TU Dortmund

Introductory Case Studies

Project 2: Comparison of Multiple distributions

Lecturers:
Prof. Dr. Katja Ickstadt
M. Sc. Zeyu Ding

Author: Bilal Tanvir Bhatti

Group number: 1

Group members: Rafay Maqsood, Maimuna Rahman, Nafisa Farhin

December 14, 2023

Contents

1 Introduction

Numerous crucial factors come into play when considering property rentals. While property price stands as a significant element, it's profoundly impacted by various variables and exhibits considerable variability. For instance, larger properties tend to command higher rents, although this can also fluctuate based on location and the surrounding neighborhood. Property prices showcase diversity from one city or area to another, with variations influenced by living costs in different regions.

This data set centers around investigating the potential correlation between landlord response times and property pricing. The focus lies on exploring how the duration it takes for a landlord to respond—categorized into three levels: "within a few hours," "within a day," and "within an hour"—impacts the logarithmically transformed price of properties. As a result, the data set now comprises two essential columns: 'log-price,' representing the transformed property prices, and 'host-response-time,' classifying the response time into distinct categories.

A comprehensive assessment is conducted using one-way analysis of variance (ANOVA) to examine whether there are differences in property price across various host response time categories. Subsequently, a two-sample t-test is employed for pairwise comparisons among the resulting property prices. To mitigate the issue of multiple testing, adjustments are made to the significance level utilizing both the Bonferroni correction and Tukey's Honest Significant Difference (HSD) method.

Section 2 provides a concise overview of the dataset, discussing its quality and outlining the structure of the descriptive analysis. Section 3 details the statistical methods applied for the comparison of multiple distributions. In Section 4, graphical representations like QQ-plots, bar charts, histograms, and various tests are employed to interpret the findings. Finally, Section 5 offers a summary encompassing all the results.

2 Problem Statement

The dataset utilized in this project is sourced from Airbnb, encompassing information about property pricing and the status of host response times. The host response time is characterized by three categories: 1 = within a few hours, 2 = within a day, and 3 = within an hour. This dataset comprises 232 observations and two variables. The property pricing variable, a numeric attribute, signifies the property's price. Property pricing varies based on the host response time category. Notably, there are no missing values within the dataset.

3 Statistical methods

Several statistical methods that are used to analyze the data are discussed in this section. For this analysis, the R software (R Development Core Team, 2020), version 4.3.0 is used with packages ggplot2 (Wickham, 2016), car (Fox and Weisberg, 2019) and tseries (Trapletti and Hornik, 2023).

3.1 Hypothesis Testing

Hypothesis testing involves using sample data to make inferences or draw conclusions about the larger population. It compares two contradictory statements about a population, aiming to determine which statement is more strongly supported by the evidence found in the sample data.

3.1.1 Null Hypothesis and Alternative Hypothesis

A hypothesis represents a speculative statement regarding the expected outcome of a study and helps describe a population parameter. The null hypothesis, designated as H_0, supports an assumption. However, rejecting the null hypothesis leads to considering the alternative hypothesis, represented by H_1. This implies that if the null hypothesis is incorrect, then the alternative hypothesis is likely to be true, and vice versa. (Akinkunmi, 2019, p. 141)

3.1.2 Significance Level

During hypothesis testing, the significance level α is predetermined. It represents the likelihood of rejecting the null hypothesis and favoring the alternative hypothesis assuming that the null hypothesis is accurate. Typically, this significance level is chosen to be 1 percent, 5 percent, or 10 percent. For instance, if the significance level is set at 5 percent, it implies that among 100 instances, only 5 would potentially lead to rejecting the null hypothesis. (Heumann et al., 2016, p. 213)

3.1.3 P-value

It represents the chance of getting results identical to or more extreme than the observed results if the null hypothesis were accurate. A lower p-value indicates evidence contradicting the null hypothesis. The smaller the p-value (approaching 0), the stronger the evidence against the null hypothesis. If the p-value is equal to or smaller than the chosen significance level α, the null hypothesis is dismissed in favor of the alternative; otherwise, the null hypothesis stands. (Heumann et al., 2016, p. 215)

3.1.4 Type I and Type II Error

In statistics, there are two potential errors. A Type-I error happens when the null hypothesis is incorrectly dismissed despite being true. Conversely, a Type-II error occurs when the null hypothesis is not dismissed even though it is false.

