# Supermarket Prices War

## Individual

Riya Minesh Amin/s3807007

## Executive Statement

The objective of the investigation is to compare the price (AUD) of different products in the two supermarkets of Australia i.e. Coles and Woolworths in order to figure out which one us cheaper.

## Introduction

Founded in Australia in 1914 and 1924 respectively, Coles Supermarkets Australia Pty Ltd (trading as Coles) and Woolworths Supermarkets (commonly referred to as "Woolworths" or "Woolies"). They make up more than 80% of the Australian industry together. Both Coles and Woolworths, while retaining high-quality products and a broad variety , market themselves in a highly cost-competitive manner. Throughout this investigation, by assessing two hypothesis, we will try to determine cost gaps between Coles and Woolworths.

## About the Sample

To acheive the goal, 100 items were choosen from both Coles and Woolsworths at random from http://www.grocerycop.com.au (http://www.grocerycop.com.au). To guarantee the cost difference (AUD) between the same brand and class products was contrasted and selected to be included in the dataset. In order to get a normal distribution, a dataset of 100 points was done. In order to achieve an unbiased dataset, the cost of products on sale and bids was ignored. According to the Central Limit Theorem(CLT), if the sample size is higher than 30 this implies that the information is about standard.

## Procedure

The dataset is dependent on the two supermarkets as it comprises of the same products. The H0 null hypothesis was regarded as a mean difference of 0 (HO = mu1-mu2) and an alternative HA hypothesis as a mean difference not equal to 0. Boxplots and qqplot have been used to determine price normality . To promote our hypothesis, we used a paired sample t-test to determine value p. If the value p is less than alpha, otherwise we dismiss the null hypothesis.

## Load Packages and Data

```
#Required Packages
library(dplyr)
library(readxl)
library(readr)
library(magrittr)
library(ggplot2)
library(gridExtra)
library("htmltools")
library(car)



# Importing data
SPW <- read_excel("D:/Supermarket Price Wars.xlsx")
View(SPW)

# Differences (d) column
SPW1 <- SPW %>% mutate(d = SPW$`Coles price`-SPW$`Woolworths price`)
```

