TU DORTMUND

INTRODUCTORY CASE STUDIES

Project 2: Comparison of multiple disribution

Lecturers:

Prof. Dr. Katja Ickstadt

M. Sc. Zeyu Ding

M. Sc. Marvin Napps

Author: Sadat Mahmud

Group number: 2

Group members: Sadat Mahmud, Ahmet Emircan Coskun, Montasir Hasan Chowdhury, Manouchehr Norouzi

December 15, 2023

Contents

1 Introduction

Landlord response time stands as a pivotal factor influencing the guest experience within the realm of property rentals. In addition to that, there are several factors related to host response time such as it helps to improve visibility, booking confidence, enhanced ratings and reviews, etc. Fast response time places a host or property owner one step ahead of others. According to one of the largest property rental online marketplaces Airbnb, " A Faster response rate helps to propel a listing further in a sea of other hosts" (IGMS (2017)). Hence, It is important to analyze the influence of response time on property pricing.

This project endeavors to investigate the relationship between host responsiveness and property pricing, recognizing its crucial significance for both hosts and guests in the context of rentals.

First, descriptive statistics is used to analyze the distribution of response time and pricing and conduct two tests to investigate the central tendency of different groups. Subsequently, a global test is conducted to identify price differences across various categories of host response times. To examine the pairwise differences in prices among separate response times, all categories were included, and two-sample t-tests were employed for each pair. Conducting multiple tests increases the likelihood of false positives. To address this issue, the Tukey's Honestly Significant Difference (HSD) and Bonferroni correction methods were applied. Ultimately, the results were compared across various correlation methods and non-adjusted tests, accompanied by an explanation for any observed differences.

In section 2, The explanation covers both the dataset and the project objectives. The detailed discussion encompasses the data source, the method employed for data collection, an evaluation of data quality, and an examination of data types.

In Section 3, all the relevant statistical methods in this report are explained. The discussion encompasses the Levene test, Kolmogorov-Smirnov test, Shapiro-Wilk test, analysis of variance (ANOVA), pairwise t-tests, Tukey's Honestly Significant Difference (HSD) test, and the Bonferroni method, all of which were used in this project.

In Section 4, The results of the various statistical methods are analyzed and visualized using different graphical plots such as scatter plots and bar plots.

In Section 5, a comprehensive summary of the entire report and its results is provided.

2 Problem statement

2.1 Overview of Dataset

The dataset used in this project was provided by the course instructors of the 'Introductory Case Studies' course at Technische Universität Dortmund during the winter of 2023/2024. The data was sourced from Inside Airbnb, a well-known platform that compiles and provides Airbnb-related datasets for various locations. Inside Airbnb is widely recognized for its commitment to transparency and accessibility in making such datasets available for research and analysis (Inside Airbnb (2016)). The dataset focuses on information related to Seattle and is dated January 2016. The provided dataset is named "Airbnb_data.csv". The dataset "Airbnb_data.csv" is a small relevant extract for this project from the original dataset. The given dataset contains information on the log price and host response time of Seattle for the year 2016. The given dataset consists of a total of 232 observations, providing a comprehensive snapshot of the Airbnb market in Seattle during the specified period. The dataset is organized into two variables and the variables are "log_price" and "host_response_time". the variable log_price represents the logarith of the price for the Airbnb listings and the variable host_response_time captures the host response time. A logarithm is the exponent to which a certain base must be raised to get a specific number. This mathematical idea shows the power required to transform one number into another through exponentiation (Coordinator (2004)). This structured format facilitates a focused analysis of the chosen aspects relevant to the project. Notably, the provided dataset is free from missing values, as any such instances have been meticulously addressed and removed by the course instructor.

2.2 Project Objective

The main objective of this project is to conduct a comprehensive analysis of the provided dataset, focusing on the variables "log_price" and "host_response_time". First, descriptive statistics were performed to gain insights into the distribution of the variables "log_price" and "host_response_time". The frequency distribution of "host_response_time" among different categories is visualized using graphical plots. In order to rigorously assess the assumptions underlying our analysis, a multi-step validation process is undertaken. First and foremost, the Levene test is employed to scrutinize potential differences in variance across distinct groups. Subsequently, the normality of variance assumption is meticulously examined through the Shapiro-Wilk test. To investigate significant differences in log_price values across diverse categories of host_response_time, a global test, specifically the one-way ANOVA, is conducted. This analysis is designed to determine whether there are overall variations in log_price among the different host response time groups. Additionally, a pairwise comparison is undertaken using T-tests to ascertain potential differences between categories concerning "log_price". To explain the results of the conducted tests, we adopted a p-value approach. Subsequently, to address issues arising from multiple comparisons, both the Bonferroni method and Tukey's Honest Significant Difference (HSD) test were employed. In conclusion, a thorough comparative analysis is undertaken to scrutinize the outcomes across multiple tests and the Bonferroni correction. The overarching aim of this project is to perform descriptive statistics to provide a nuanced understanding of the relationship between "log_price" and "host_response_time".

