MATH1324 Assignment 2

Code

Supermarket Price Wars

Group/Individual Details

- Abdul Raheman Nasir Gazi(s3705448)

Executive Statement

Coles and Woolworths are arguabely the two biggest supermarkets in Australia. Because they're the most famous supermarkets it's natrual to have a competition betwween them. Both Coles and Woolworths are always in competition based on prices, quality and exclusive prducts. Both supermakets aim to provide products with superior quality in a reasonable price. To get an idea about the difference in price for similar products provided by Coles and Woolworths both I used Random sample technique to get prices for 37 products which are both available at Woolworths and Coles.

I noted 50 products down and numbered them from 1 to 50, using online random number generator I picked 37 products out of the lot and created a dataset consisting of those 37 observations. The dataset "pricewars.csv" contains four variables(Product, Category, Prices_Coles, Price_Woolworths). The data has been collected over just a single day(18th May, 2019)

Summary statistics, box-plot were worked upon to get insights about the difference in price before the actual t-test. Q-Q plot was used to see if the prices were normally distributed.

The t-test done below revealed that there is no significant difference between the product prices of Coles and Woolworths. The null hypothesis failed to get rejected.

Load Packages and Data

Hide

```
library(knitr)
library(magrittr)
library(dplyr)
library(readr)
pricewars <- read_csv("pricewars .csv")
```

```
Parsed with column specification:
cols(
  Product = [31mcol_character()[39m,
  Category = [31mcol_character()[39m,
  Price_Coles = [32mcol_double()[39m,
  Price_Woolworths = [32mcol_double()[39m
)
```

Hide

```
pricewars
```

Product <chr>	Category <chr>	Price_Coles <dbl>	Price_Woolworths <dbl>
Wholemeal Bread	Bakery	2.70	3.00
Toscano Pizza Base	Bakery	3.95	5.50
Helga's Low Carb Wraps	Bakery	5.00	4.00
Tip Top Raisin Roast Toast	Bakery	3.20	4.00

Product <chr>	Category <chr>	Price_Coles <dbl>	Price_Woolworths <dbl>
Top Top Sandwich Thins Original	Bakery	3.50	2.80
Tip Top Sandwich Thins Wholemeal	Bakery	3.50	2.80
Hot Dog Rolls	Bakery	4.00	4.00
English Muffins	Bakery	2.10	2.10
Top Top Original English Muffins	Bakery	5.00	3.00
Free Range Eggs	Dairy, Eggs	4.20	4.20

1-10 of 36 rows Previous **1** 2 3 4 Next

Summary Statistics

After investigating the data collected and finding the summary and bloxpot for the prices of products for both Coles and Woolworths we can see that Coles has a lower mean price than Woolworhts. Coles mean price is AU5.28 and Woolworths mean price is AU5.51. There is also a small difference between the median price of both the supermaket i.e Coles median price is 3.97 and Woolworths median price is 4. Although there is a difference in the summary of both the supermakets, the difference is not significant enough.

After checking Q-Q plot for both the supermarket prices we can see that in both the plots the points are fitting close to the distribution line but only until halway between 0 and 1. We can see distribution have far outliers that are away from the normal distribution line.

Hide

```
pricewars %>% summarise(Min = min(Price_Coles,na.rm = TRUE),
                        Median = median(Price_Coles, na.rm = TRUE),
                        Max = max(Price_Coles,na.rm = TRUE),
                        Mean = mean(Price_Coles, na.rm = TRUE),
                        SD = sd(Price_Coles, na.rm = TRUE),
                        n = n(),
                        Missing = sum(is.na(Price_Coles)))
```

Min <dbl>	Median <dbl>	Max <dbl>	Mean <dbl>	SD <dbl>	n <int>	Missing <int>
1.2	3.975	22	5.288889	4.162076	36	0

1 row

Hide

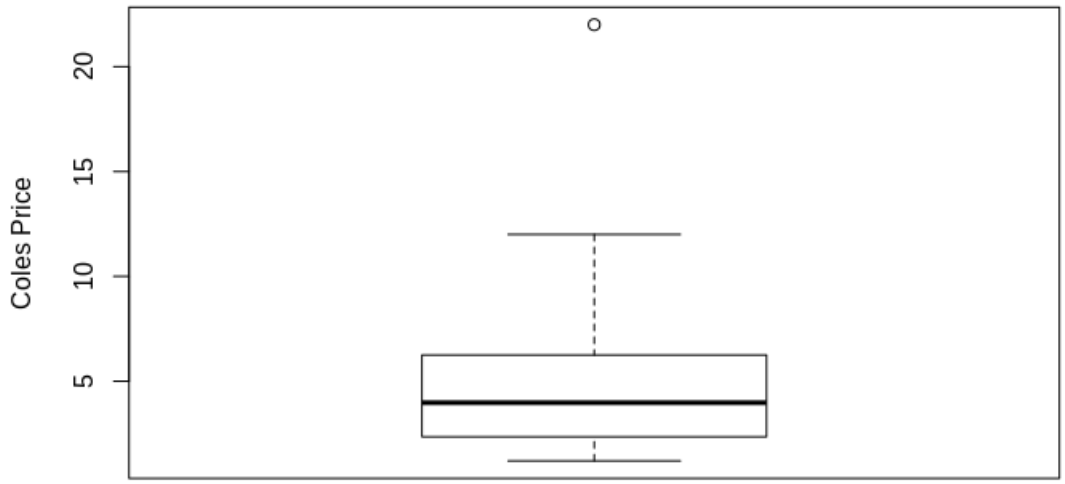
```
pricewars %>% summarise(Min = min(Price_Woolworths,na.rm = TRUE),
                  Median = median(Price_Woolworths, na.rm = TRUE),
                  Max = max(Price_Woolworths,na.rm = TRUE),
                  Mean = mean(Price_Woolworths, na.rm = TRUE),
                  SD = sd(Price_Woolworths, na.rm = TRUE),
                  n = n(),
                  Missing = sum(is.na(Price_Woolworths)))
```

