MATH1324 Assignment 2

Code ▾

Supermarket Price Wars

Group/Individual Details

- Phalgun Haribabu Chintal (s3702107)
- Syed Junaid Ahmed (s3731300)

Executive Statement

The aim of this investigation is to look at costs between two major Australian supermarkets, Coles or Woolworths is cheaper. With a specific end goal to discover it out, the products information is gathered through http://www.grocerycop.com.au/ (http://www.grocerycop.com.au/). A sum of 81 products was selected in a view of the product names, units that match both the supermarkets. The dataset contains six variables (Sl.no, Product name, Units, Category, Coles_Price, Woolworths_Price). All the product prices are in Australian Dollars (AUD). The logic behind taking the large sample is to limit standard error. Summary Statistics, box plot and qqplot visuals are created from the dataset. Moreover, a paired t-test is used to check the mean difference between Coles and Woolworths. The result of t-test indicates a statistically significant difference between the price of products at Coles and Woolworths.

Load Packages and Data

Hide

```
library(dplyr)
library(granova)
library(readr)
library(magrittr)
library(car)
price <- read_csv("pricewar.csv")
```

```
Parsed with column specification:
cols(
  Sl.no = col_integer(),
  `Product Name` = col_character(),
  Units = col_character(),
  Category = col_character(),
  Coles_Price = col_double(),
  Woolworths_Price = col_double()
)
```

Summary Statistics

The statistical summary and box plots for each supermarket, Coles, and Woolworths are obtained from the data collected. The price per unit (kilogram/litre/each) are used in this investigation. The price brought from Woolworths has a lower median price of A$ 4.50 and a lower mean price of A$ 5.36, contrasted with Coles median of A$ 4.79 and mean price of A$ 5.99. The qqplot of both supermarkets did not follow a normal distribution.

```
pricelist <- data.frame(Stores = c("Coles", "Woolworths"),
                        min = c(min(price$Coles_Price, na.rm = TRUE), min(price$Woolworths_Pr
ice, na.rm = TRUE)),
                        Q1 = c(quantile(price$Coles_Price, probs = 0.25, na.rm = TRUE), quant
ile(price$Woolworths_Price, probs = 0.25, na.rm = TRUE)),
                        Median = c(median(price$Coles_Price, na.rm = TRUE), median(price$Wool
worths_Price, na.rm = TRUE)),
                        Q3 = c(quantile(price$Coles_Price, probs = 0.75, na.rm = TRUE), quant
ile(price$Woolworths_Price, probs = 0.75, na.rm = TRUE)),
                        Max = c(max(price$Coles_Price, na.rm = TRUE), max(price$Woolworths_Pr
ice, na.rm = TRUE)),
                        Mean = c(mean(price$Coles_Price, na.rm = TRUE), mean(price$Woolworths
_Price, na.rm = TRUE)),
                        SD = c(sd(price$Coles_Price, na.rm = TRUE), sd(price$Woolworths_Pric
e, na.rm = TRUE)),
                        n = c(length(price$Coles_Price), c(length(price$Woolworths_Price))),
                        Missing = c(sum(is.na(price$Coles_Price)), sum(is.na(price$Woolworths
_Price))))
pricelist
```

Stores <fctr>	min <dbl>	Q1 <dbl>	Median <dbl>	Q3 <dbl>	Max <dbl>	Mean <dbl>	SD <dbl>	n <int>	Missing <int>
Coles	1.0	3.5	4.79	6.9	14.0	5.994568	3.589065	81	0
Woolworths	0.2	3.1	4.50	6.9	15.2	5.359383	3.049801	81	0

2 rows

```
pricelist_difference <- price %>% mutate(Price_Difference = r$Coles_price - r$Woolworths_pric
e)
pricelist_difference
```