3 Statistical methods

In this chapter, all the statistical methods are discussed which we used in this project. The Python programming language (Python Software Foundation (2023)) was employed, while the specific computational environment utilized was Jupyter Notebook (Jupyter Development Team (2022)). This selection allowed for the seamless integration of key libraries and packages such as Pandas (pandas development team (2023)), NumPy (NumPy Community (2023)), Matplotlib (Matplotlib Development Team (2023), Seaborn (Seaborn Development Team (2023)), Scipy (Jones et al. (2001)), Itertools (Itertools Software Foundation (2003)), Statsmodels (Statsmodels Development Team (2022)), and Pingouin (Pingouin Development Team (2022)) fostering an environment conducive to interactive development and comprehensive code documentation.is juyter notebook (Jupyter Development Team (2022)).

3.1 Null Hypothesis (H_0) and alternative hypothesis (H_1)

In many statistical analyses, the formulation and testing of hypotheses play a crucial and common role, offering valuable insights into the data. Typically, data is analyzed under the assumption of a specific outcome. This assumption, known as a hypothesis, serves as a foundational element in statistical analysis, enabling us to assess the reasonableness of the assumed result. If the analysis reveals that the assumption is untenable, we reject the hypothesis and explore alternative hypotheses. It's a systematic process wherein the validity of our assumptions is rigorously examined (Rees (2000) p. 139-141). Statistical hypothesis tests are employed to assess and validate these hypotheses. There are two types of hypothesis commonly referred to as null hypothesis H_0 and alternative hypothesis H_1. The null hypothesis H_0 generally posits the absence of difference. When the observed value aligns with the assumed result, it is defined as the null hypothesis H_0 (Rees (2000) p.139-141). Conversely, the alternative hypothesis H_1 suggests a difference between the assumed and actual results (Rees (2000) p. 139-141). If null hypothesis H_0 is rejected, an alternative hypothesis H_1 must be considered, and vice versa.

3.2 Significance value (α) and p-value

A significance level represents the threshold at which a statistician is willing to accept the risk of rejecting the null hypothesis, H_0 (Rees (2000) p. 141). It serves as the predetermined criterion, guiding the decision-making process in hypothesis testing. A significance level is usually defined before conducting the test. The significance level is set at 0.05%, signifying the threshold at which the null hypothesis is considered for rejection. There is a 5% probability of opting for the alternative hypothesis under the assumption that the null hypothesis is true (Rees (2000) p. 141-142). In this project, the significance level is 0.05 and is denoted by α.
The p-value is a statistical measure used to determine whether the null hypothesis can be accepted or rejected. If the p-value is less than the significance level, then the null hypothesis is rejected. Conversely, if the p-value is greater than the significance level, the null hypothesis is accepted (Rees (2000) p. 145-16). The significance level and the p-value collectively serve as the deciding factor for a given hypothesis. The rejection of a null hypothesis implies a statistically significant result, while the failure to reject a null hypothesis suggests a lack of statistical significance (Rees (2000) p. 145-146).

3.3 Statistical tests and assumptions

In this project, two distinct tests were conducted: a one-way ANOVA and a t-test. To refine the results from both tests, Bonferroni correction and Tukey's Honestly Significant Difference (HSD) were applied. Ensuring the validity of the tests required careful control of assumptions. Consequently, the equality of variances across groups was essential, and the residuals needed to exhibit a normal distribution. A thorough exploration of the intricacies of these tests will be presented in the following sections.

3.4 Shapiro-Wilk Test

The Shapiro-Wilk is a formal test to check residual normality (Elliott et al. (2017) p. 380). The Shapiro-Wilk test is a hypothesis test used to assess whether a sample originates from a normal distribution. The null hypothesis posits that the sample follows a normal distribution. A low p-value leads to rejecting this null hypothesis, indicating that the sample does not exhibit characteristics of a normal distribution. The Shapiro-Wilk test is considered a straightforward tool for assessing normality. However, one limitation of this test is its inability to handle large datasets (Malato (2023)).

