

MATH1324 Assignment 3

Code ▼

Supermarket Price Wars: Coles vs Woolworths

Group/Individual Details

- Parikshit Sreedhar (s3643796)
- Arjun Balaji (s3629999)
- Ravi Madhurchand Pandey (s3638787)

Executive Statement

Given a choice of the retailer and budget constraint on the customer's pocket, which retailer would customers prefer to fill their basket with ?

- In working through the context; data on products, brands, unit of measurement and price had been collected from the retailers: *Coles* and *Woolworths*
- Products from both retailers were matched with same unit of measurement
- The original prices of the products were sampled
- A test was carried out to determine any statistically significant difference in their product prices
- There seems to be a marginal difference in product prices which does not influence the choice of the retailer

Data collection

How the data was collected ?

Stratified Sampling

- We a team of 3 people, strolled across the aisles at *COLES* and *WOOLWORTHS* supermarket; randomly sampled products and noted down their prices.
- The variables that were taken into consideration are as follows:
 - *Product Name*
 - *Brand*
 - *COLES Price*
 - *WOOLWORTHS Price*
 - *UNIT OF MEASUREMENT*
 - *Aisle*
 - *Price_bucket* - Additional variable

Load Packages and Data

Hide

```

library(dplyr)
library(Rcmdr)

setwd("C:\\Users\\Wel\\Downloads\\Intro to stats\\Assignment 3")
price_war<-read.csv("assign3_final_data_updt.csv")

price_war<-price_war%>%select(2:7)
price_war<-price_war[complete.cases(price_war),]

##Price data for COLES
coles<-price_war%>%select(1:3,5,6)
coles<-mutate(coles,Store="Coles")
names(coles)[3]<-"Price"

##Price data for woolworths
woolworths<-price_war%>%select(1:2,4:6)
woolworths<-mutate(woolworths,Store="Woolworths")
names(woolworths)[3]<-"Price"

##Combining both the data sets
price_war<-rbind(coles,woolworths)
price_war$Price<-round(price_war$Price,2)
price_war$Store<-as.factor(price_war$Store)      #coercing the Store column t
o a factor

```

Summary Statistics

Code

No.of products available at *Coles* and *Woolworths* based on Price Bucket

Hide

```

#An additional feature called "price_bucket" was derived from the price variable.
table(price_war$Store,price_war$price_bucket)

```

	0-5	5-10	10-15	15-20	20-25	25-30	30+
Coles	76	23	2	1	1	1	0
Woolworths	74	24	3	1	1	1	0

Code

Store <fctr>	Min <dbl>	Q1 <dbl>	Median <dbl>	Q3 <dbl>	M.. <dbl>	Mean <dbl>	SD <dbl>	n <int>	Missing <int>
Coles	0.75	2.85	3.75	5.4225	29	4.574327	3.706597	104	0
Woolworths	0.06	2.80	3.65	5.4900	29	4.650577	3.866743	104	0

