

Price difference between the initial price and auction price of Nintendo Wii Mario Kart

Tanya Chen, Allison Bleicher

Introduction

Have you ever played Mario Kart on a Nintendo Wii? Many children in the US played it growing up. While you can still buy the game today, older versions are only available on second sale sites, like Ebay. Ebay is interesting because you can sell things through an online auction. You set the start price, but others can bid higher prices to get the item. In this data analysis, we look into the correlation between the start price, shipping price, and total price of Mario Kart games sold on Ebay. The data set used took all the auctions of Mario Kart games that happened during one week in October 2009 [mario]. Through this data analysis, we are able to see the correlation of the prices for Mario Kart games, but we can also apply these findings to other things for auction on Ebay to help people make decisions about what they are going to pay.

Exploratory Analysis

Dataset

The data was collected from Ebay and includes all auctions for a full week in early October of 2009. Auctions were only included if they fulfilled certain requirements. The auctions were in a search for “wii mario kart” on ebay.com; items were under the Video Games -> Games -> Nintendo Wii section of Ebay; the listing was an auction and not only a “Buy It Now” listing; the item listed was the actual Mario Kart game; the item was being sold from the United States; the item had at least one bidder; there were no other items included in the auction listing (except for racing wheels); and the auction did not end with a “Buy It Now” option.

There is code that creates a new table from the data that only includes the variables that we will be looking at and analyzing. The original data set includes 12 variables and 143 observations, but we only look at 3 specific variables. The variables studied are the start_pr (starting price of the auction in USD), ship_pr (shipping price of the item in USD), and total_pr (total price of the item in USD).

Distribution of Variables

Figure 1 is a boxplot that shows the distribution of the start prices. The mean start price is \$8.78.

Figure 2 is a boxplot showing the distribution of the shipping prices. The mean shipping price is \$3.14.