3.5 Assessing Homogeneity of Variances with Levene Test

The Levene's is a very commonly used statistical method to check the homogeneity of variances with different groups or samples. It use to determine whether there are any notable differences in variability among groups. The Levene test holds significance as it constitutes an assumption in numerous statistical examinations, including the t-test and ANOVA. The null hypothesis, symbolized as H_0, posits that the variances within the populations of all k groups are identical. Conversely, the alternative hypothesis, denoted as H_1 states that there exists at least one group with a population variance distinct from the others. The mathematical expression of the hypothesis is presented as follows:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$$

$$H_1 : \sigma_i^2 \neq \sigma_j^2 \quad \text{for at least one pair } i \text{ and } j$$

Suppose j ranges from 1 to k ($j = 1, 2, 3, \ldots, k$), illustrating each sample. The size of the j^{th} sample is denoted by n_j, and y_{ij} signifies the i^{th} observation within the j^{th} sample. The mean of the j^{th} sample is denoted as $\bar{y}_j$. Additionally, the absolute deviation is calculated as $\Delta_{ij} = |y_{ij} - \bar{y}_j|$. Assuming $n = \sum_{j=1}^{k} n_j$, which signifies the overall size encompassing all samples combined. Let $\bar{\Delta}$ stand for the average absolute deviation, whereas the sample mean and variance for the absolute deviations are expressed as $\bar{\Delta}_j$

and $s^2_{\Delta j}$ respectively. The test statistic is formulated as follows:

$$F^* = \frac{\sum_{j=1}^{k} n_j(\bar{\Delta}_j - \bar{\Delta})^2(n-k)}{\sum_{j=1}^{k}(n_j-1)s^2_{\Delta j}(k-1)}$$

In this context, the p-value, denoted as $P(F \geq F^*)$, is determined based on the distribution of F^*, which follows an F-distribution with degrees of freedom $k-1$ and $n-k$. If the computed p-value exceeds the pre-defined significance level of 0.05, we retain the null hypothesis, indicating that there is no sufficient evidence to reject it. This suggests the presence of equal variances among the groups. Conversely, if the p-value is less than or equal to 0.05, we reject the null hypothesis, signifying unequal variances among the groups (Hay-Jahans p. 247-248).

3.6 ANOVA

ANOVA is a statistical analysis employed to assess the variance in means among multiple data samples or groups, typically exceeding two. The test can be categorized as either one-way or two-way ANOVA. In this report, we will opt for one-way ANOVA, given the presence of only one quantitative independent variable. Two-way ANOVA is applied for two independent variables. In the context of ANOVA with k samples, the examination involves testing the following hypotheses:

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

$$H_1 : \text{At least one of the means is different from the others.}$$

If the mean of one sample differs from the others, we reject the null hypothesis, indicating a significant difference among data samples. To perform the test, We initially assessed variances both between and within groups. The variance between groups (SSB) is obtained by subtracting the variance within groups (SSW) from the total variance (SST).

$$SST = \sum_{i=1}^{n}\sum_{j=1}^{k}(x_{i,j} - \bar{X})^2$$

$$SSW = \sum_{k}^{j=1}\sum_{n}^{i=1}(x_{i,j} - \bar{x}_j)^2$$

Here, represents the grand mean. The grand mean is formulated as:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k} x_{i,j}$$

Let x_{ij} denote the individual data points for each group, where k represents the count of groups or data samples. The symbol $\bar{x}_j$ stands for the mean value within group j, and n signifies the overall sample size. Moving forward, the Mean Square Between Groups (MSB) and Mean Square Within Groups (MSW) are determined through the subsequent formula:

$$MSW = \frac{SSW}{\sum_{j=1}^{k}(n_j - 1)} = \frac{SSW}{n-k} = \frac{SSW}{df_w}$$

$$MSB = \frac{SSB}{k-1} = \frac{SSB}{df_b}$$

In this context, the concept of freedom degree is defined as the utmost count of logically autonomous values permitted to exhibit variation. The count of freedom degrees among groups, labeled as df_b, equals $k-1$, with k denoting the group count. Conversely, df_w stands for the within-group freedom degrees, corresponding to $n-k$, where n signifies the overall sample size (Tukey (1949) p. 230-244). Ultimately, the F-statistic, denoted as F^*, is the ratio of the two variances.

$$F^* = \frac{MSB}{MSW}$$

To assess whether the null hypothesis is declined, the p-value method proves useful. In the context of ANOVA, the p-value is associated with the probability $P(F \geq F^*|H_0)$, where F is a random variable drawn from the $F(df_w, df_b)$ distribution at the significance level α. If the chosen significance level exceeds the obtained p-value ($\alpha > p$), the null hypothesis is rejected, indicating a statistically significant result. Conversely, if the significance level is less than the p-value ($\alpha < p$), the null hypothesis is retained, and consequently, the outcome is deemed statistically insignificant (Black et al. (2018), p. 404-4114).

