the p-value is > 0.05.

## Question 2: Levene’s Test

Q 2 Comments:

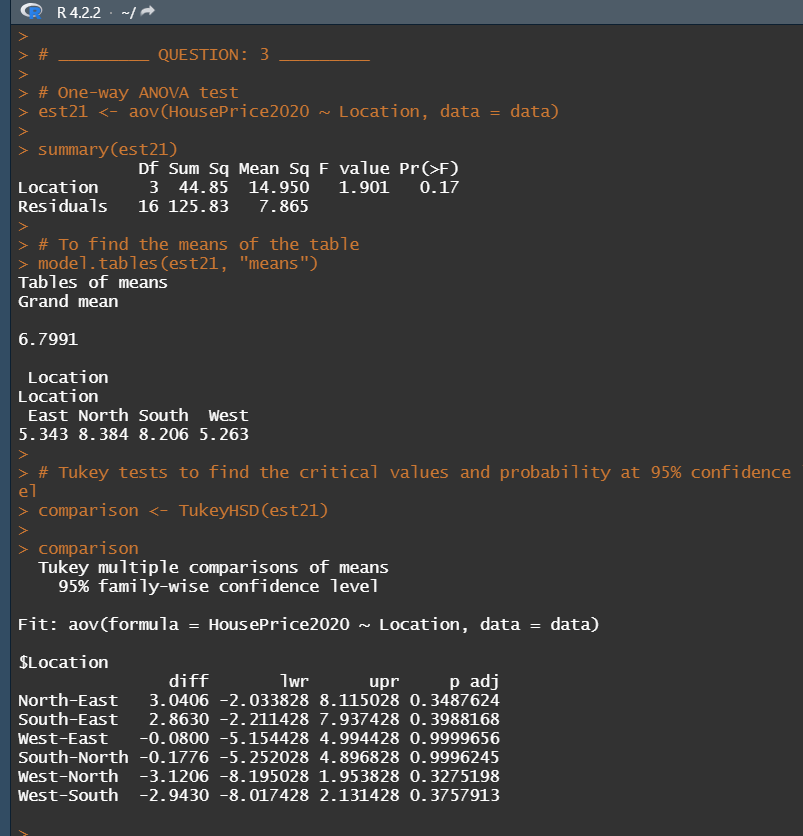
Below Figure explains that we can proceed to perform a one-way ANOVA as the p-value > 0.05, indicating that our analysis has satisfied the requirement of homogeneity of variance. In second case, we can apply a non-parametric test to conduct the study if the p-value < 0.05, indicating that the assumption of homogeneity of variance has been broken.