# Summary Statistics

Comparing the sales of the products in both stores (Coles and Woolworths) was performed with the summary statistics. Boxplots have been created to comprehend price normality. It was noted that rates fell below the CI of 95 percent.

```
# Coles Summary Statistics
SPW %>% summarise(Mean = mean(SPW$`Coles price`, na.rm = TRUE),
                          Median = median(`Coles price`, na.rm = TRUE),
                          SD = sd(`Coles price`, na.rm = TRUE),
                          Q1 = quantile(`Coles price`, probs = .25, na.rm = TRUE),
                          Q3 = quantile(`Coles price`, probs = .75, na.rm = TRUE),
                          Min = min(`Coles price`, na.rm = TRUE),
                          Max = max(`Coles price`, na.rm = TRUE),n = n())
```

```
## # A tibble: 1 x 8
##    Mean Median    SD    Q1    Q3   Min   Max     n
##   <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <int>
## 1  4.54   4.08  2.30     3  5.84     1    14   100
```

```
#  Woolworths Summary Statistics
SPW %>% summarise(Mean = mean(`Woolworths price`, na.rm = TRUE),
                          Median = median(`Woolworths price`, na.rm = TRUE),
                          SD = sd(`Woolworths price`, na.rm = TRUE),
                          Q1 = quantile(`Woolworths price`,probs=.25,na.rm=TRUE),
                          Q3 = quantile(`Woolworths price`,probs=.75,na.rm =TRUE),
                          Min = min(`Woolworths price`, na.rm = TRUE),
                          Max = max(`Woolworths price`, na.rm = TRUE),n = n())
```

```
## # A tibble: 1 x 8
##    Mean Median    SD    Q1    Q3   Min   Max     n
##   <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <int>
## 1  4.56      4  2.23     3  5.85   1.3  11.8   100
```

```
SPW %>% summarise(Mean_Coles = mean(`Coles price`, na.rm = TRUE),
                  SD_Coles = sd(`Coles price`, na.rm = TRUE),
                  Mean_Woolworths = mean(`Woolworths price`, na.rm = TRUE),