2 rows

Hide

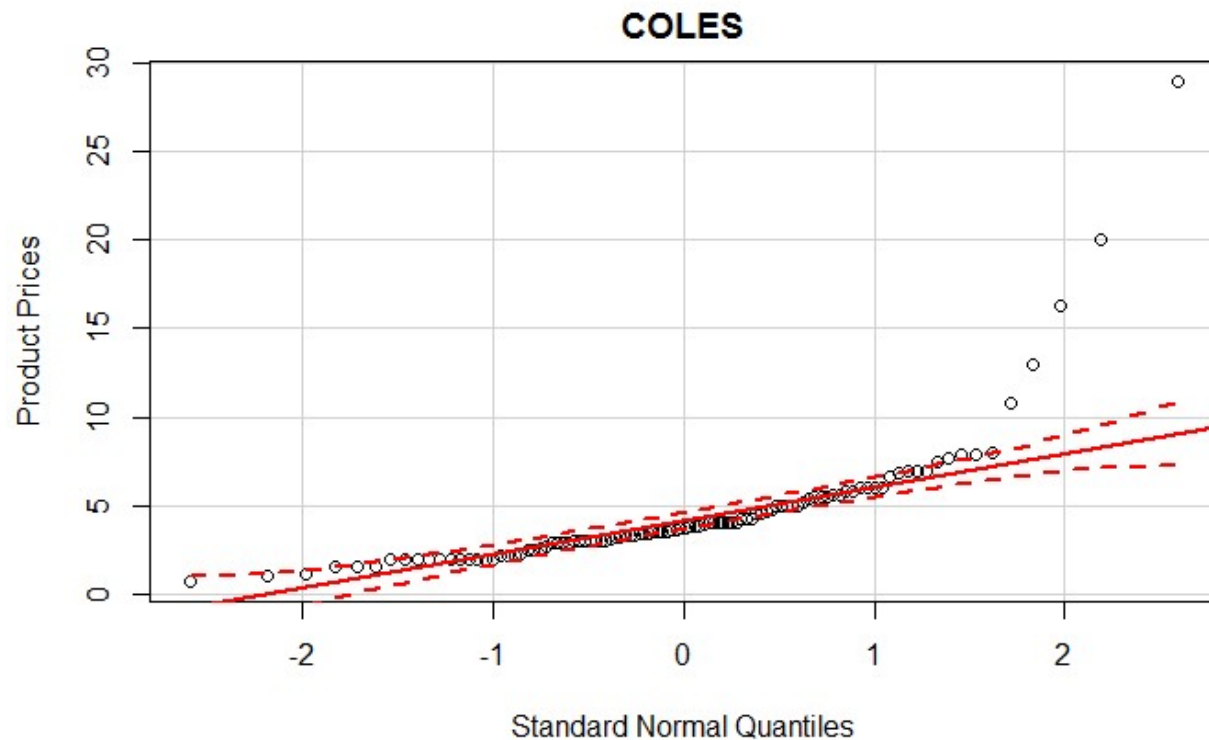
```
##Boxplot of the PRICE variable
boxplot(price_war$Price~price_war$Store,ylab="Product Prices"
        ,xlab="Retailer"
        ,main="PRICE WAR\nComparison of product prices between COLES and WOOLWO
RTHS"
        ,col="Green")
```



The above box plot confirms a near to same distribution of product prices in both the stores

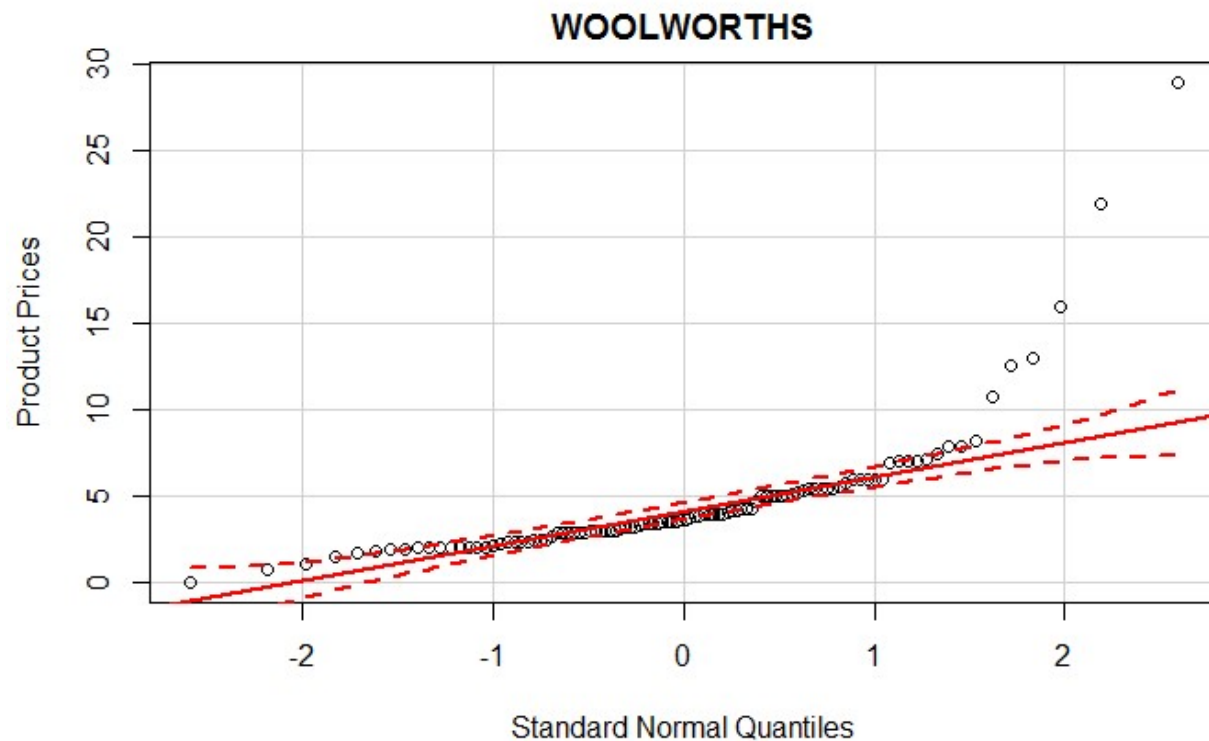
Hide

```
##Eventhough we have a large sample, we may want to check if our data comes from a normal distribution
##We shall examine this using a Q-Q plot
coles<-price_war%>%filter(Store=="Coles")
woolworths<-price_war%>%filter(Store=="Woolworths")
coles$Price%>%qqPlot(dist="norm",main="COLES",xlab="Standard Normal Quantiles",ylab="Product Prices")
```