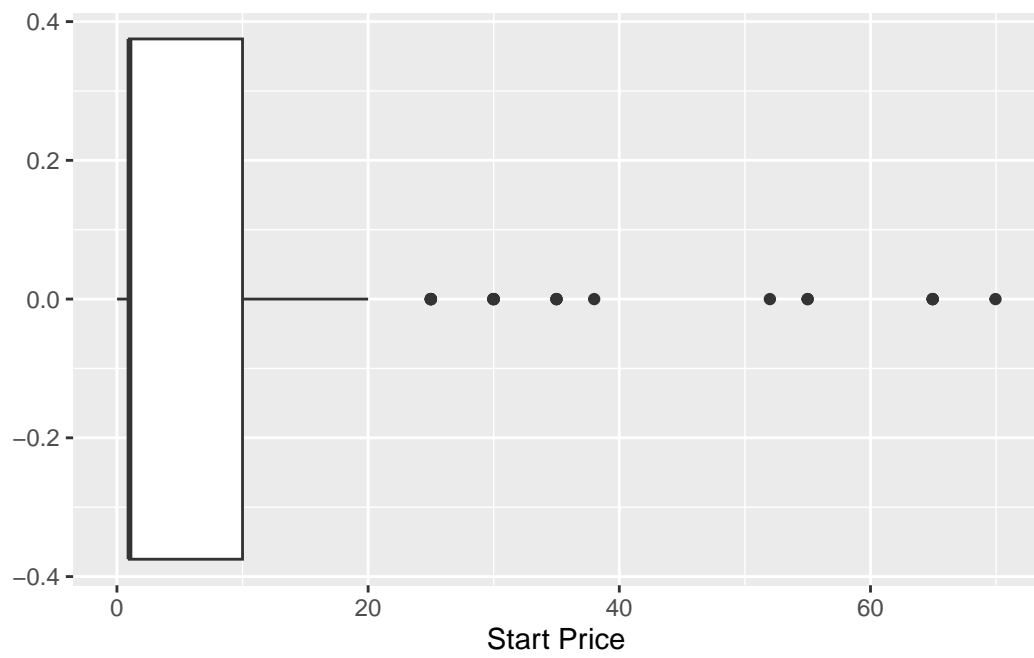


Figure 1: Distribution of the Start Prices

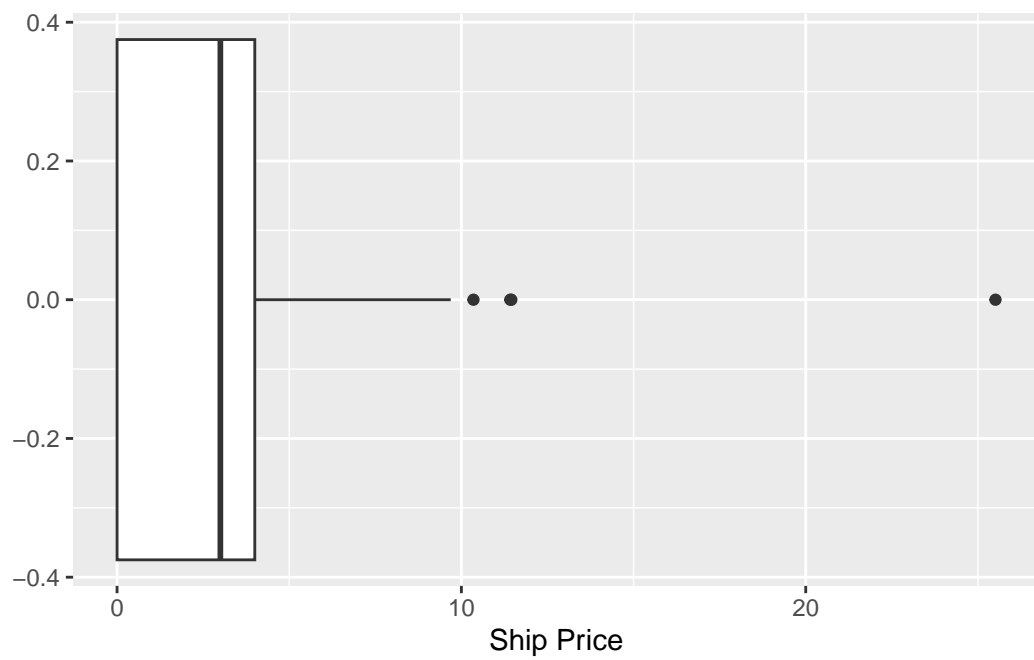


Figure 2: Distribution of the Ship Prices

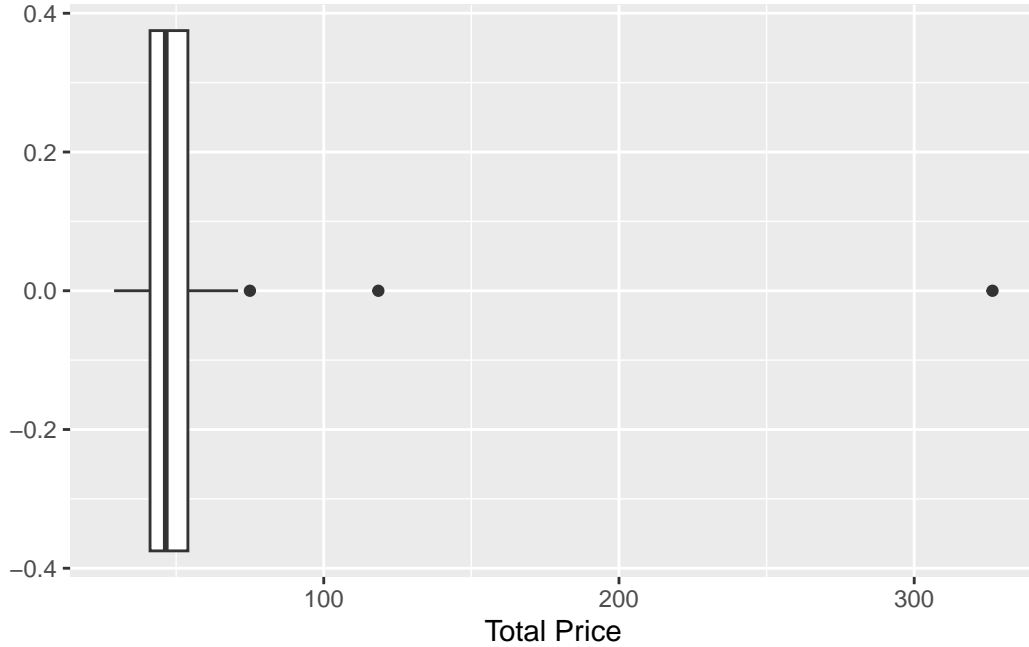


Figure 3: Distribution of the Total Prices

Figure 3 is a boxplot that shows the distribution of the total prices. The mean total price is \$49.88.

Relationships Between Variables

Figure 4 is a scatter plot of the relationship between the start price and the total price. The line of best fit shows a slightly positive relationship between the two variables.

Figure 5 is a scatter plot of the relationship between the shipping price and total price. The line of best fit shows a moderate positive relationship between the two variables.

Result

Hypotheses

Before starting hypothesis testing, we need to bring out null hypothesis and alternative hypothesis of our research question. Our null hypothesis is that there is no difference between the mean initial total price and the mean dealing price for game Mario Kart on EBay. Our alternative hypothesis is that there is a difference between the mean initial total price and the mean dealing price of Mario Kart on EBay. In symbols:

$$H_0 : \mu_{diff} = 0$$

$$H_A : \mu_{diff} \neq 0$$

Our goal is to use hypothesis testing to reject our null hypothesis. To investigate whether our null hypothesis is false, we need to conduct a **Paired T-Test**.