                  SD_Woolworths = sd(`Woolworths price`, na.rm = TRUE),
                  Mean_Difference = Mean_Coles - Mean_Woolworths,
                  SD_Difference = sd(`Coles price`- `Woolworths price`, na.rm = TRUE),
                  n = n())
```

```
## # A tibble: 1 x 7
##   Mean_Coles SD_Coles Mean_Woolworths SD_Woolworths Mean_Difference
##        <dbl>    <dbl>           <dbl>         <dbl>           <dbl>
## 1       4.54     2.30            4.56          2.23         -0.0169
## # ... with 2 more variables: SD_Difference <dbl>, n <int>
```
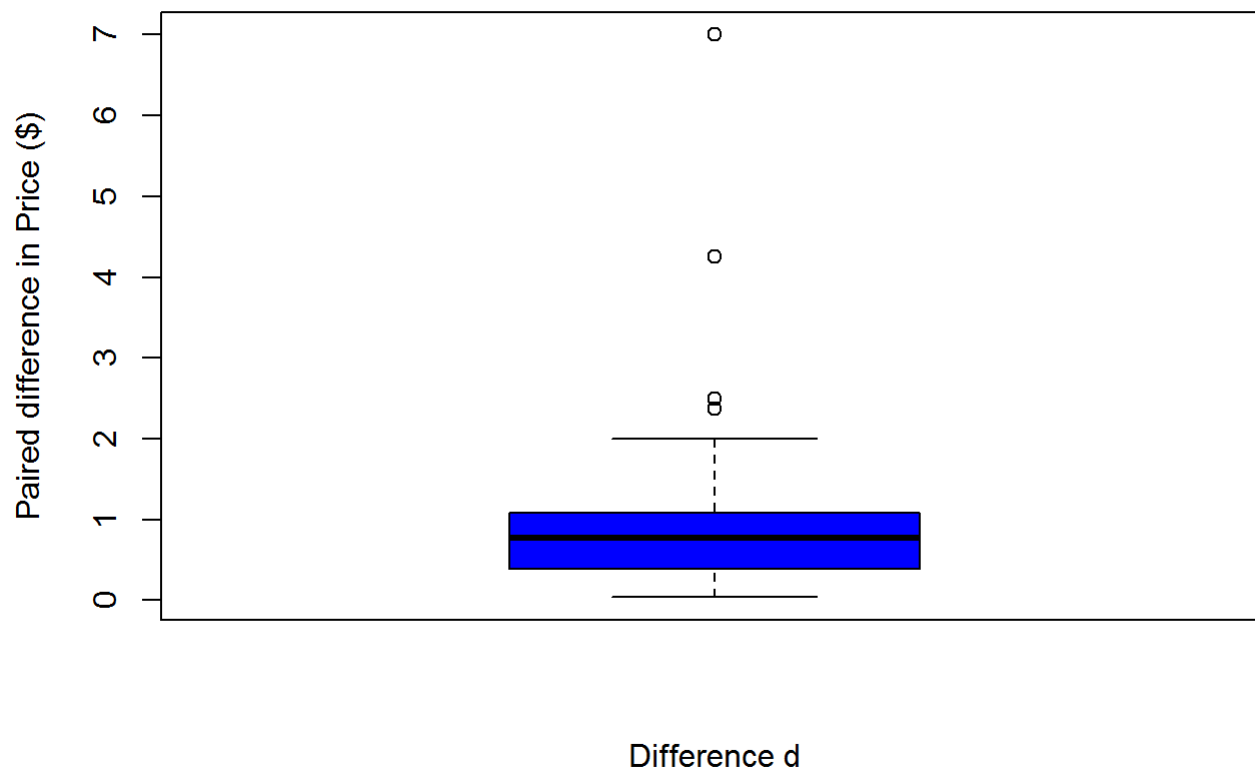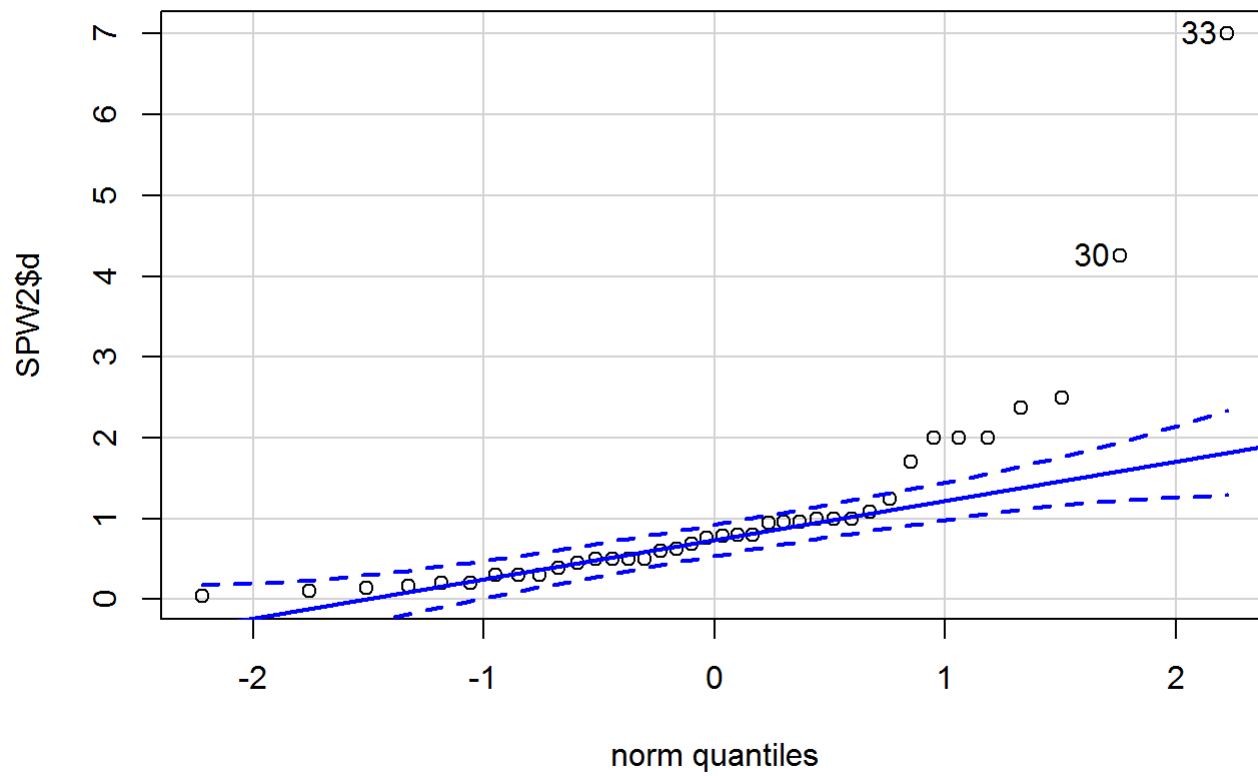
```
# Visualisation
boxplot(SPW$`Coles price`, SPW$`Woolworths price`,ylab = "Price",col=c('red','green'))
axis(1, at = 1:2,labels = c("coles", "woolworths"))
```
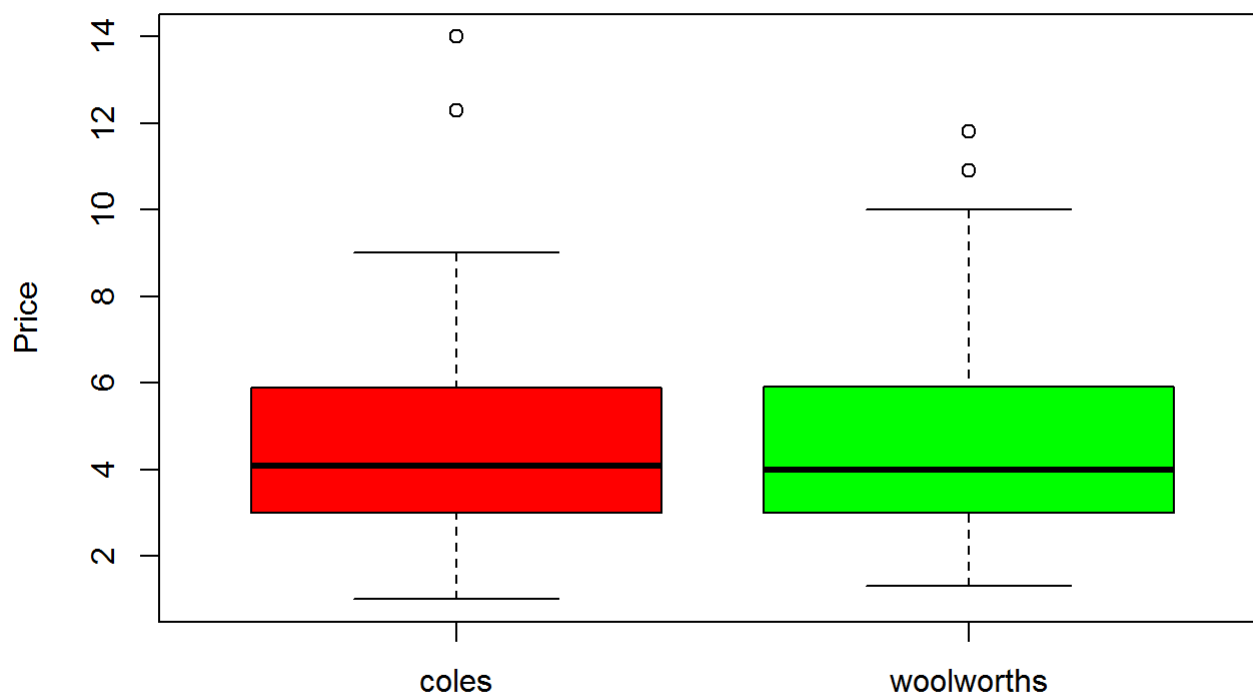
```
#Filtering the outliers
SPW2 <- SPW1 %>% filter(d >0)
boxplot(SPW2$d,col='Blue',ylab="Paired difference in Price ($)", xlab= "Difference d")
```
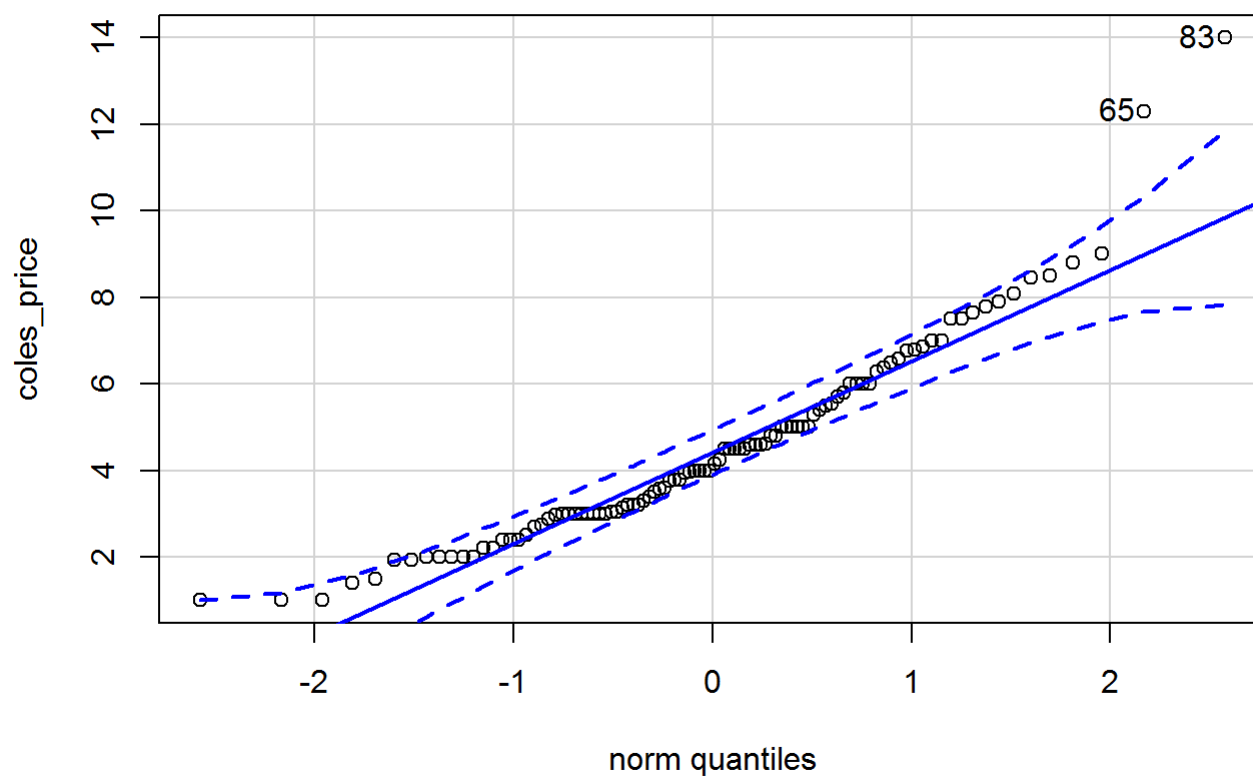