When testing a hypothesis, several outcomes can arise:

	H_0 is true	H_0 is false
H_0 is accepted	Correct Choice	Type II error
H_0 is rejected	Type I error	Correct Choice

Table 1: Hypothesis Testing possible outcomes.(Heumann et al., 2016, p. 213)

3.2 Two Sample Test

During hypothesis testing, the t-test is employed to determine if the means of two groups within a population are identical or differ. It examines the null hypothesis specifically when the variances of key variables in both populations are unknown, which is a common scenario. However, it also assumes equality of these variances across the two populations.

Let us suppose two random variables X_1 and X_2 of size n_1 and n_2 are drawn from populations 1 and 2. The underlying parameters are μ_1 and μ_2, and σ_1^2 and σ_2^2, where μ_1 and μ_2 are the means of the two samples and σ_1^2 and σ_2^2 are the variances of the two samples. The following null hypothesis is tested:

$$H_0 : \mu_1 = \mu_2$$

against the alternative hypotheses.

$$H_1 : \mu_1 \neq \mu_2$$

Test statistics are given by:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1).s_1^2+(n_2-1).s_2^2}{n_1+n_2-2}}} . \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$\bar{x}_1$ is the mean of X_1, $\bar{x}_2$ is the mean of the sample X_2, s_1^2 and s_2^2 are the variances of X_1 and X_2 respectively. The test statistic follows a t-distribution with $n_1 + n_2 - 2$ degrees of freedom, allowing us to either support or reject the null hypothesis based on this distribution. (Rasch et al., 2020, p. 63, p. 64)

3.2.1 Multiple Testing Problem

The multiple testing problem arises when a set of m hypotheses are tested simultaneously. Relying solely on un-adjusted individual p-values for decision-making increases the likelihood of mistakenly rejecting true null hypotheses. For instance, if we set a significance level of 0.05 and the null hypothesis is true, there's a 5% chance of a Type I error (false positive). This means that in 1 - 0.05 cases where the null hypothesis is true, we won't make a Type I error. When conducting two independent tests, the probability of making no false positives is $0.95^2 = 0.9$. Extending this to 12 comparisons yields $0.95^{12} = 0.54$, indicating a 46% chance of encountering at least one false alarm across these 12 comparisons. Consequently, as the number of comparisons increases, the likelihood of false alarms also escalates.

Mathematically, the probability of encountering a single Type I error among a set of m tests is expressed as:

$$1 - (1 - \alpha)^m \tag{1}$$

or $1 - (1 - 0.05)^m$ for $\alpha = 0.05$.

Here m is the number of tests performed. (Herzog et al., 2016, p. 63, p. 64)

3.2.2 Bonferroni Corrections and Adjustment of Significance Level

To tackle multiple testing issues, the project employs the Bonferroni method, a straightforward approach aimed at minimizing Type-I errors typically by adjusting the significance level. For m independent tests, if the desired significance level is set at 0.05, the method involves equating Eq.1 to 0.05 to determine the value of α.

$$\alpha = 1 - (1 - 0.05)^{\frac{1}{m}}$$
$$\alpha = 1 - (0.95)^{\frac{1}{m}} \cong \frac{0.05}{m}$$

To have a Type-I error rate of 0.05 for all m tests, we need:

$$p - value < \frac{0.05}{m}$$

Utilizing the Bonferroni correction necessitates a reduced *p-value* for a statistical test to achieve significance. This method relies on assumptions: data independence, normal distribution, and equality of variance among the data.

3.3 ONE-WAY ANOVA Method

The one-way ANOVA, known as Analysis of Variance, assesses if there's statistical support for significant differences among the means of two or more independent groups within a population. This method divides the data into distinct groups according to a single grouping variable, often referred to as the factor variable.

If we're examining k groups, our hypothesis can be formulated as follows.

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k$$
$$H_1 : \text{At least one of } \mu_i \text{ is different}$$

Like the t-test, ANOVA operates under the assumption of equal variances among all the studied groups. If the null hypothesis holds, it suggests that the population means across all groups are identical, and any observed discrepancies in sample means are attributed to variance.