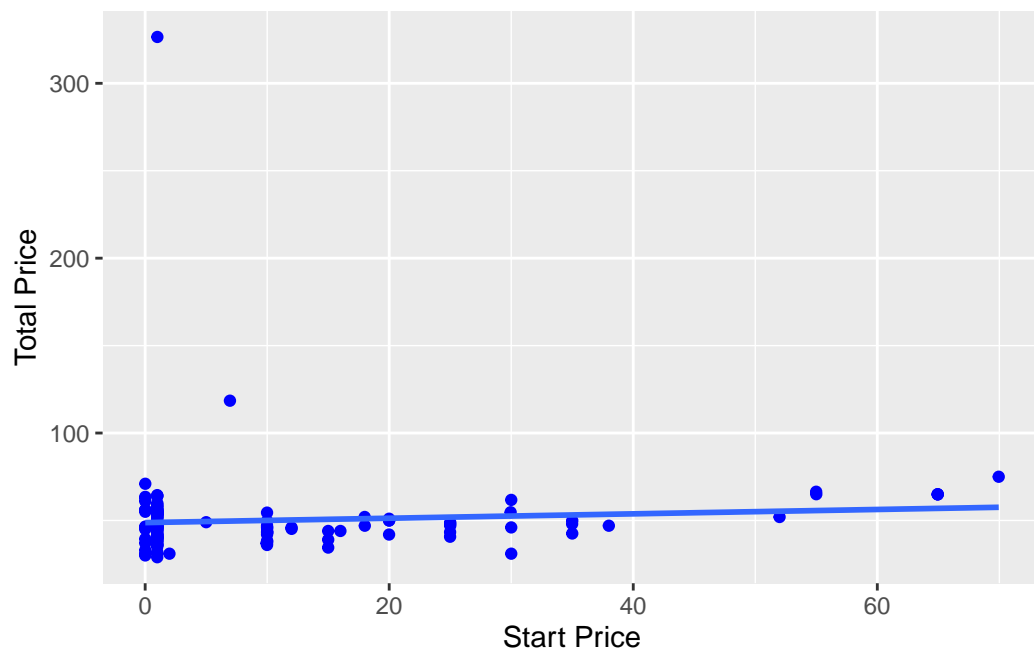


Figure 4: Relationship Between Start Price and Total Price

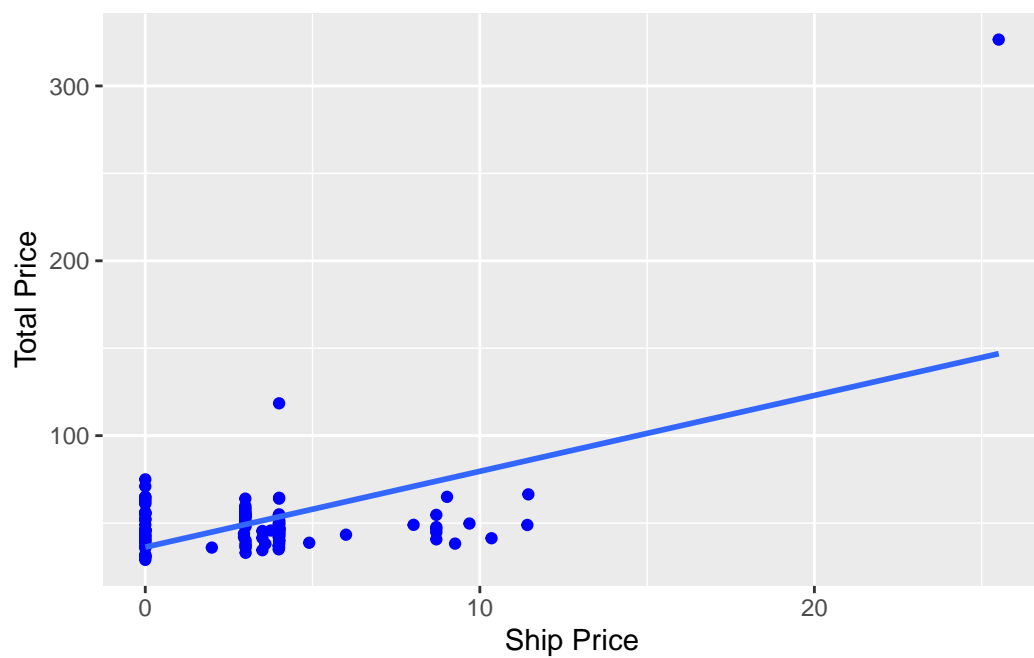


Figure 5: Relationship Between Ship Price and Total Price

Checking Conditions

Before conducting the hypothesis testing, we need to check two conditions: **independence** and **normality**.

Independence checks whether the data within the groups are independent. Since the `mariokart` dataset is collected, it is a simple random sample, and we can assume that each pair of the initial total price and the dealing price are independent.

Normality checks whether the difference of the paired data is roughly symmetric without large outliers.