Difference d

```
#Q-Q plot
qqPlot(SPW2$d,dist="norm")
```

```
## [1] 33 30
```

```
#Coles Q-Q plot
qqPlot(SPW$`Coles price`,dist="norm",main="Coles",ylab="coles_price")
```
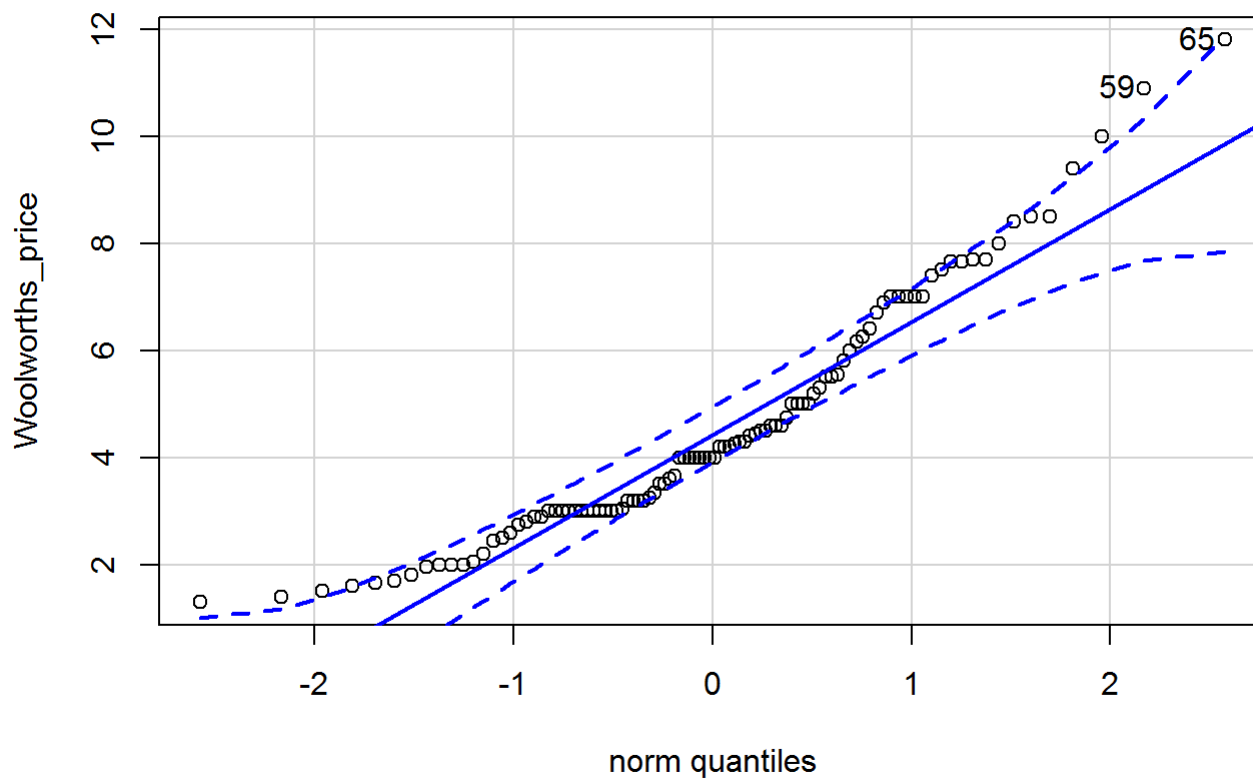
## Coles



```
## [1] 83 65
```

```
#woolworths Q-Q plot
qqPlot(SPW$`Woolworths price`,dist="norm",main="Woolworths",ylab="Woolworths_price")
```