When the means of all groups vary from each other, individual observations are affected by both within-group variance and the differences between group means. In such instances, the estimated variance based on the variability between group means tends to be larger than the true value. Conversely, the estimate derived from within-group variability is closer to the actual value. In ANOVA, these two estimates are divided to calculate the F-value.

$$F = \frac{\text{Variation between group means}}{\text{Variation within groups}}$$

Mathematically, it is written as:

$$F = \frac{\frac{\sum_{j=1}^{k} n_j (M_j - M_G)^2}{k-1}}{\sum_{j=1}^{k} \frac{\sum_{i=1}^{n_j} (x_{ij} - M_j)^2}{n_j - 1}}$$

In this context, k denotes the number of groups, n_j signifies the count of observations within group j, M_G stands for the mean of all groups combined, M_j represents the mean of group j, and x_{ij} indicates the i^{th} observation within group j.

3.4 Quantile-Quantile Plots (QQ-Plots)

The Quantile-Quantile plot, abbreviated as QQ-plot, is a graphical method employed to assess if a dataset aligns with a theoretical distribution. It serves as a visual indicator to determine if two variables likely share a common distribution. For example, when checking if a variable originates from a normal distribution, one can use a Normal Q-Q plot to validate this hypothesis.

A Q-Q plot displays a scatter of theoretical quantiles against sample quantiles. If both sets originate from the same distribution, we would expect to see the data points aligning relatively close to a straight line.(Heiberger and Burt, 2015, p. 152)

3.5 Levene's Test

Levene's test assesses if the distributions of underlying random variables deviate from symmetry and normality.

Consider a set of examples, each of size n (where j represents $j = 1, 2, \ldots, k$ and for $i = 1, 2, \ldots, n_j$), with observations x_{ij} be the i-th observation for the j-th sample. The mean of the j-th sample, denoted as $\bar{x}_j$ calculates absolute deviations.

$$d_i j = |x_i j - \bar{x_j}|$$

Let $\bar{d}$, represent the mean of all these absolute deviations, $s^2_{d_j}$ stand for variances, and $\bar{d}_j$ denote the sample means of the absolute deviations. The test statistic for Levene's test is then calculated as follows:

$$F^* = \frac{\frac{\sum\limits_{j=1}^{k} n_j (\bar{d}_j - \bar{d})^2}{k-1}}{\frac{\sum\limits_{j=1}^{k} (n_j - 1) s_d j^2}{n-k}}$$

where $F^* \sim F(k-1, n-k)$, and p-value $= P(F \geq F^*)$ (represents the number of tests performed (Levene, H., 1960, p. 278))

3.6 Shapiro-Wilk Test

The Shapiro-Wilk test assesses whether a random sample X_i, where (i ranges from 1 to n), is drawn from a Gaussian probability distribution with a true mean μ_i and variance σ^2, denoted as $X \sim N(\mu_i, \sigma^2)$.

Here, we example the following hypothesis:

$$H_0 : \text{The sample originates from a normal population, } N(\mu_i, \sigma^2)$$
$$\textbf{Vs}$$
$$H_a : \text{The sample does not adhere to } N(\mu_i, \sigma^2)$$

The Shapiro-Wilk test statistic is employed for testing the hypothesis as expressed by:

$$W = \frac{(\sum\limits_{i=1}^{n} a_i x_{(i)})^2}{\sum\limits_{i=1}^{n} (x_i - \bar{x})}$$

Here, $x_{(i)}$ represents the arranged sample values, while a_i denotes constants derived from the given expression.

$$(a_1, a_2, \ldots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} m)^{1/2}}$$

Here,

$$(m = m_1, m_2, ..., m_n)^T$$

'm' in the Shapiro-Wilk test denotes the number of observations in the dataset, and it plays a role in the calculation of the test statistic used to assess the normality of the data.

It represents the anticipated values of the arranged statistics, which are independent and identically distributed random variables following the standard normal distribution $N(0,1)$, while V stands for the covariance matrix of these ordered statistics. (Shapiro, S. S. and M. B. Wilk (1965)).