Sl.no <int>	Product Name <chr>	Units <chr>
1	Lite Wrap 10 inch	567g
2	Wholemeal Pizza Bases 12 Inch 2 pack	440g
3	Gourmet Pizza Bases 2 pack	440g
4	Spinach & Herb Wraps 8 pack	567g
5	Dollar Sweets Decoration Rainbow Poppng Topping	150g
6	Frozen Puff Pastry Sheets 10 pack	2kg
7	Seafood & Vegetable Frozen Sea Shantys 12 pack	420g
8	Frozen Party Pack 30 pack	1kg
9	Ice Cream Honeycomb Crunch 4 pack	428m
10	Ice Cream Classic 4 pack	428m

Box plot is visualized to show the price comparison of Coles and Woolworths. In the dataset, both the supermarkets seem to be quite similar except outliers.

Hide

```
boxplot(price$Coles_Price, price$Woolworths_Price, ylab = "Prices", xlab = "Supermarkets", na
mes = labels,
        main="Boxplot of Prices in AUD of Coles and Woolworths", col = c("blue", "green"))
```

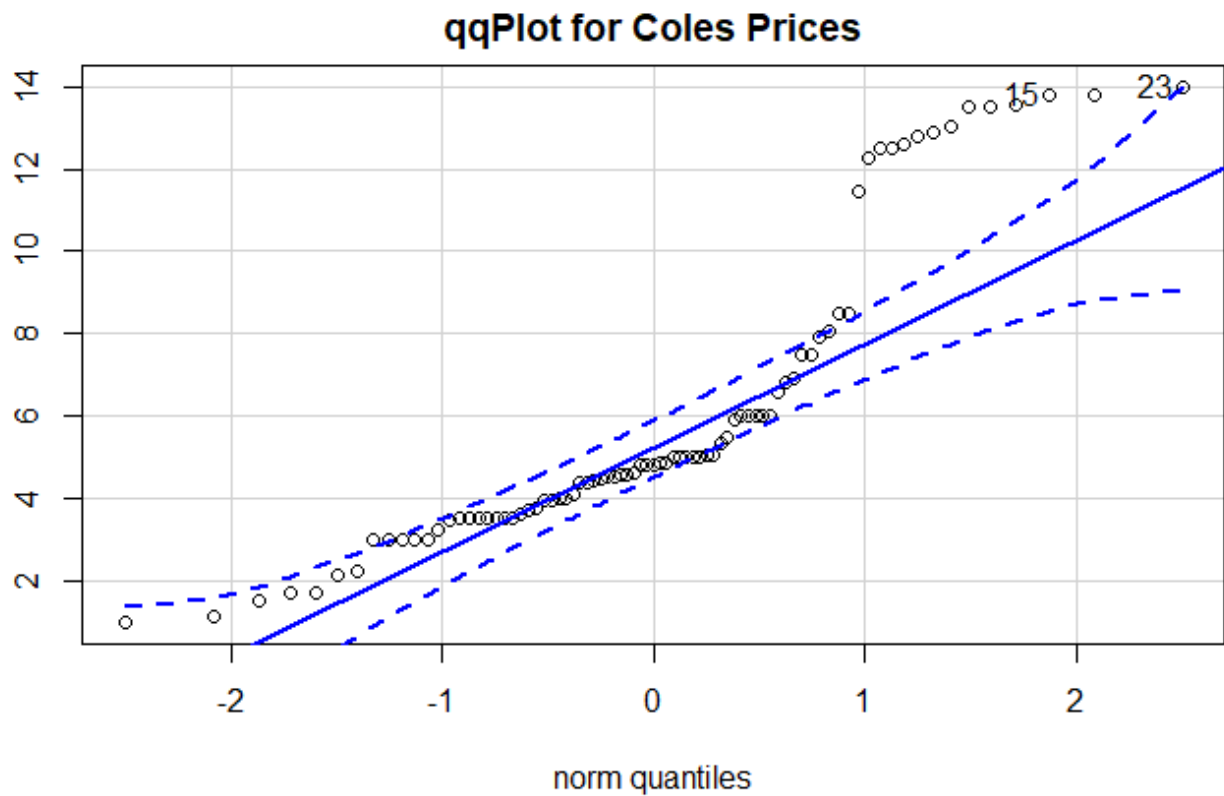## Boxplot of Prices in AUD of Coles and Woolworths