3.7 Pairwise t-test

Through the examination of the mean values across groups in the preceding section, the ANOVA test primarily determines whether it rejects the null hypothesis. However, this test does not furnish details on which specific pairs of groups exhibit similar mean values and which pairs differ. To discern significant differences in mean values between pairwise groups or categories, a pairwise t-test becomes imperative. Analogous to the ANOVA test, the assumptions for this test are established at the outset of this section. For all pairs i, j, where $i, j = 1, \ldots, k$, the following hypothesis characterizes this test:

$$H_0 : \mu_i = \mu_j$$

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

To derive the t-statistic, it is necessary to calculate a pooled sample standard deviation (S_p), representing a weighted average of standard deviations drawn from two or more independent groups or data samples. The formula for the pooled standard deviation, applicable to more than two independent data samples, is expressed as follows:

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2}{n_1 + n_2 + \ldots + n_k - k}$$

If $n_1, n_2, \ldots, n_k$ represent the sample sizes of group k, and $s_1, s_2, \ldots, s_k$ denote the standard deviations of the k groups, assuming the equality of pooled standard deviations ($\sigma_1 = \sigma_2$), The pairwise t-test can be formulates as:

$$t^* = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2+\ldots+(n_k-1)s_k^2}{n_1+n_2+\ldots+n_k-k}} \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}$$

The variable $\bar{x}_i - \bar{x}_j$ represents the mean difference between two distinct groups. Upon obtaining the t-statistic, to ascertain if the null hypothesis can be rejected, one may employ the p-value method. In the context of a two-tailed pairwise t-test, the p-value is associated with the probability $2P(T \geq |t^*| \mid H_0)$, where T is a random variable following the $T_{n_1+n_2+\ldots+n_k-k}$ distribution at the significance level α. If the preferred significance level surpasses the computed p-value ($\alpha > p$), the null hypothesis is rejected, signifying a statistically significant outcome. Conversely, if the significance level is below the p-value ($\alpha < p$), the null hypothesis is retained, and thus, the result is deemed statistically inconclusive (Hay-Jahans p. 260-262).

3.8 Bonferroni method

In the context of hypothesis testing, the erroneous rejection of the null hypothesis is referred to as a Type I error or false positive. Conversely, when we fail to reject the null hypothesis erroneously, such a circumstance is designated as a Type II error or false negative El-gohary (2019). When numerous statistical tests, whether dependent or independent, are conducted simultaneously, there is a substantial escalation in the false positive rate. Consequently, the Bonferroni correction is employed to mitigate the occurrence of false positives. Consider a collection of hypotheses denoted as $H_1, \ldots, H_m$, each associated with corresponding p-values $p_1, \ldots, p_m$. Let m_0 represent the count of true null hypotheses, where m signifies the overall number of null hypotheses. The familywise error rate (FWER) pertains to the probability of committing at least one Type I error, specifically, the probability of erroneously rejecting at least one true hypothesis H_i. The calculation of the familywise error rate is expressed as follows:

$$\text{FWER} = P(\text{at least one Type I error}) = 1 - P(\text{no Type I errors})$$

The familywise error rate (FWER) is defined as the probability of committing at least one Type I error. To manage the Type I error rate in the context of multiple tests, a recalibrated significance level (α_{adj}) is determined for each individual test:

$$\alpha_{\text{adj}} = \frac{\alpha}{m}$$

Here, α represents the overall significance level, and m is the total number of hypotheses under examination. Employing this adjusted significance level (α_{adj}) for each specific hypothesis test aids in minimizing the risk of a Type I error while simultaneously governing the collective familywise error rate. The familywise error rate (FWER) is effectively managed by recalibrating the significance level (α_{adj}) for individual tests. However, it is essential to acknowledge that this method, while significantly reducing Type I errors, may elevate the susceptibility to Type II errors. To navigate this trade-off, where the reduction in Type I errors may increase the likelihood of Type II errors, a careful consideration of the adjusted significance level (α_{adj}) becomes pivotal for the validity of hypothesis testing in multiple scenarios (Field (2013) p. 275-276).

