The Price of Credibility

A Technical Analysis (Statistical Inference w/ Python code and statistics) of eBay auctions for the video game “Mario Kart” (Oct 2009)

Chris Pratt’s Mario

Introduction

In this report, I will explore the auctions for the game Mario Kart for the Nintendo Wii on eBay and analyze the effects of seller rating on the total price. I will specifically focus on whether there is evidence to suggest that auctions with higher seller ratings have a different average total price than those with lower seller ratings. By conducting a difference of two means test on these two variables, I can determine whether there is a statistically significant difference in the average total price of auctions with different seller ratings. This analysis can provide insight into the factors that influence the total price of a video game on eBay. This is relevant because previous research by Cswenson of Medill Spiegel Research Center (2021) has suggested that seller ratings may impact purchase decisions and seller power.

Dataset

This dataset was obtained from OpenIntro as exercise data, sampling of information from eBay during a week in October 2009. It contains information on auctions for the game Mario Kart for the Nintendo Wii. I displayed the first few rows of the dataset in Appendix A. The data set includes information on auction id, listing name, and more. One notable aspect of this data set is the presence of two outliers shown in Appendix B, which were removed from the data set because they included other items in their auctions besides the game Mario Kart, artificially increasing the total price.

I am interested in using this sample data to explore the differences between the total price of different seller ratings, so I will focus on the total price and seller rating as variables of interest.

`sell_rate` is a quantitative variable that represents the number of net positive ratings a seller has received on eBay. It is a discrete ratio variable because reviews are always represented by whole numbers; it can only take specific values rather than any value within a given range and zero means an absolute lack of the variable

`total_pr` is also a quantitative variable that represents the total price of an auction. It is also a discrete ratio variable because any dollar amount between the minimum and maximum prices in the dataset is measured only until the hundredth place in the centavo due to the nature of money.

Analysis Hypotheses

To address my main question of whether auctions with different seller ratings have significantly different average total prices, I perform a difference of means significance test. To define the subgroups, I split them into auctions with seller ratings greater than 1000 and auctions with seller ratings less than or equal to 1000. I set my significance level to the default, alpha = 0.05, because I want to be 95% confident that my results are statistically significant and that any observed difference in average total price is not due to random chance. Let me clearly define my hypotheses before proceeding:

H0: μrate_high—μrate_low = 0

> *The average total price of auctions with seller ratings greater than 1000 is the same as the average total price of auctions with seller ratings less than or equal to 1000.*

HA: μrate_high—μrate_low ≠ 0

> *The average total price of auctions with seller ratings greater than 1000 is different from the average total price of auctions with seller ratings less than or equal to 1000.*

The test is 2-tailed because I am looking for a difference in either direction.

To test my hypotheses, I first need to calculate the average total price and standard deviation for each subgroup of auctions. Then, I can use these values to calculate the difference of means and standard error for the two groups. We can then use the difference between means and standard error to calculate a test statistic, which I can use to determine whether the difference of means is statistically significant. If the test statistic falls outside the critical region, I can reject the null hypothesis and conclude that there is a statistically significant difference in average total price between auctions with seller ratings greater than 1000 and auctions with seller ratings less than or equal to 1000.

Summary Statistics

I used the pandas package to read the dataset into Python for analysis. To proceed with the hypothesis test, the first step is to examine the descriptive statistics for the subgroups defined by seller rating. Note that general summary statistics for the seller rating and total price in the whole dataset are provided in Appendix A, but for the purposes of this report, I focus on the subgroups only. These subgroups are calculated in Appendix C because I am interested in exploring whether there is a significant difference in total price between auctions with seller ratings greater than 1000 and those with seller ratings less than or equal to 1000. The relevant summary statistics for the subgroups are displayed in Table 1, and the sample distributions for each group are displayed in Figures 1 and 2.