Min <dbl>	Median <dbl>	Max <dbl>	Mean <dbl>	SD <dbl>	n <int>	Missing <int>
1.25	4	25	5.511111	4.648888	36	0

1 row

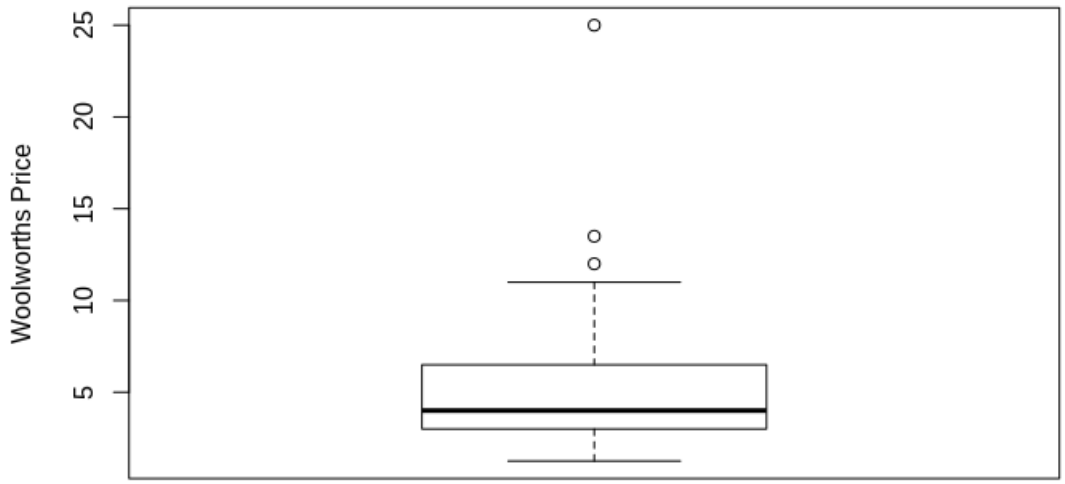
Hide

```
pricewars$Price_Coles %>% boxplot(ylab = "Coles Price")
```



Hide

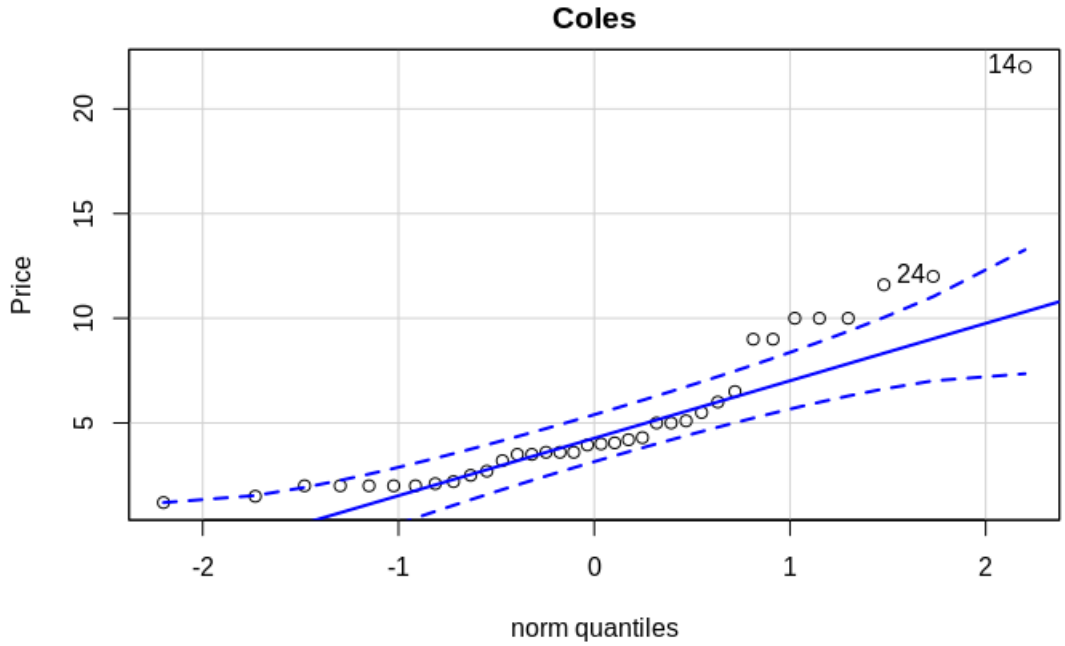
```
pricewars$Price_Woolworths %>% boxplot(ylab = "Woolworths Price")
```