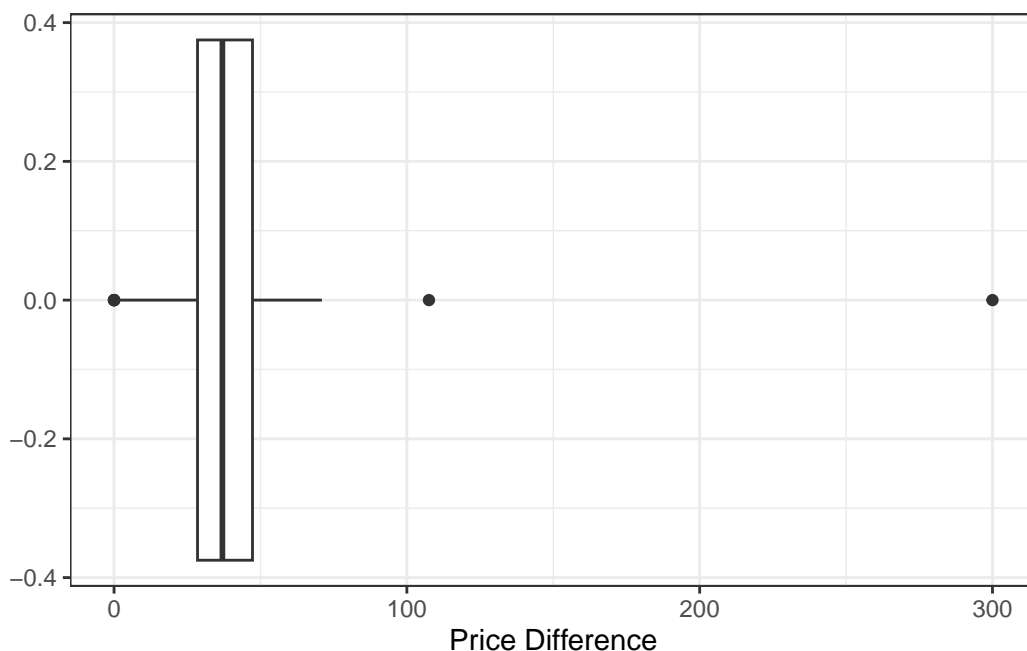


Figure 6: Distribution of price difference between initial and dealing price of Mario Kart on EBay

Figure 6 is a boxplot representing the distribution of the price difference of game Mario Kart on EBay in 2009. Based on the boxplot above, the distribution of the difference of initial total price and dealing price are roughly normal with three outliers. There is no outlier that significantly differs from the distribution and therefore satisfies the condition.

Calculation

Using t test for paired means, we will compute T-Statistics, degrees of freedom, p-value, and confidence interval with a significance level $\alpha = 0.05$. The significance level means that

T statistics stands for the ratio of how the sample mean difference varies from zero as compared to how the observations vary. The T test statistics obtained from formula

$$T = \frac{\bar{x}_{paired}}{\sqrt{\frac{S^2_{paired}}{n}}}$$

is -16.354. The large absolute value of t test statistics means that if our null hypothesis that there is no difference between the initial total price and the dealing price for Wii Mario Kart on EBay is true, the result is relatively largely deviated from what we expect based on our null hypothesis.

The degrees of freedom is used to describe the shape of t-distribution. Its formula is $df = n - 1$. In this case, the degrees of freedom is 142. The large degrees of freedom means that the t-distribution is closer to normal distribution.

After calculating the t statistics and degrees of freedom, we can calculate **p-value**. P-value is the probability that views the alternative hypothesis for the data least agreeable if the null hypothesis is true. Since our alternative hypothesis is $H_A : \mu_{diff} \neq 0$, our p-value is two-tailed. Based on R code `t.test(initial_total_pr,dealing_pr, paired = TRUE, alternative = "two.sided")`, the p-value is to $2.2 * 10^{-16}$. Since our p-value is very small and approximately equals to zero, our p-value is smaller than the significance level $\alpha = 0.05$. Therefore, we reject the null hypothesis that there is no difference between the mean initial total price and the mean dealing price of game Mario Kart on EBay on 2009.

To find a range of the true difference between initial price and dealing price, we compute **confidence interval**. Confidence interval is a range of plausible values where we can find true population value. In this case, we are using confidence interval to calculate a range of plausible values of true difference of paired means. A Based on R Code `t.test(initial_total_pr,dealing_pr, paired = TRUE, alternative = "two.sided")`, The 95% confidence interval is $(-42.49814, -33.42102)$. It means that we are 95% confident that true mean difference between initial price and dealing price is in the range $(-42.49814, -33.42102)$.

Discussion

This study investigate whether there is a difference between the initial price and the dealing price of game Mario Kart on EBay. Based on the result we gathered, there is a difference between the initial total price and the dealing price of game Mario Kart on EBay in 2009. The true mean difference between the initial total price and dealing price of Nintendo Wii Mario Kart lies in the range $(-42.49814, -33.42102)$. Based on our result, the dealing price of game Mario Kart is more expensive than its initial total price.

There are some limitations in this study. 1. We cannot use the data from 2009 to match what is the current second-hand price for Wii Mario kart. 2. This data is not a simple random sample. The data set used took all the auctions of Mario Kart games that happened during one week in October 2009. This sample is not representative. Different time (discount seasons might make a difference for the dealing price/auction price). If we want to investigate more 1. Gather new data of current second-hand Nintendo Wii Mario Kart price with larger sample size.