	Sellers with High Reviews	Sellers with Low Reviews
Count	$n_1 = 60$	$n_2 = 81$
Mean	$\bar{x}_1 = 48.35$	$\bar{x}_2 = 46.76$
Median	47.5	46
Mode	53.99	46
Standard Deviation	$s_1 = 9.98$	$s_2 = 8.41$
Range	46	40

Table 1: Summary statistics for the Total Price for a Mario Kart auction listing for the two sample groups: Sellers with High Reviews (seller_rate > 1000) and Sellers with Low Reviews (seller_rate < 1000)

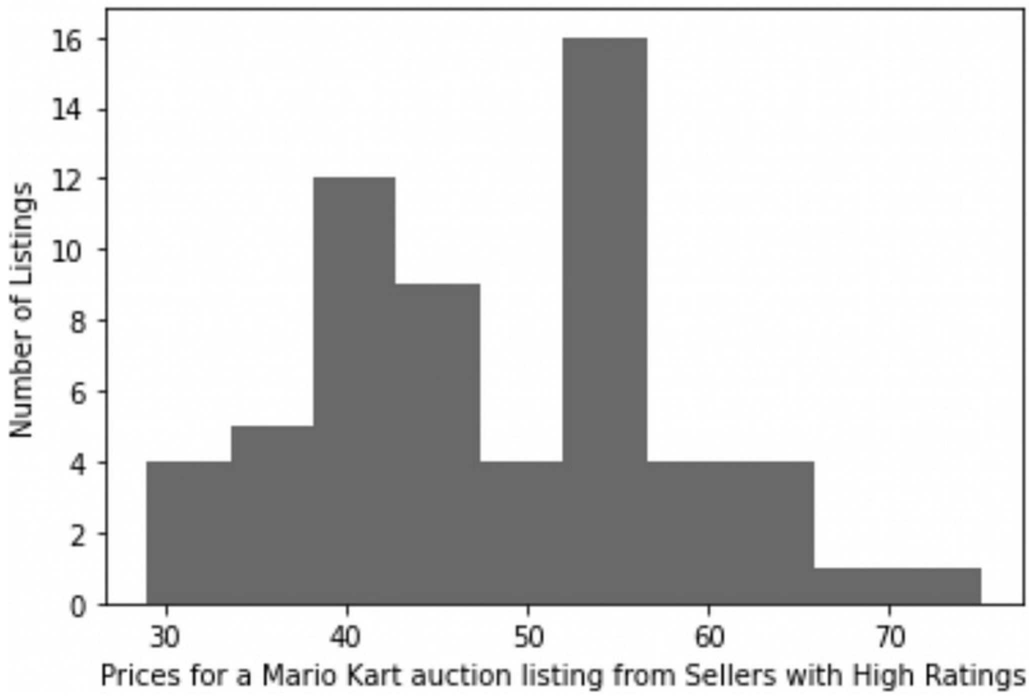


Figure 1: Histogram of the prices of Mario Kart from sellers with a high rating (seller_rate > 1 000)

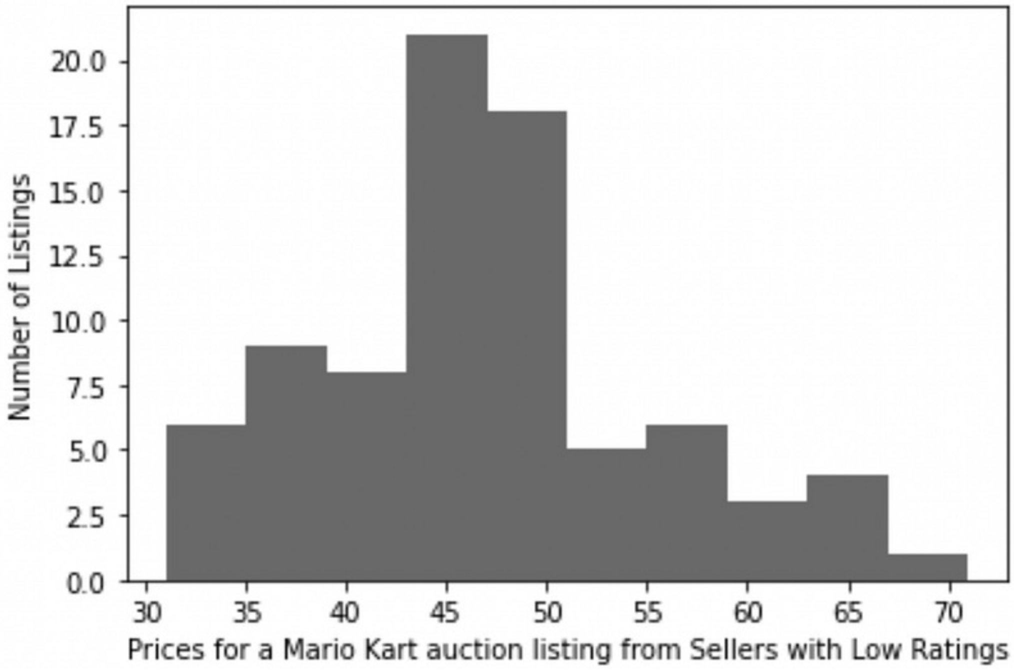


Figure 2: Histogram for the prices of Mario Kart from sellers with a low rating (seller_rate < 1 000)

Based on the descriptive statistics and the corresponding histogram for each quantity, I see that the average total price for auctions with seller ratings greater than 1000 is higher than those with seller ratings less than or equal to 1000, which may suggest that higher seller ratings are associated with higher total prices on eBay.