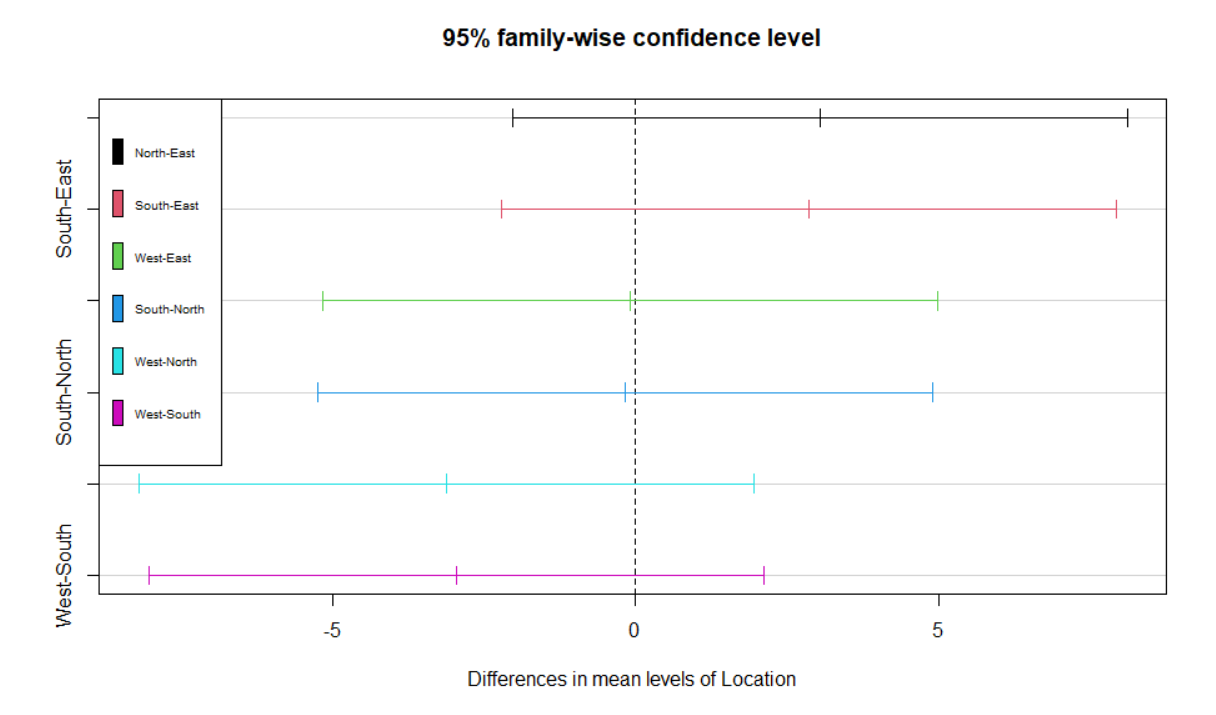
## Question 3: One-way ANOVA Analysis

Q 3 Comments: Null Hypothesis p-value is 3.63 which is greater than 0.05, Thus our model is fit to analyse.

I have carried out a one-way analysis because it has the ability to regulate the overall Type or error rate, ANOVA is used to compare multiple groups. Therefore, we offer the general test for group mean equality. It allows me to regulate the overall inaccuracy rate or kind (i.e. false positive finding). We can see in the results below the Sum sq and Mean sq. These variances are computed by dividing each sum of squares by the associated degrees of freedom. They are referred to as Mean Squares. Result: Nothing but dividing the within-group mean square by the between-group mean square and ensuring the data is statistically significant. By Examining and comparing these means we can choose the best-fit model for our data.



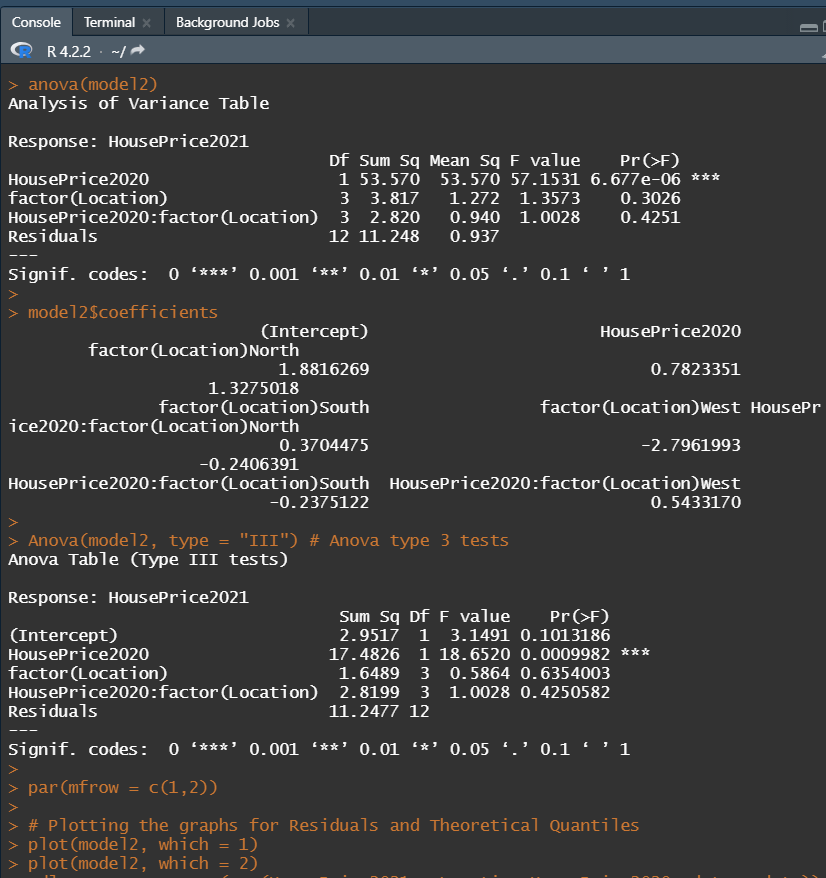
Q 3 Comments: The mean intervals are greater than the significant level(p>0.05) which is normally distributed and the model is fit for linear modeling.



## Question 4: ANCOVA Analysis

Q 4 Comments: Null Hypothesis p-value is 3.38 which is greater than 0.05, Thus our model is fit to analyze.

I have carried out an ANCOVA analysis to regulate the overall Type or error rate, ANOVA is used to compare multiple groups. Therefore, the experiment's significant P-value and correlation coefficient for such dependent variable and covariate.  stating that there is no significant association between the covariate and the independent variable, and claiming that the slope of the regression lines is identical. Here in the below graph the p-value is



Q4 Plots:

Fig: 1 Below Graph is for HousePrice2021 to identify the Residuals and fitted values for and Theoretical Quantiles using Normal Q-Q of for the dependent variables to identify and choose the best fit model