Hide

```
        labels = c("Coles", "Woolworths")
```

Hide

```
price$Coles_Price %>% qqPlot(dist = "norm", main = "qqPlot for Coles Prices")
```

```
[1] 23 15
```

qqPlot for Coles Prices

norm quantiles

Hide

```
price$Woolworths_Price %>% qqPlot(dist = "norm", main = "qqPlot for Woolworths Prices")
```

```
[1] 41 50
```

qqPlot for Woolworths Prices

norm quantiles
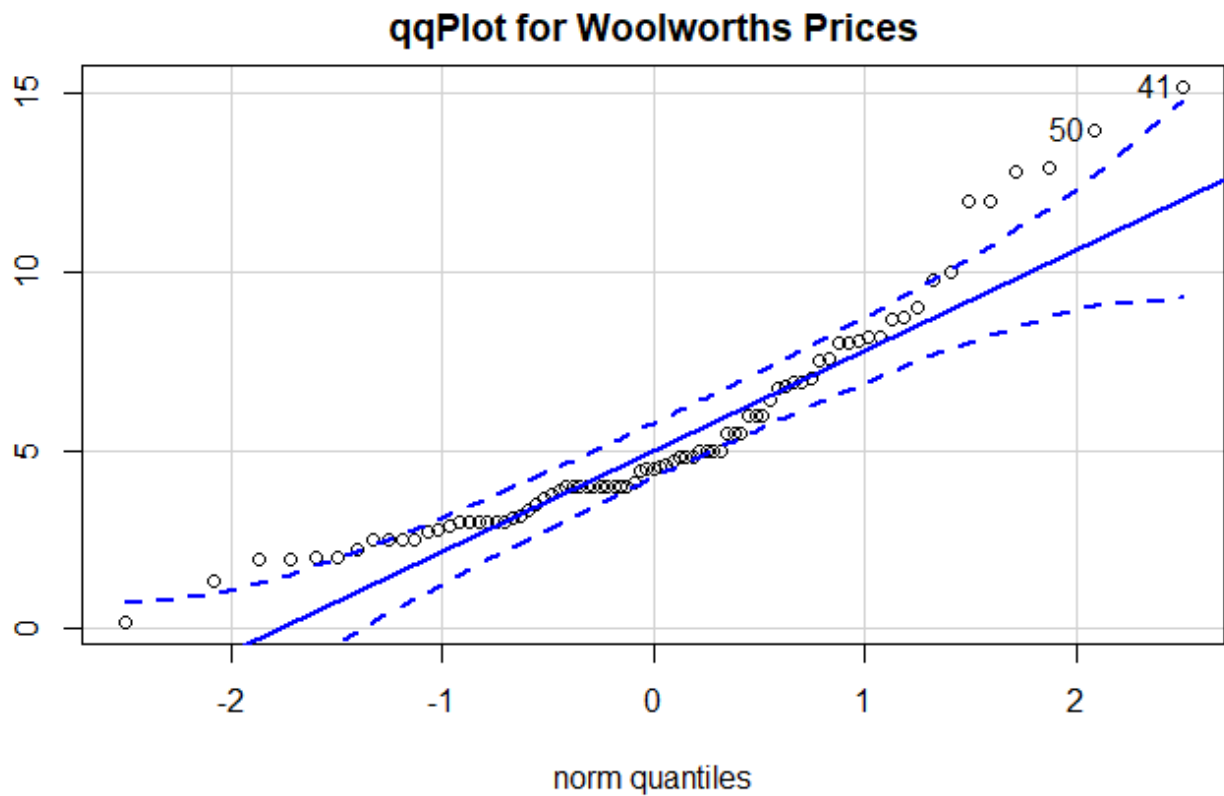
Hypothesis Test

A paired sample t-test is used in this test because all the variables are continuous and observations are depended. A confidence interval of 95% is used, with significance level 0.05. 1. H0 : $\mu\Delta$=0 2. HA : $\mu\Delta\neq$0

H0: There is no price difference between Coles and Woolworths. HA: There is a price difference between Coles and Woolworths.