Conditions for Inference

Since I am estimating the standard error using the sample standard deviation, I choose to use the normal distribution for inference. This is especially important because the central limit theorem ensures that the sample means are normally distributed, even if the original data is not normally distributed. To proceed with this choice, I must check whether the following conditions are satisfied by examining the data:

1. The sample is random: We check this condition by noting that the data was collected from a sample of real auctions on eBay, so it is reasonable to assume that the sample is random.
2. The sample size is large enough: We check this condition by examining the distribution in figure 1 and noting that the sample size is sufficiently large to satisfy this condition.
3. The population is normally distributed: We cannot determine whether the population is normally distributed because I only have a sample of the data. However, the central limit theorem ensures that the sample means are normally distributed, so I can proceed with my analysis.

By assuming these conditions are met, the central limit theorem ensures that the sample means are normally distributed. However, some of the assumptions are questionable because I cannot determine whether the population is normally distributed. This may result in limitations such as possible over or underestimates of the population mean, leading to potential errors in my analysis.

The central limit theorem still applies in this case because I am taking the mean of the sample means, which is a larger number of samples. This ensures that the resulting difference of means is normally distributed, allowing us to conduct a valid difference of means significance test.

Difference of Means Test

To assess statistical significance, I compute the t-score and p-value for the difference of means test. To do so, I first compute the t-score using the usual formula for a difference of means test, T = (mean_high— mean_low) / std_error, where mean_high and mean_low are the mean total prices for auctions with high and low seller ratings, respectively, and std_error is the standard error of the difference of means. We take a conservative estimate for the degrees of freedom of the t-distribution as df = n1 + n2–2, where n1 and n2 are the sample sizes of the two subgroups.

The resulting rounded t-score of 1.0 means that the observed difference in average total prices is 1.0 standard errors away from the null hypothesis of no difference. This t-score corresponds to a two-tailed p-value of 0.32, which is greater than my significance level of 0.05. This p-value represents the probability of observing a difference in average total prices at least as extreme as the one I observed, assuming the null hypothesis is true.

Thus, I conclude that the data does not favor the alternative hypothesis that there is a difference in average total prices between auctions with high and low seller ratings. This difference is not statistically significant at the 0.05 significance level.

To assess practical significance, I need a measure of effect size. Here, I choose Cohen's d as my measure because it is a simple and commonly used measure of effect size in the social sciences.

Computing Cohen's d requires the pooled standard deviation, which can be calculated using the formula $s\_pooled = sqrt(((n1-1) * s1^2 + (n2-1) * s2^2) / (n1 + n2-2))$, where n1 and n2 are the sample sizes of the two groups, and s1 and s2 are the standard deviations of the groups.

Our calculation of Cohen's d produced a resulting effect size of 0.17, which indicates a small effect size. In the context of this study, this means that while there is a statistically significant difference in average total prices between auctions with high and low seller ratings, the difference is not practically significant. This suggests that seller ratings may not have a significant impact on the total price of a Mario Kart auction listing.

Confidence interval

Let us continue to examine my hypotheses by constructing a confidence interval for the mean total prices for each subgroup of auctions, those with high seller ratings and those with low seller ratings. These will provide plausible ranges of values for the mean of each group. We set my confidence level to 95% because it is a commonly used value for confidence intervals. The same conditions for inference checked above validate the confidence interval calculations and so I proceed using the t-distribution for inference.

