

Final Project

Tanya Chen, Allison Bleicher

Introduction

Exploratory Analysis

```
Rows: 143
Columns: 2
$ initial_total_pr <dbl> 4.99, 4.98, 4.49, 0.99, 0.01, 4.99, 0.01, 3.99, 4.99,~
$ dealing_pr      <dbl> 51.55, 37.04, 45.50, 44.00, 71.00, 45.00, 37.02, 53.9~
```

Result

To investigate whether the average initial total price (the sum of starting price and shipping price) equals to the average dealing price, we need to conduct a two sample mean T test.

Hypotheses

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

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

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

To investigate whether our null hypothesis is true, we need to conduct a **T Test for paired means**.

Checking Conditions

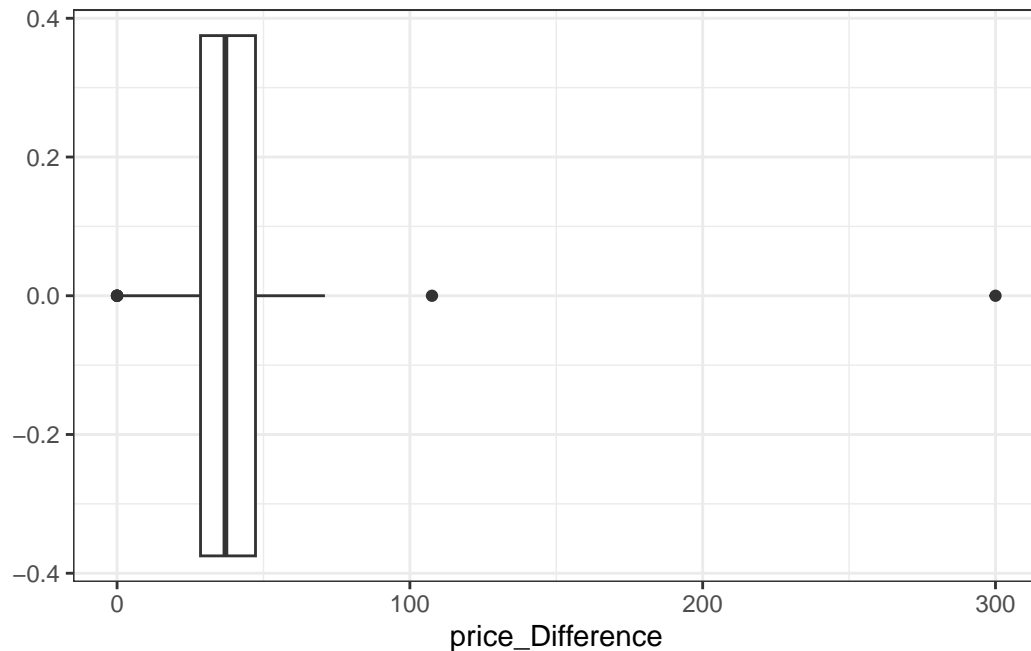
Before conducting the hypothesis testing, we need to check the conditions: **independence** and **normality**.

Independence

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

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



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 an outlier that significantly differs from the distribution.

Calculation

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

R Code Overview

In RStudio, we can use `t.test(initial_total_pr,dealing_pr, paired = TRUE, alternative = "two.sided")` to calculate t statistics, degrees of freedom, p-value, and confidence interval of the difference between the initial and dealing price of Mario Kart game on EBay:

Paired t-test

```

data: initial_total_pr and dealing_pr
t = -16.534, df = 142, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -42.49814 -33.42102
sample estimates:
mean difference
 -37.95958

```

Aside from R code, we also can compute T-statistics, degrees of freedom, p-value, and confidence interval based on their mathematical formulas.

T-Statistics

T statistics means the ratio of how the sample mean difference varies from zero as compared to how the observations vary. We compute T statistics using the formula below:

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

```

# A tibble: 1 x 4
      n x_bar_paired s_paired t_statistics
  <dbl>   <dbl>   <dbl>   <dbl>
1   143    -38.0    27.5    -16.5

```

Since $n = 143$, $\bar{x}_{paired} = -37.9596$, and $S_{paired} = 27.4549$, the T-statistics equals to:

$$T = \frac{\bar{x}_{paired}}{\sqrt{\frac{S_{paired}^2}{n}}} = \frac{-37.9596}{\sqrt{\frac{27.4549^2}{143}}} = -16.534$$

Degrees of Freedom The degrees of freedom, meaning that, equals to $df = n - 1 = 143 - 1 = 142$

P-Value

After calculating the t statistics and degrees of freedom, we can calculate p-value using R code `2*pt(t_statistics,df,lower.tail = TRUE)`. Since our alternative hypothesis is $H_A : \mu_{diff} \neq 0$, our p-value is two-tailed.

Based on R code, the p-value equals to 6.584×10^{-35} . 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, there is a strong evidence to reject the null hypothesis that there is no difference between the mean initial total price and the mean dealing price of Mario Kart in EBay.

Confidence Interval

Confidence interval is

Discussion