3.9 Tukey's HSD and Confidence level

Tukey's HSD (Honestly Significant Difference) test is a post-hoc analysis designed for conducting numerous pairwise comparisons among group means while effectively managing the overall type 1 error rate. Formulated for all pairs i, j where $i, j = 1, ..., k$, the hypothesis for this examination is expressed as:

$$H_0 : \mu_i = \mu_j$$

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

To determine Tukey's Honestly Significant Difference (HSD), the formula for the test statistic is given by:

$$Q_{i,j} = \frac{\sqrt{2}(\mu_i - \mu_j)}{s^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \quad \sim q(k, n-1)$$

This formula computes the test statistic $Q_{i,j}$, where μ_i and μ_j represent the means of groups i and j, respectively. The variables s, r, n_i, and n_j correspond to the standard deviation, number of treatments, and sample sizes for groups i and j. The test statistic follows a $q(k, n-k)$ distribution. To assess the p-value ($P(q \geq Q_{i,j})$), the test statistic relies on the studentized range distribution, denoted by q. The studentized range distribution, widely employed in statistical hypothesis testing, especially for the comparison of multiple groups as seen in procedures like the Tukey's Honestly Significant Difference (HSD) test, is instrumental in statistical analyses (StatisticsHowTo.com (December 15, 2023)). The resulting p-value is below the predetermined significance level of 0.05, it suggests a statistically significant distinction between the two specific groups. Conversely, if the p-value exceeds 0.05, no significant difference is observed. The formulation for this is expressed as:

$$\text{p-value} = P(q \geq Q_{i,j})$$

A confidence interval (CI) stands as a vital statistical instrument, furnishing a range of values wherein the true population parameter is reasonably anticipated to reside. A 95% confidence interval is commonly employed, signifying a 95% level of assurance. This analysis is contextualized by the widespread use of a 95% confidence interval, denoting a 95% level of assurance in our estimation of the true population mean difference in host response time. In the absence of standard deviation data, alternative approaches can be employed to compute confidence intervals (CIs). One such method is the Tukey's Honestly Significant Difference (HSD) test. This test computes confidence intervals for pairwise comparisons of group means and centers its attention on the distribution of the studentized range, commonly denoted as the 'q' distribution. When $\bar{y}_i > \bar{y}_j$, the Tukey's confidence interval for the difference $(\bar{y}_i - \bar{y}_j) - \mu_i - \mu_j$ is typically calculated as follows:

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

The expression

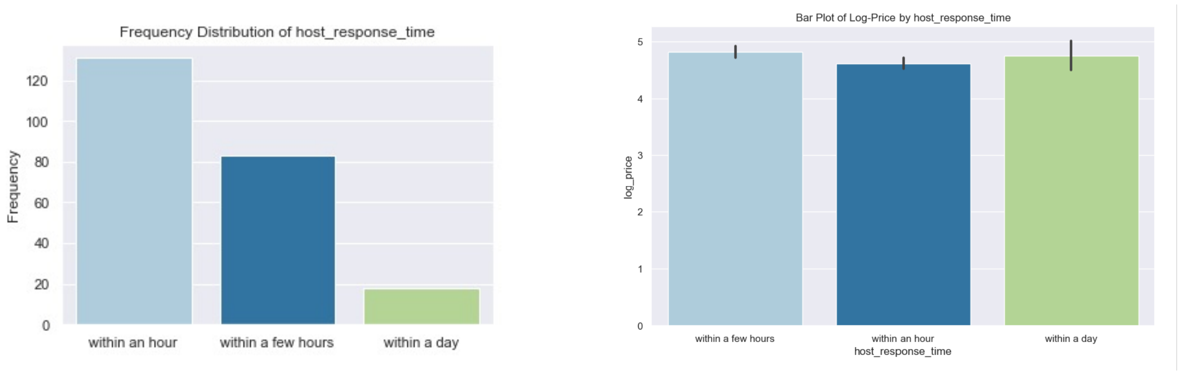
$$(\bar{y}_j - \bar{y}_k) \pm \text{margin of error}$$

establishes a confidence interval for the difference between the population means μ_j and μ_k. Key components include the sample means $\bar{y}_j$ and $\bar{y}_k$, the critical value q_α accounting for multiple comparisons, the standard error s measuring mean differences' variability, and the adjustment for different sample sizes. This interval offers a reliable range where the true difference between population means is likely to reside, adjusted for statistical variability and multiple comparisons. The critical value $q_\alpha > 0$ is determined by the condition $P(q \geq q_\alpha) = \alpha$, where q follows a $q \sim q(p, n-p)$ distribution. This ensures that the probability of observing a value q greater than or equal to q_α is equal to the chosen significance level α (Hay-Jahans p. 276).

4 Statistical Analysis

4.1 Descriptive analysis of log price and host response time

Descriptive statistics were conducted to characterize the variables log price and host response time within the dataset. The analysis utilized two bar plots to visually represent the findings. Figure 1 illustrates the frequency distribution of response times across different categories. The bar chart reveals that "Within an Hour" is the most frequent response time, comprising 56.47%. Following closely is "Within a Few Hours" at 38.78%, and "Within a Day" has the lowest frequency, accounting for 7.76%. Figure 1 Bar chart log price by host response time displays the mean log price for each response time category. "Within a Few Hours" exhibits the highest mean log price, succeeded by "Within a Day" and "Within an Hour." The bar plot depicting the relationship between price and host response time exhibits three small lines positioned at the top-middle section of each bar. These lines represent the confidence intervals around the mean log price for respective response time categories. The length of each error bar is determined by the standard deviation (std), which is calculated as 0.55. Additionally, the sample size (count) utilized for this analysis is 238.0. This presentation offers insights into the variability and precision of the mean log prices within each response time category. Descriptive statistics for log prices are presented in Table 1.