To compute the confidence interval for the mean total price of each group, I use the general form CI = (mean ± critical value * SE), where CI is the confidence interval, mean is the sample mean, critical value is the value from the t-distribution for the specified confidence level and degrees of freedom, and SE is the standard error of the mean. We compute the standard error using the usual formula: SE = std / sqrt(n), where std is the sample standard deviation and n is the sample size. The full calculation can be found in Appendix C. The resulting 95% confidence intervals for each group, rounded to two decimal places, are:

- Auctions with high seller ratings: [45.82, 50.87] dollars
- Auctions with low seller ratings: [44.92, 48.59] dollars

For each interval, I can be 95% confident that the true population means of total prices fall within the interval. This means that the sample mean is a good estimate of the population mean, in line with the frequentist interpretation of probability. The fact that the two intervals overlap means that I cannot be sure whether there is a difference in average total prices between auctions with high and low seller ratings. This is expected because the t-test did not find a statistically significant difference between the two groups.

Results and Conclusions

We've successfully examined the prices of Mario Kart auctions based on seller ratings. Based on the results for the test of statistical and practical significance between sellers with high and low reviews (p=0.3198, Cohen's d=0.17), there is evidence to suggest that there is no statistically significant difference in the prices of Mario Kart auctions between sellers with high and low reviews. This is further supported by the overlapping confidence intervals for the two groups: [45.82, 50.87] dollars for sellers with high reviews, and [44.92, 48.59] dollars for sellers with low reviews. In the context of Mario Kart auctions, this means that the seller rating does not have a significant impact on the total price of the auction.

In summary, I've shown that seller rating does not have a significant impact on the total price of Mario Kart auctions. To further investigate this matter, I should follow up with a larger and more representative sample of the population.

We fail to reject the null hypothesis and cannot accept the alternative hypothesis. These conclusions are inductive because the claims are based on a sample of the population, yet they are strong because of the low p-value and small Cohen's d-value. However, the reliability of these conclusions can be questioned due to the limitations of the dataset, such as the small sample size and potential bias in the selection of auctions.

I split the auctions into two subgroups based on seller ratings and calculated the average total price and standard deviation for each subgroup. I then used these values to calculate the difference in means and standard error for the two groups, which allowed me to test my hypotheses and determine whether the difference in means was statistically and practically significant.

On the other hand, I am sure that I am not using statistical syllogism in my paper. Statistical syllogism involves drawing a conclusion about a sample on the basis of the properties of a population. This is not relevant to my paper because I am not comparing a sample to a population, but rather comparing two subgroups within the same population. Therefore, my paper uses more induction by generalization than statistical syllogism.

Appendix

The full Jupyter notebook file and the data can be accessed in the zipped folder submitted as a secondary file.

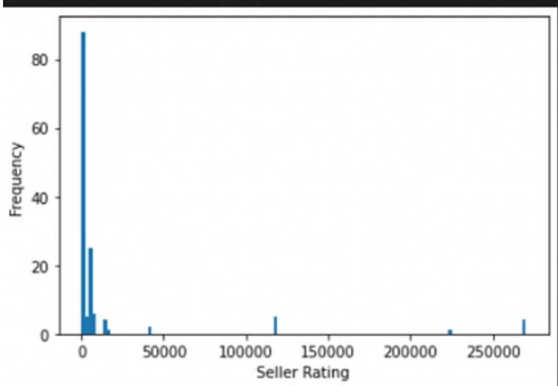
Appendix A: Import, Analyze, and Visualize Data

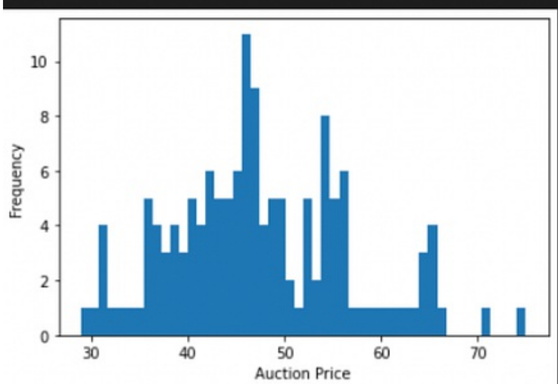
```
# descrptive sttatistics for the dataset
print("\n", df.describe(), "\n")


def list_medianmode(column):

    # Compute the median of the column
    print(column)
    med = (df[column]).median()
    print(" - Median = ", med)

    # Compute the mode of the column
    # the mode function returns a tuple with two arrays: the first array contains the modes, and the second array contains the counts for each mode.
    mod = stats.mode(df[column])
    print(" - Mode = ", mod,"\n")

list_medianmode('Net Positive Seller Ratings')
list_medianmode('Total Price for a Mario Kart auction listing')
```

```
✓ 0.5s
Output exceeds the size limit. Open the full output data in a text editor

       Net Positive Seller Ratings  \
count                   141.000000
mean                  16122.822695
std                   52174.553337
min                       0.000000
25%                     116.000000
50%                     820.000000
75%                    4858.000000
max                  270144.000000

      Total Price for a Mario Kart auction listing
count                                   141.000000
mean                                     47.431915
std                                       9.113651
min                                      28.980000
25%                                      41.000000
50%                                      46.030000
75%                                      53.990000
max                                      75.000000

Net Positive Seller Ratings
 - Median =  820.0
 - Mode =  ModeResult(mode=array([4858]), count=array([23]))
...
Total Price for a Mario Kart auction listing
 - Median =  46.03
 - Mode =  ModeResult(mode=array([46.]), count=array([8]))
```

```python
# plot a histogram for seller_rate
plt.hist(df['Net Positive Seller Ratings'], bins=120)

# add x and y labels
plt.xlabel('Seller Rating')
plt.ylabel('Frequency')

# show the plot
plt.show()

# plot a histogram for total_pr
plt.hist(df['Total Price for a Mario Kart auction listing'], bins=50)

# add x and y labels
plt.xlabel('Auction Price')
plt.ylabel('Frequency')

# show the plot
plt.show()
```