3.7 Tukey's Test

For each unique pair j and k ranging from 1 to p, where $\bar{y_j} \geq \bar{y_k}$, the equation for all potential $100(1-\alpha)\%$ simultaneous Tukey confidence intervals for the disparities $\mu_j - \mu_k$ is given as follows:

$$(\bar{y_j} - \bar{y_k}) - \frac{1}{\sqrt{2}} q_\alpha s \sqrt{\frac{1}{n_j} + \sqrt{\frac{1}{n_k}}} < \mu_j - \mu_k < (\bar{y_j} - \bar{y_k}) + \frac{1}{\sqrt{2}} q_\alpha s \sqrt{\frac{1}{n_j} + \sqrt{\frac{1}{n_k}}}$$

where s stand for standard error, The critical value

$$q_\alpha > 0 \text{ is such that } P(q \geq q_\alpha) = \alpha \text{ with } q \sim q(p, n-p).$$

'q' in Tukey's test serves as a threshold or critical value used to determine which differences between group means are statistically significant after considering the overall error rate in multiple comparisons.

On the other hand, the relevant test statistic:

$$q*_{ij} = \frac{\sqrt{2}(\bar{y_i} - \bar{y_j})}{s\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}$$

can be used to calculate p-value = P(q q_{jk}). In Tukeys method, both the individual and combined significance levels are deliberately identical. Consequently, the p-value for each pair comparison is assessed against the value α (John W. Tukey, 2003, p. 278).

3.8 Pairwise t Test

The paired t-test also referred to as a dependent test, evaluates the means and standard deviations of two associated groups to establish whether a noteworthy difference exists between them. This difference is considered significant when the variations observed between groups are improbable due to random sampling errors (Dunnett, C. W. 1980). The null and alternative hypotheses for all pairs are presented as follows:

$$H_0 : \mu_i = \mu_j \quad \text{with} \quad i \neq j$$

The alternate hypothesis is defined as:

$$\text{H1} : \mu_i \neq \mu_j \text{ with } i \neq j$$

4 Statistical analysis

In this section, the statistical techniques described earlier are applied to analyze the provided dataset and comprehend the outcomes.

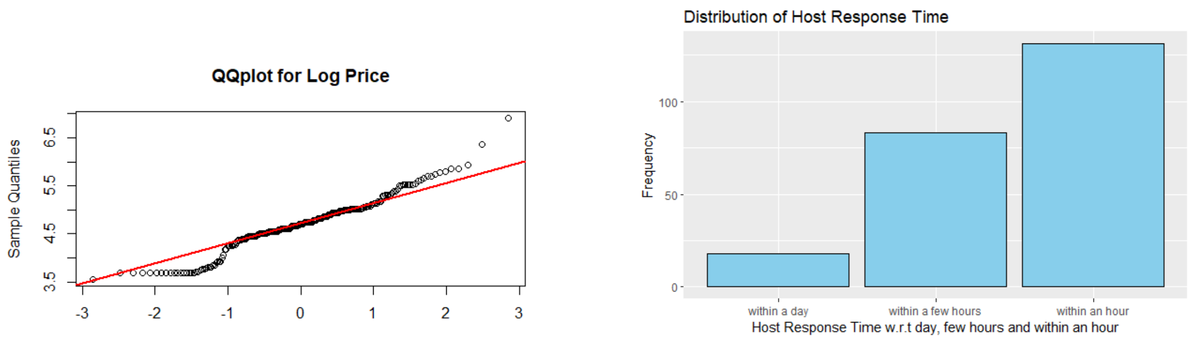
4.1 Descriptive Analysis of the Variables

In this section, we're conducting a descriptive examination of two variables: logprice and host response time. The dataset provides insights into the relationship between log price and different levels of host response time. Table.2 presents a summary of the data showcasing the variations in log price across different response time levels exhibited by hosts.

Table 2: Data summary

Variable name	Min	1st Qu	Median	Mean	3rd Qu	Max	Count
Host response time w.r.t few hours	3.55	4.511	4.654	4.816	5.027	6.907	83
Host response time w.r.t in a day	3.689	4.300	4.725	4.743	5.116	5.768	18
Host response time w.r.t in an hour	3.689	4.309	4.745	4.612	4.977	5.940	131

The summary provided presents key statistics for the log price across different categories. The "Min" column signifies the lowest log price recorded across all categories, standing at 3.55. The 1st quartile, representing the lower range of values, is exemplified by the value below which 25% of the data lies. For instance, within the "host response within a few hours" category, this quartile sits at 4.511. The median and mean values of the log price vary among categories, ranging from 4.645 to 4.745 for the median and from 4.612 to 4.816 for the mean. The 3rd quartile, signifying the upper range of values, is where 75% of the data falls. For instance, within the "host response within an hour" category, this quartile is at 4.977. The "Max" column records the highest log price within any category, ranging from 5.768 to 6.907. The "count" column denotes the frequency of each category within the dataset, with "Host response time within an hour" having the highest count of 131.Fig.1(a) depicts the distribution of log prices, showing nearly linear alignment and most points aligning closely to a reference line, suggesting a normal distribution.