Hide

```
pricewars$Price_Coles %>% qqPlot (dist="norm", main = "Coles", ylab = "Price")
```

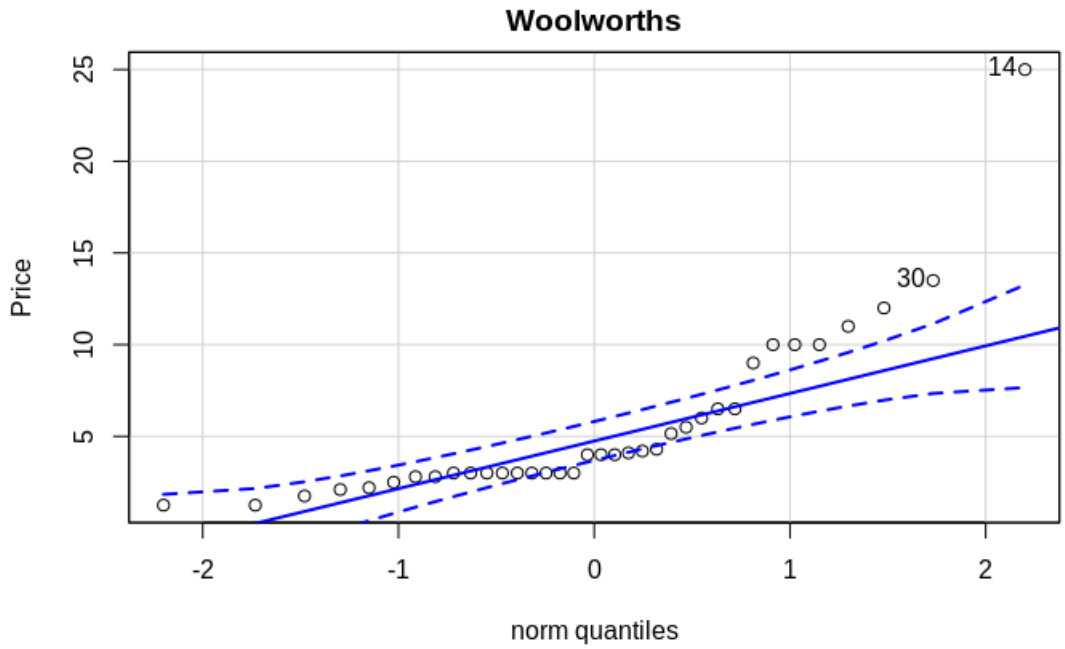
```
[1] 14 24
```



Hide

```
pricewars$Price_Woolworths %>% qqPlot (dist="norm", main = "Woolworths", ylab = "Price")
```

```
[1] 14 30
```



Hypothesis Test

Two-Sample T-Test was used to determine the difference between prices of products for the two supermarkets. Significance level of 0.05 has been used to reject or not reject the null hypothesis. H0 : μPrice_Coles - μPrice_Woolworths = 0, HA: μPrice_Coles - μPrice_Woolworths ≠ 0

Hide

```
t.test(pricewars$Price_Coles, pricewars$Price_Woolworths)
```

```
    Welch Two Sample t-test

data:  pricewars$Price_Coles and pricewars$Price_Woolworths
t = -0.21368, df = 69.161, p-value = 0.8314
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.296811  1.852367
sample estimates:
mean of x mean of y
 5.288889  5.511111
```

Interpretation

As per the t-test done above, the p-value of the test turned out to be 0.83 which is greater than the significance level of 5% hence we fail to reject H0. We can see a difference between mean of Price_Coles and mean of Price_Woolworths but it's not significant enough.

Discussion

The investigation suggests that there is a slight variance in price between both the supermarkets Coles and Woolworths. As per the caluclations there is a statistical evidence that there is no significant difference between both the supermarkets. We also need to consider the fact that the sample data was small and there were certain limitations which if worked upon may change the price difference between the two.

During the creation of the sample dataset it was ensured that the product size and brand is also taken into the factor. It was observed that when the product was of the brand Coles and Woolworths when in comparison, there was a big variance in price. in future just working on the products made by the supermarkets themselves might reveal a different scenario.

In conclusion, this report exhibits a slight price difference(not significant) between Coles and Woolworths.