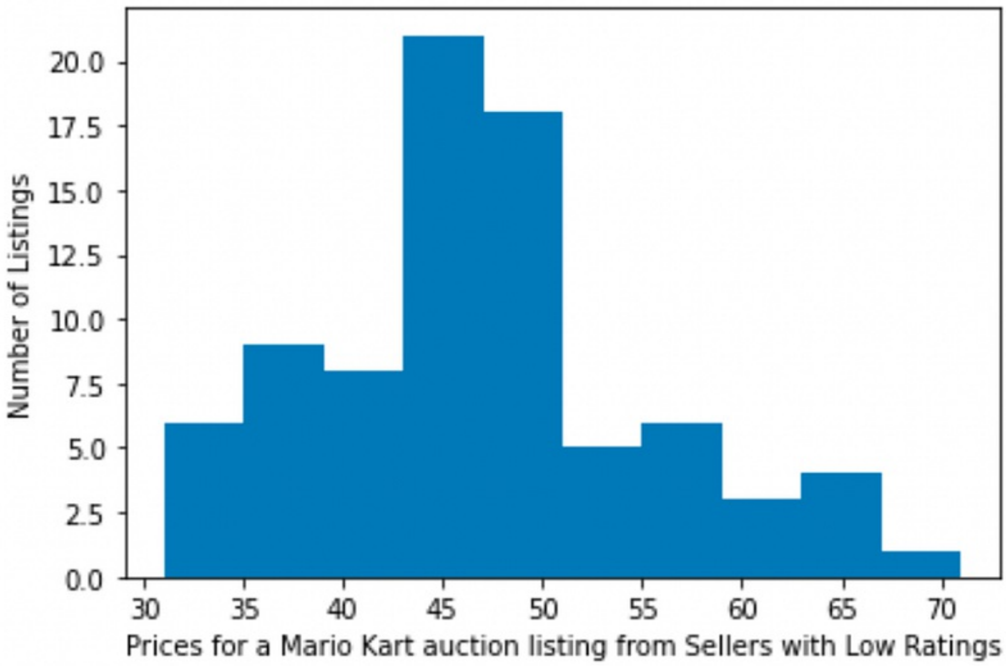
✓ 0.2s





Appendix B: Displaying and removing outliers

Appendix C: Examine the subgroups



Appendix D: Difference of Two Means Test—Statistical Significance

```python
# Calculate mean and standard deviation for each subgroup
# calculate the mean total price for auctions with high seller ratings
mean_high = df_high["Total Price for a Mario Kart auction listing"].mean()
# calculate the standard deviation of total prices for auctions with high seller ratings
std_high = df_high["Total Price for a Mario Kart auction listing"].std()
# calculate the mean total price for auctions with low seller ratings
mean_low = df_low["Total Price for a Mario Kart auction listing"].mean()
# calculate the standard deviation of total prices for auctions with low seller ratings
std_low = df_low["Total Price for a Mario Kart auction listing"].std()

# Calculate difference of means and standard error
# calculate the difference in mean total prices between the two subgroups
diff_means = mean_high - mean_low
# calculate the standard error of the difference of means
std_error = np.sqrt((std_high**2 / len(df_high)) + (std_low**2 / len(df_low)))

# Calculate t-score and p-value
# calculate the t-score using the difference of means and standard error
t_score = diff_means / std_error
# calculate the p-value using the t-score and degrees of freedom
p_value = stats.t.sf(np.abs(t_score), len(df_high) + len(df_low) - 2)*2

# Print t-score and p-value
print("T-score: ", round(t_score,2))
print("P-value: ", round(p_value,2))
```