Hide

```
t.test(price$Coles_Price, price$Woolworths_Price, paired = TRUE, alternative = "two.sided")
```

```
    Paired t-test

data:  price$Coles_Price and price$Woolworths_Price
t = 3.2253, df = 80, p-value = 0.001824
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.2432614 1.0271090
sample estimates:
mean of the differences
              0.6351852
```

Hide

```
granova.ds(data.frame(price$Coles_Price, price$Woolworths_Price), xlab = "Coles", ylab = "Woo
lworths")
```

```
         Summary Stats
n                81.000
mean(x)           5.995
mean(y)           5.359
mean(D=x-y)       0.635
SD(D)             1.772
ES(D)             0.358
r(x,y)            0.870
r(x+y,d)          0.315
LL 95%CI          0.243
UL 95%CI          1.027
t(D-bar)          3.225
df.t             80.000
pval.t            0.002
```

Dependent Sample Assessment Plot
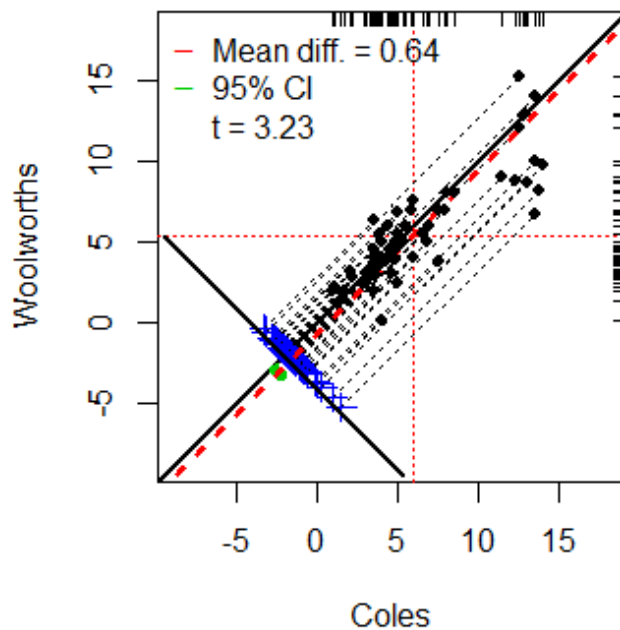
Interpretation

A paired-samples t-test was used to check for a statistically significant mean difference between the price of Coles and Woolworths. 0.64 was the mean difference along with a standard deviation of 1.772, t = 3.23, df = 80 between both the supermarkets. Non-normality was the result upon visualizing the qq plot of both the retailers' prices. The t-test resulted in a p-value of 0.001824, that is lesser than 0.05. The confidence interval turned out to be as [0.243, 1.027] does not capture the mean difference H0: μΔ = 0. The result of the investigation found to be statistically significant, or reject null hypothesis.

Discussion

The findings from the above investigation, a paired t-test found a statistically significant mean difference between the prices of products at Coles and Woolworths. Thus, the investigation concludes, Woolworths is cheaper compared to Coles.

Strengths:

The strength of this investigation ensures the right product names and its units at both the supermarkets used. The dataset was easy to observe and look into many categories.

Limitations:

- Numerous items and categories are not covered during this investigation.
- The price list of products at Coles and Woolworths had a big challenge as the costs fluctuate from one store to another.
- Regular prices of products are included while, excluding the discounted prices.

Future attempts:

Furthermore, an investigation in the future should have an increase in dataset. It must include all genre and items along with the discounted products.

Conclusion:

Overall, from the results achieved, it shows that Woolworths are cheaper compared to Coles.