(a) Frequency dist. of host response time (b) Bar plots of price by host response time

Figure 1: Bar plots of Frequency distribution of host response and price by host response

Variable	Count	Mean	Std	Min	25%	50%	75%	Max
log_price	232.0	4.7	0.55	3.56	4.44	4.7	5.01	6.91

Table 1: Descriptive Statistics for log_price

4.2 Comparison of log price across host response time categories using global test

A global test called one-way ANOVA is used to analyze the differences between log price and categories of host response time. Before conducting the global test it is necessary to do assumption checks such as equal variance assumption and normality assumption. Levens test is used to investigate the equal variance assumption and Shapiro-Wilk test is conducted to verify normality check. Levene's test was conducted to assess whether the variance of log price differs between the host response time categories. The results show no significant difference in variances across categories (Levene's Test Statistic = 1.755, p-value = 0.175). Shapiro-Wilk tests were performed for each host response time category to check the normality of log price within each category. Results show that log price is normally distributed for 'within an hour' (p-value = 0.000004) and 'within a day' (p-value = 0.821), while 'within a few hours' deviates from normality (p-value = 0.000002). The counts corresponding to different categories of host response time are as follows: "Within a Few Hours" has 18 occurrences, "Within an Hour" has 83 occurrences, and "Within a Day" has 131 occurrences. This information provides a succinct representation of the sample sizes within each host response time category. A one-way ANOVA was conducted to test whether there is a significant difference in the mean log price across the three host response time categories. The ANOVA results indicate a significant difference (F-statistic = 3.648, p-value = 0.028), rejecting the null hypothesis. Table 2 presents the results of the ANOVA test.

Parameter	**Value**
Number of Groups (k)	3
Degrees of Freedom Between Groups (df_1)	2
Total Number of Observations (n)	232
Degrees of Freedom Within Groups (df_2)	229
Significance Level (α)	0.05
Critical F-value	3.00

Table 2: Result of ANOVA Test

4.3 Pairwise Differences in Log Prices by Host Response Time Categories

To explore potential pairwise differences in log prices across various host response time categories, we conducted two-sample tests, adjusting the outcomes using both the Bonferroni correction and Tukey's Honest Significant Difference (HSD) method. The key findings are summarized in Table 3 and Table 4. Tukey's HSD test revealed a significant difference in mean log prices specifically between "within a few hours" and "within an hour." However, all pairwise comparisons failed to show statistically significant differences after Bonferroni correction. The Bonferroni-corrected T-tests did not identify statistically significant variations among host response time categories. These results offer nuanced insights into the pairwise fluctuations in log prices, underscoring the importance of meticulous consideration and correction for multiple comparisons.

Table 3: Pairwise T-Test Results for Host Response Time Categories

Comparison	P-value	Bonferroni Corr. P-value	Reject Null Hypothesis
Within a Day vs. Within a Few Hours	0.633176	1.0	False
Within a Day vs. Within an Hour	0.38457	1.0	False
Within a Few Hours vs. Within an Hour	0.006415	0.019244	True

Table 4: Tukey's HSD Test Results for Host Response Time Categories

Group 1	Group 2	Mean Diff.	P-Adj.	Reject Null Hypothesis
Within a Day	Within a Few Hours	0.0726	0.8647	False
Within a Day	Within an Hour	-0.1312	0.6027	False
Within Few Hours	Within an Hour	-0.2039	0.0219	True

4.4 Comparing the results among two correlation methods

To compare the results of the two correction methods (Bonferroni and Tukey's Honest Significant Difference) with the non-adjusted test, we can analyze the outcomes of the pairwise tests conducted among different host response time categories. The statistical tests aimed to identify potential differences in mean log prices. After the Bonferroni correction, no statistically significant differences in mean log prices were found among host response time categories. $p > 0.05$ for all comparisons. The adjusted p-values surpassed the chosen significance level, indicating a conservative correction that mitigates Type I error risk. Emphasizing a cautious approach, the Bonferroni correction yielded no significant differences, prioritizing the control of Type I errors but potentially leading to an increase in Type II errors.