(a) Log Price Distribution

(b) Count - host response time w.r.t categories

Figure 1: Descriptive analysis of variables

Fig.1(b) illustrates the frequency distribution of host response time across its different categories. The category "host response within a day" has the lowest count of 18, while "host response within an hour" has the highest count of 131.

4.2 Global Test using One-Way ANOVA Method

As detailed in section 3.3, a one-way ANOVA necessitates an independent variable with a minimum of three levels and a continuous dependent variable. The dataset contains a categorical variable with three distinct categories and log price as the continuous dependent variable. In this scenario, the null hypothesis suggests no notable distinction among the mean durations of various host response times, while the alternative hypothesis asserts differences exist in the mean durations across one or more types of host response times.

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

Alternative hypothesis can anyone of the following:

$$H_1 : \text{pair}(i, j) \text{ with } i \neq j \text{ so that } \mu_i \neq \mu_j$$

$$H_1 : \mu_i \neq \mu_j$$

Here i and j=1,2,3 which represent host response time within a few hours, within a day and within an hour.

4.2.1 Validating the Assumptions

To verify the assumptions underlying the ANOVA test, the outcomes of both the Levene test and QQ plots are taken into account. The initial assumption of uniform variance is confirmed through the Levene test outcome, which is detailed in Table.3.

Table 3: Results of Levene's test

	Df	F-value	P-value
group	2	1.7549	0.1752

The p-value observed in Table.3, which is 0.1752, exceeds 0.05. Hence, we retain the null hypothesis, indicating uniform variances across groups.

To meet the second ANOVA assumption of observations being independent and identically distributed, entries in the dataset that belonged to multiple categories were eliminated. Lastly, the assessment of normality is addressed through the utilization of QQ-plots.

Figures 2, 3, and 4 in the appendix indicate that a majority of observations align closely with the reference line but not perfectly. This deviation might be attributed to the dataset's limited size. However, with a larger dataset, such deviations could be more distinctly understood. Larger sample sizes typically lead to more consistent plots. Despite not aligning precisely with the reference line, the proximity of data points suggests conformity to a normal distribution within the provided dataset. To complement the QQ-plot, the Shapiro-Wilk normality test is employed. It indicates that if the p-value exceeds 0.05, there's inadequate evidence to dismiss the null hypothesis, suggesting the data reasonably conforms to a normal distribution. However, if the p-value is less than or equal to 0.05, it signifies substantial deviation of the data from a normal distribution.

Table 4: Results of Shapiro-Wilk test

	Shapiro-Wilk test	P-value
few_hours	0.88581	2.246×10^{-6}
in_an_hour	0.9298	3.958×10^{-6}
in_a_day	0.97127	0.8212

For the category "few_hours" of host response time, the p-value stands at 2.246e-06, and for "in_an_hour," it registers at 3.958e-06. As both p-values are below 0.05, there's substantial evidence to reject the null hypothesis, indicating a significant deviation from a normal distribution in both groups. However, for the "in_a_day" category, the p-value surpasses 0.05, implying inadequate evidence to dismiss the null hypothesis. Therefore, the data within this category can be reasonably assumed to adhere to a normal distribution.

After confirming the assumptions, a comprehensive test is carried out to investigate potential weight variations, setting the significance level at 0.05.

From the provided Table.5, it's evident that the F-value stands at 3.648 with a corresponding P-value of 0.0276, which is below 0.05. Therefore, the null hypothesis, asserting

Table 5: Results of ANOVA - Global Test

	DF	Sum Sq	Mean Sq	F-value	P-Value
host response time	2	2.16	1.0785	3.648	0.0276
Residuals	229	67.70	0.2956	-	-

equal mean log prices across all host response time categories, is rejected. This outcome indicates a distinction in the mean log prices within at least one host response time category.