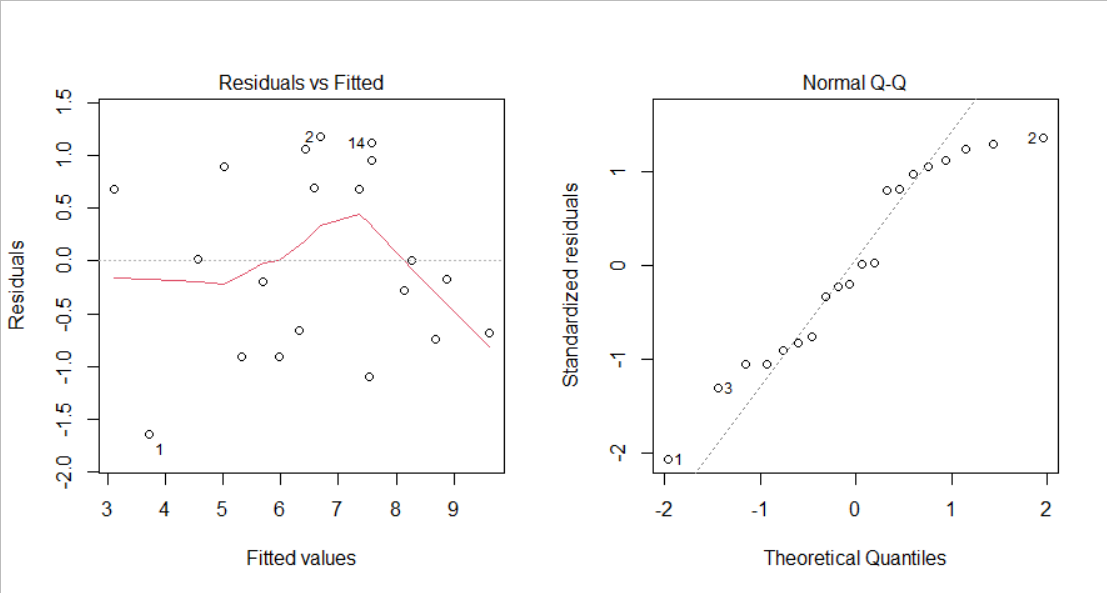
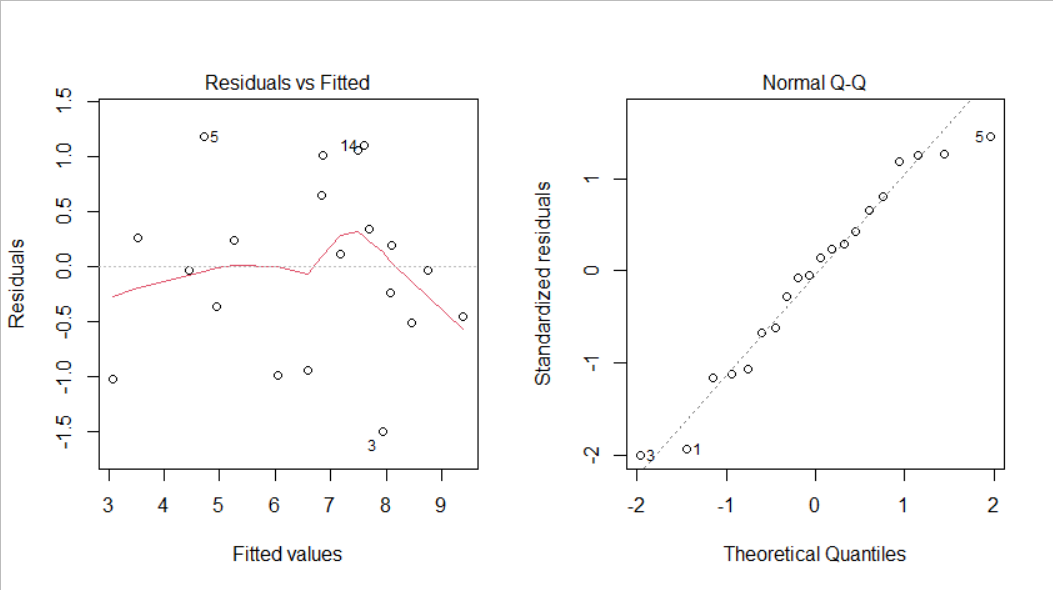


Fig: 2 Below Graph is for HousePrice2020 to identify the Residuals and fitted values for and Theoretical Quantiles using Normal Q-Q of for the dependent variables to identify and choose the best fit model.



## Question 5: ANCOVA Plot

Q 5 Comments:

Plotted below are the four Regression Lines of House Prices increased as per year within locations (East, West, North and South). By running this model I have concluded that Similar to multiple linear regression, ANCOVA is more appropriate when comparing groups based on one of the independent variables, whereas regression is more appropriate when the focus is on the dependent outcome variable. So, we can report the equality of the slope of regression lines due to lack of significance between the covariate and the independent variable. Interpreting the results between the two measuring variables is ease.