Tukey's HSD test identified a significant difference solely between "within a few hours" and "within an hour" as $p = 0.0192$. This pinpointed distinction implies that, even with correction for multiple comparisons, a notable divergence in mean log prices exists for this specific pairwise comparison. Less conservative than Bonferroni, Tukey's HSD method identified a significant difference between "within a few hours" and "within an hour," emphasizing its ability to detect specific contrasts even with correction for multiple comparisons.

The initial non-adjusted test indicated a significant difference in mean log prices between host response time categories with $p = 0.064$. However, without correction for multiple comparisons, there is an increased risk of Type I errors, potentially yielding false positives. The initial non-adjusted test pointed to significant differences but is prone to inflated Type I errors. The observed significance may be influenced by the increased risk of false positives inherent in multiple testing. The nuanced differences in outcomes emphasize the delicate balance between statistical rigor and sensitivity when applying correction techniques in the context of multiple comparisons.

4.5 Summary

In this project, we investigated the relationship between host response time categories and log prices of accommodations using data from a website called Inside Airbnb, affiliated with the online property rental marketplace Airbnb. With 232 observations, the dataset categorized response times as 'within a few hours,' 'within an hour,' and 'within a day.' Our one-way ANOVA test indicated a significant difference in mean log prices among response time categories ($p < 0.05$). Pairwise t-tests revealed specific differences, notably between 'within a few hours' and 'within an hour' ($p = 0.0064$). To address Type I error risks in multiple testing, we applied Bonferroni and Tukey's HSD corrections. Bonferroni showed no significant differences post-adjustment, reflecting its conservative nature. Conversely, Tukey's HSD identified a significant difference between 'within a few hours' and 'within an hour' ($p = 0.0192$), demonstrating sensitivity even with correction. The non-adjusted test suggested a significant difference in mean log prices ($p = 0.064$), cautioning against increased Type I errors. The study sheds light on nuanced log price variations based on host response times, emphasizing the importance of correction techniques. While Bonferroni controls Type I errors, Tukey's HSD exhibits greater sensitivity. Although providing valuable insights for Airbnb hosts and users, the study is constrained to 'log price' and 'host response time.' To enrich findings and comprehend pricing dynamics comprehensively, future studies should explore variables like distance from the city center, category amenities, and more.

Bibliography

Ken Black, John Asafu-Adjaye, Paul F Burke, Nazim Khan, Gerard King, Nelson Perera, Andrew Papadimos, Carl Sherwood, and Saleh Ahmed Wasimi. *Business Analytics and Statistics*. John Wiley & Sons Australia, Limited, 2018.

Distant Education Coordinator. Unversity of minnesota, 2004. URL `http://www.mclph.umn.edu/mathrefresh/logs.html`. Accessed: December 15, 2023.

Tarek M El-gohary. Hypothesis testing, type i and type ii errors: Expert discussion with didactic clinical scenarios. *International Journal of Health and Rehabilitation Sciences (IJHRS)*, 8(3):132, 2019.

Alan C. Elliott, Harry L. Gray, and Wayne A. Woodward. *Applied Time Series Analysis with R*. 2017. ISBN 9781498734226, 1498734227.

Andy Field. *Discovering Statistics Using IBM SPSS Statistics*. SAGE, 2013.

Christopher Hay-Jahans. *R Companion to Elementary Applied Statistics*. Chapman Hall/CRC. ISBN 9781138329164.

IGMS. Airbnb response rate, 2017. URL `https://www.igms.com/airbnb-response-rate/`. Accessed: December 15, 2023.

Inside Airbnb. Inside Airbnb, 2016. URL `http://insideairbnb.com/`. Accessed: December 15, 2023.

Itertools Software Foundation. Python itertools Module, 2003. URL `https://docs.python.org/3/library/itertools.html`.

Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for python. `https://www.scipy.org/`, 2001.

Jupyter Development Team. Jupyter metapackage. Install all the Jupyter components in one go, 2022. URL `http://jupyter.org`. Jupyter Notebook Version: 1.0.0.

Gianluca Malato. Shapiro-wilk test, 2023. URL `https://builtin.com/datascience/shapiro-wilktest`. Published on May 2, 2023.

Matplotlib Development Team. *Matplotlib Documentation*. Matplotlib Development Team, 2023. URL `https://matplotlib.org/stable/contents.html`.

NumPy Community. *NumPy Documentation*. NumPy Community, 2023. URL `https://numpy.org/doc/stable/`.

pandas development team. *pandas Documentation*. pandas development team, 2023. URL `https://pandas.pydata.org/pandas-docs/stable/`.

Pingouin Development Team. Pingouin: Statistical package for Python, 2022. URL https://pingouin-stats.org/.