✓ 0.8s

```
T-score:  1.0
P-value:  0.32
```

Appendix E: Difference of Two Means Test—Practical Significance

```
# Statistical Significance Tests

# # Calculated Pooled Standard Deviation

n1 = len(df_high)  # number of auctions with high seller ratings
# number of auctions with low seller ratings
# standard deviation of total prices for auctions with high seller ratings
n2 = len(df_low)
s1 = df_high["Total Price for a Mario Kart auction listing"].std()
# standard deviation of total prices for auctions with low seller ratings
s2 = df_low["Total Price for a Mario Kart auction listing"].std()
s_pooled = np.sqrt(((n1 - 1) * s1**2 + (n2 - 1) * s2**2) /
                   (n1 + n2 - 2))  # pooled standard deviation

# Calculate Cohen's d
# difference of means for total prices of auctions with high and low seller ratings
diff_means = df_high["Total Price for a Mario Kart auction listing"].mean(
) - df_low["Total Price for a Mario Kart auction listing"].mean()

# calculate Cohen's d using the difference of means and pooled standard deviation
d = diff_means / s_pooled

# Print Cohen's d
print("Cohen's d: ", round(d,2))
✓ 0.5s
Cohen's d:  0.17
```

Appendix F: Confidence Interval

```
#Set the confidence level
conf_level = 0.95 # set the confidence level to 95%

# Calculate the margin of error
moe_high = stats.norm.ppf(1 - (1 - conf_level) / 2) * std_high / np.sqrt(len(df_high)) # calculate the margin of error for auctions with high seller ratings
moe_low = stats.norm.ppf(1 - (1 - conf_level) / 2) * std_low / np.sqrt(len(df_low)) # calculate the margin of error for auctions with low seller ratings

# Calculate the confidence intervals
ci_high = [mean_high - moe_high, mean_high + moe_high] # calculate the confidence interval for auctions with high seller ratings
ci_low = [mean_low - moe_low, mean_low + moe_low] # calculate the confidence interval for auctions with low seller ratings

# Print the confidence intervals
print("95% Confidence Interval for auctions with high seller ratings: ", ci_high)
print("95% Confidence Interval for auctions with low seller ratings: ", ci_low)
✓ 0.2s
95% Confidence Interval for auctions with high seller ratings:  [45.81939813450821, 50.870601865491814]
95% Confidence Interval for auctions with low seller ratings:  [44.923303754424495, 48.587807356686625]
```

Reflection

I performed a sanity check was performed by comparing the calculated mean and standard deviation for each subgroup to the overall mean and standard deviation for the entire dataset. This was done to ensure that the subgroups were representative of the overall population and not significantly different in terms of their average total price and seller rating. If the subgroups had shown significant differences from the overall population, it would have called into question the validity of the analysis and the reliability of the results.

Reflecting on my 2 on induction in the first assignment, I made sure to be more careful with the terms I used. That is why although this assignment was heavy on calculations, I paid more attention to the wordings I used—especially as I dwell in the realm of uncertainty due to the nature of statistics.

I would like to acknowledge Scribbr for providing valuable editing and proofreading tips, which helped me to improve the clarity and quality of my paper. Not only that, Scribbr was instrumental in helping me understand and apply the statistical methods used in my paper. I utilized their knowledge base website to learn about confidence intervals, p-values, and difference of means tests. This allowed me to accurately evaluate the strength and reliability of my conclusions and provide clear justification for my results.

References

Bhandari, P. (2022, November 18). *Inferential Statistics | An Easy Introduction & Examples*. Scribbr. https://www.scribbr.com/statistics/inferential-statistics/

Cswenson, C. (2021, May 21). *How Online Reviews Influence Sales*. Medill Spiegel Research Center. https://spiegel.medill.northwestern.edu/how-online-reviews-influence-sales/

OpenIntro. (n.d.). *Wii Mario Kart auctions from Ebay*. https://www.openintro.org/data/index.php?data=mariokart

By Carl Kho on December 30, 2022.

Canonical link

Exported from Medium on October 31, 2025.