## Woolworths



```
## [1] 65 59
```

# Hypothesis Test

The information depends on the comparison between the same products. The paired sample t-test was performed to get the p value to reach the conclusion of the hypotheis.

#Null Hypothesis Both supermarkets' mean value is equal (Ho = mu1-mu2 = 0)

#Alternate Hypothesis Both the supermarkets mean value are not equal.(HA = mu1-mu2 ≠ 0)

```
# hypothesis testing.
# Paired t-test
paired_ttest <- t.test(SPW$`Coles price`, SPW$`Woolworths price`,
                paired = TRUE,alternative = "two.sided",conf.level = .95)
paired_ttest
```

```
## 
##  Paired t-test
## 
## data:  SPW$`Coles price` and SPW$`Woolworths price`
## t = -0.11964, df = 99, p-value = 0.905
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.297193  0.263393
## sample estimates:
## mean of the differences
##                 -0.0169
```

# Interpretation

The outcomes of the hypothesis test by obtaining the intervals of p-value and confidence intervals shows that in order to analyze the paired sample t-test, the t-test led to 0.905 p-value higher than 0.05 alpha. In 95% of CI, the average distribution drops and so statistically, the outcome is not important.

# Conclusion

- A paired-sample t-test was used to test a substantial mean price distinction in Coles and Woolworths. The mean data set gap was discovered to be -0.0169 (SD = 1.412613).
- Visual inspection of the Q-Q differential scores plot suggested that the data were distributed approximately as usual.
- The t-test paired samples discovered a non-statistically important mean price distinction, t(df=100) is -0.0169, p-value is 0.905(>0.05) and trust interval of 95 percent[-0.297193 0.263393 ].
- We do not dismiss the null hypothesis, therefore.
- There is insufficient evidence to promote our alternative hypothesis and the outcome is therefore not statistically important.
- We conclude, therefore, by stating that there is no significant statistical evidence to show which supermarket is cheaper.

# Discussion

The analysis has the following findings :

- The price average was made and discovered to be comparable. In both Coles and Woolworths, there is very little distinction in product rates.

# Strengths and Limitations while collecting the data.

- Here the data is randomly collected. A sample of 100 similar products was used to conduct the investigation. The dataset taken by choosing different products from both Coles and Woolworths helps study the comparison of both the supermarkets. The products are compared and found to have very minimal differences in the prices of each product between bith the Supermarkets. Occasionally, prices may differ. If we perform distinct time period hypothesis testing, the findings might be more effective.