Python Software Foundation. *Python Language Reference*. Python Software Foundation, 3.11.3 edition, Apr 2023. URL https://docs.python.org/3/reference/. Version 3.11.3 (tags/v3.11.3:f3909b8, Apr 4 2023, 23:49:59) [MSC v.1934 64 bit (AMD64)].

D.G. Rees. *Essential Statistics*. Chapman & Hall Texts in Statistical Science Series. Chapman & Hall, 4th edition, 2000. ISBN 978-1-5848-8007-3. Includes bibliographical references and index.

Seaborn Development Team. *Seaborn Documentation*. Seaborn Development Team, 2023. URL https://seaborn.pydata.org/.

StatisticsHowTo.com. Tukey test / tukey procedure / honest significant difference, December 15, 2023. URL https://www.statisticshowto.com/probability- and-statistics/statistics-definitions/post-hoc/tukey-test-honest-significant-difference/.

Statsmodels Development Team. Statsmodels: Econometric and statistical modeling with Python, 2022. URL https://www.statsmodels.org/stable/index.html.

John W. Tukey. One degree of freedom for non-additivity. *Biometrics*, 5(3), 1949.

Appendix

A Additional figures

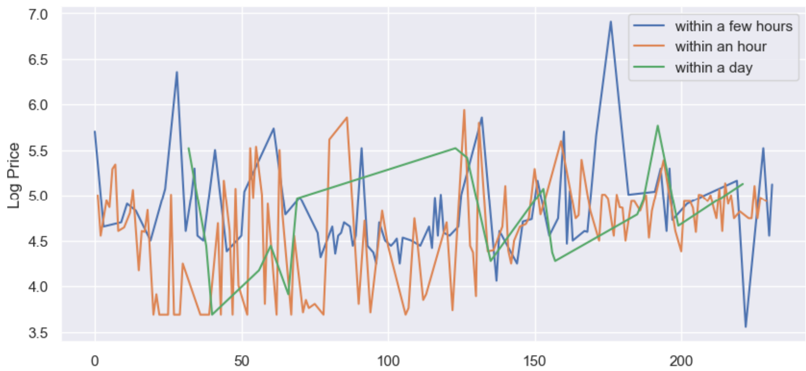


Figure 2: Log-price for each category to show the effect on the log transformation on the data variation reduction and data stabilization

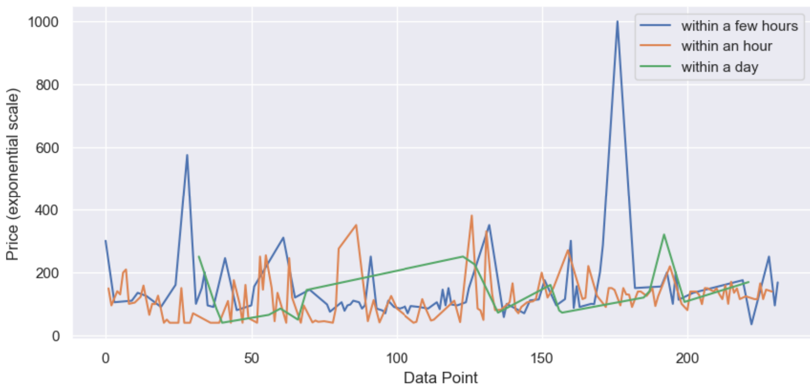


Figure 3: Log-price for each category to show the effect of the log transformation on the data variation reduction and data stabilization

B Additional tables

Table 5: Log Prices Across Host Response Time Categories

Host Response Time	**Log Price**
Within a Few Hours	5.703782
Within an Hour	5.003946
Within an Hour	4.553877
Within a Few Hours	4.653960
Within an Hour	4.941642

Table 6: Summary Statistics for Log Prices Across Host Response Time Categories

Host Response Time	**Count**	**Mean**	**Std**	**Min**	**25%**	**50%**	**75%**	**Max**
Within a Day	18.0	4.74	0.59	3.69	4.30	4.73	5.12	5.77
Within a Few Hours	83.0	4.82	0.50	3.56	4.51	4.65	5.03	6.91
Within an Hour	131.0	4.61	0.56	3.69	4.31	4.74	4.98	5.94

Table 7: Tukey's HSD Test Results for Host Response Time Categories

Group 1	Group 2	Mean Difference	P-Adjusted	**Lower**	**Upper**	**Reject**
Within a Day	Within a Few Hours	0.0726	0.8647	-0.2609	0.4061	False
Within a Day	Within an Hour	-0.1312	0.6027	-0.4537	0.1912	False
Within a Few Hours	ithin an Hour	-0.2039	0.0219	-0.3838	-0.0239	True