4.3 Pairwise Testing

The outcomes obtained through the one-way ANOVA displayed a statistically noteworthy difference among the group means. Yet, these results don't specify which particular group demonstrates a distinct mean. To address this, a pairwise test is conducted to determine the specific group that differs from the others. The null and alternative hypotheses for all pairs are formulated as follows:

$$H_0 : \mu_i = \mu_j \text{ with } i \neq j$$

Alternate hypothesis is defined as:

$$H_1 : \mu_i \neq \mu_j \text{ with } i \neq j$$

where i and j = 1, 2, 3 are the different categories in host response time.

In Section 4.2.1, we validated assumptions, which are also expected for the pairwise t-test, enabling its execution.

Tukey's test, a method for comparing all potential pairs of means, is utilized. The outcomes of Tukey's test are detailed in Table 6 below.

Table 6 exhibits the outcomes from numerous pairwise comparisons conducted via the Tukey test. Each row signifies a comparison between two groups, while the columns present the difference (diff), lower bound (lwr), upper bound (upr), and adjusted p-value (p adj) for each specific comparison. The analysis indicates that except for the

Table 6: Tukey multiple comparisons of means

	diff	lwr	upr	p adj
within a few hours - within a day	0.07262275	0.2608697	0.40611515	0.8647409
within an hour - within a day	-0.13124826	-0.4536681	0.19117161	0.6027068
within an hour - within a few hours	-0.20387100	-0.3838131	-0.02392892	0.0219001

comparison "within an hour - within a day", all other pairs showcase substantial differences in their means. These pairs exhibit adjusted p-values surpassing 0.05, leading to the failure to reject the null hypothesis for these comparisons.

Table 7: Results of pairwise t-test with P value adjustment method: none

	within a day	within a few hours
within a few hours	0.6079	-
within an hour	0.3379	0.0081

Table 8: Results of pairwise t-test using Bonferroni Correction

	within a day	within a few hours
within a few hours	1.000	-
within an hour	1.000	0.0024

As explained in section 3.2.1, when conducting multiple pairwise tests, there's an increased likelihood of committing a type-I error. To address this concern in the analysis, both Bonferroni Corrections and Tukey's Honest Significant Difference (HSD) method are utilized.

Tables 7 and Table 8 display the results derived from employing pairwise t-tests with p.adjust.method = "none" and Bonferroni Correction tests.

In both Tables 8 and 9, the majority of pair comparisons exhibit statistically significant p-values, indicating substantial differences in their mean values. However, notable observations emerge in the pairs "within a day - within a few hours" and "within a day - within an hour," where the p-values remain insignificant even after the application of correction methods. This suggests that the means of these pairs are comparable, irrespective of the correction applied. It's important to highlight that while the p-values have increased due to the correction methods, there isn't enough evidence to support the null hypothesis for the other pair, specifically "within a few hours - within an hour".

5 Summary

The dataset includes information regarding landlord response time and property pricing. The dataset, sourced from the Introductory Case Studies instructors at TU-Dortmund, consists of 232 observations for two explanatory variables. The analysis investigates the variation in property pricing based on different levels of host response time—categorized as 1 = within a few hours, 2 = within a day, and 3 = within an hour—to determine if significant differences exist in property pricing across these groups.

Initially, the analysis focused on investigating the property pricing distribution concerning host response time through descriptive analysis. Moreover, the frequency of host response time was evaluated across its various categories. Subsequently, a comprehensive test, specifically a one-way ANOVA, was employed to ascertain whether significant mean differences existed among the groups. For pairwise comparisons, both Tukey's Honest Significant Difference (HSD) test and pairwise t-test were utilized. To tackle the challenge of multiple comparisons, Bonferroni Correction and Tukey's HSD were applied. The outcomes from the one-way ANOVA revealed dissimilar means among the groups, resulting in the rejection of the null hypothesis.

During the pairwise tests, it was evident that most pairs, excluding "within a day - within a few hours" and "within a day - within an hour," showed notable disparities in mean weights. After employing Bonferroni Correction and Tukey's HSD, the hypothesis remained consistent with the findings from the pairwise tests. Specifically, only the pairs "within a day - within a few hours" and "within a day - within an hour" displayed analogous mean weights, while all other pairs showcased distinct mean weight variations.