Hide

```
woolworths$Price%>%qqPlot(dist="norm",main="WOOLWORTHS",xlab="Standard Normal Quantiles",ylab="Product Prices") ##There is a mild departure from normality
```



Hypothesis Test

- **Assumption:** The sample of product prices of *COLES* and *WOOLWORTHS* are independent of each other
- The NULL hypothesis is that the mean difference in prices of *COLES* and *WOOLWORTHS* is zero
- The alternate hypothesis is that the mean difference in prices of *COLES* and *WOOLWORTHS* is NOT zero
- A two sample t-test will be carried out with a confidence level of 95% and significance level of 5%
- We use the p-value approach to interpret the *t-test statistic*
- Since, there are more than 30 data points in both the samples, we need not assume the population distribution of product prices of *COLES* and *WOOLWORTHS* to be normally distributed

Product prices of both *COLES* and *WOOLWORTHS* have the same variance. This is also verified using the Levenes test conducted below:-

Hide

```
##Homogeneity of Variance test  
leveneTest(Price~Store,data=price_war)
```

Hide

```
##Performing two sample t-test
t.test(price_war$Price~price_war$Store
       ,paired=FALSE
       ,data = price_war
       ,var.equal=TRUE
       ,alternative="two.sided")
```

Interpretation

- The test is not statistically significant
- The p-value is greater than 0.05.
- The probability of observing a difference in product prices as high as the t-test statistic value is greater than the 5% significance level
- The confidence interval houses the null hypothesised value of zero
- There is now insufficient evidence to reject the *NULL HYPOTHESIS*

Hence the customers need not be weary of their choice of the retailer based on price point of the respective products

Both COLES' and WOOLWORTHS' product prices come from the same distribution of prices

Discussion

Findings and Conclusion:

- Both *COLES* and *WOOLWORTHS* seems to have similar pricing strategy
- In order to conduct an efficient analysis of this kind, the quantitative variables from both the samples(*COLES* and *WOOLWORTHS* in our case) should have the same unit of measurement for the respective products
- The whole analysis hinges on product matching and standardisation of product prices and unit of measurement
- Based on the study conducted, product prices at *COLES* and *WOOLWORTHS* have a negligible difference

Limitation - Price Point: Price point is not the only factor having a bearing on the customer's choice of retailer. There are other factors such as brand spread offered at either of the stores, discounts, loyalty programs, location, etc.. that could influence the buying decision/customer footfalls.

Strengths: The size and equivalence of our respective data points is the core strength for this analysis.

Analysis tweaking: Another way to investigate the problem is to collect in-house products of *COLES* and *WOOLWORTHS* respectively and carry out the same analysis.