In future studies, the approaches utilized in this project could be broadened to accommodate a more extensive dataset, incorporating extra factors influencing property pricing alterations. For instance, extending the analysis to include property location, property type, and supplementary amenities could enhance the comparison of outcomes. Incorporating these extra variables would contribute to a more holistic comprehension of the factors impacting property pricing changes.

Bibliography

Mustapha Akinkunmi. *Introduction to Statistics Using R.* Morgan and Claypool Publishers, 2019.

John Fox and Sanford Weisberg. *An R Companion to Applied Regression.* Sage, Thousand Oaks CA, third edition, 2019. URL `https://socialsciences.mcmaster.ca/jfox/Books/Companion/`.

Richard M Heiberger and Holland Burt. *Statistical Analysis and Data Display.* Springer, New York, NY, 2015. doi: https://doi.org/10.1007/978-1-4939-2122-5.

Michael Herzog, Gregory Francis, and Aaron Clarke. *Understanding Statistics and Experimental Design.* Springer International Publishing, 2016. doi: 10.1007/978-3-030-03499-3.

Christian Heumann, Michael Schomaker, and Shalabh. *Introduction to Statistics and Data Analysis.* Springer International Publishing, 2016. doi: 10.1007/978-3-319-46162-5.

R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2020.

Dieter Rasch, Rob Verdooren, and Jürgen Pilz. *Applied Statistics: Theory and Problem Solutions with R.* John Wiley and Sons Ltd, 2020.

Adrian Trapletti and Kurt Hornik. *tseries: Time Series Analysis and Computational Finance*, 2023. URL `https://CRAN.R-project.org/package=tseries`. R package version 0.10-55.

Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL `https://ggplot2.tidyverse.org`.

Appendix

A Additional figures

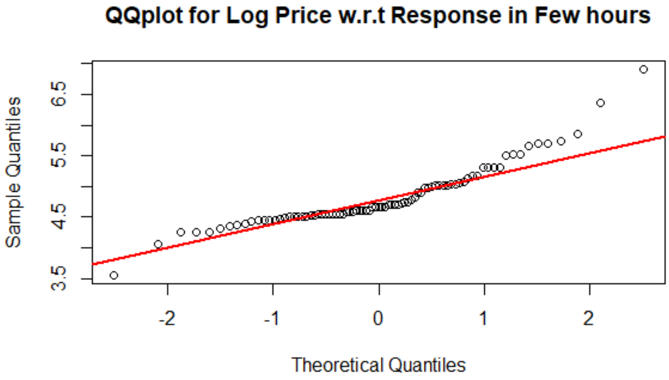


Figure 2: QQplot - host response few hours

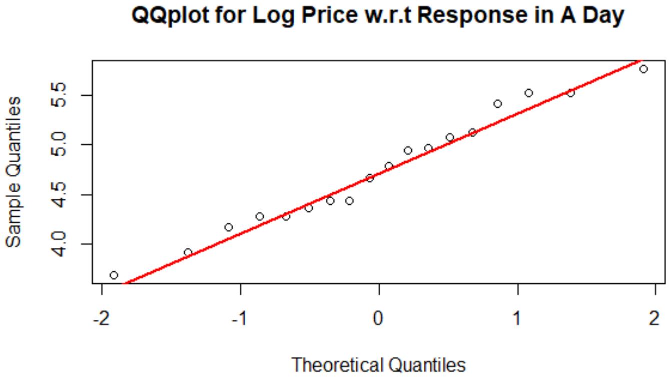


Figure 3: QQplot - host response in a day

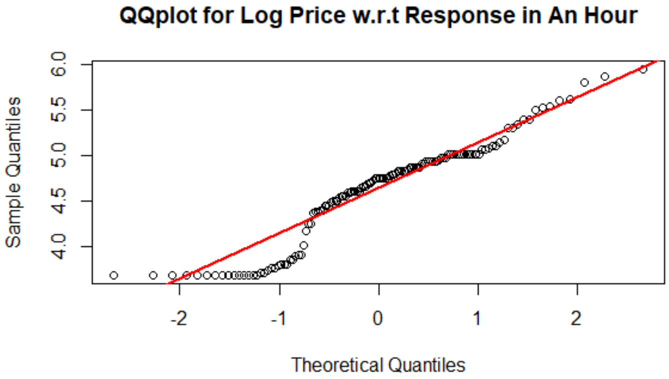


Figure 4: QQplot - host response in an hour