ECSA Graduate attribute project

Part 1-6

23572000- A Oosthuizen

Contents

Introduction **2**

1.Data Wrangling **2**

2.Descriptive statistics **2**

2.1. Analysis of valid data set 2

2.2. Visualisation of categorical features 3

2.3. Visualisation of continuous features 4

2.4. Process Capabilities Calculations 8

3.Statistical Process Control **8**

3.1. X&S- Charts and limits 8

3.1.1.X Chart Limits- 8

3.1.2. S Chart Limits- 9

3.1.3. Chart Analysis (First 30 samples) 9

4. Optimising the delivery process **13**

4.1. Inspection of X-Charts 13

A - Sample means outside Outer Control Limits 13

B- Most consecutive samples of s-bar between -0.3 and +0.4 sigma control limits 14

4.2. Type 1 Error 15

4.3.Minimizing technology delivery costs 15

4.4. Type 2 Error 16

5. DOE & MANOVA **16**

5.1 First Hypotheses 17

5.2. Second hypotheses 19

5.3. Third hypotheses 20

6. Reliability of service & products **22**

6.1. Taguchi Loss 22

6.2. Reliability of machines 23

6.3. Expecting reliable delivery 23

6.3.1. Reliable deliveries 23

6.3.2. Increase number of vehicle by 1 to 22 24

Conclusion **25**

References **26**

Introduction

We are to analyze, manipulate and provide feedback on 180 000 entries of sales data for a specific retailer. Exploratory data analysis was used to perform initial examinations on the data in order to discover patterns,to spot abnormalities,to test hypotheses and to explore assumptions by means of descriptive statistics and graphical representations. The retailer are interested in factors, such as product pricing, classes of products sold, reasons for buying a product and the delivery times of their products, therefore, we will identify trends and relationships between these features.

In order to process and analyse the data, it has to be pre-processed. The data is cleaned by removing all invalid records and separated into valid and invalid data sets. The valid data set is used to familiarize ourselves with the relationships in the data in order to formulate x-charts for the different classes and to determine optimal delivery times,costs and reliability of the products.

1.Data Wrangling

In the data set “salesTable2022”, there are invalid instances that need to be removed, therefore, valid and invalid data needs to be separated from each other to be able to perform accurate analysis on the data. Differences are found within the invalid data set. 17 missing values are found in the Price feature column, no calculation can be done with these instances and the analysis will be inaccurate if these values were to be included, therefore they are removed. There are 5 negative values found in the Price feature. The price of an item cannot be a negative value, which is why these instances are invalid and must be removed to gather an accurate analysis of the data set. There were no values in the price feature that were extremely high,except for the different classes,however, this is normal as different categories are priced differently; therefore, it is assumed that there are no visible outliers that will greatly impact the results. After identifying all the invalid instances, 179978 instances remain in the data set that are documented correctly. These instances form the ‘Valid’ dataset that will be used for further analysis.

2.Descriptive statistics

2.1. Analysis of valid data set

A data quality report was generated from the valid data set as a basis for understanding the data set and the types of features:

Analysis of continuous features-

Descriptive statistics were implemented by means of rcode and the following statistical estimates are obtained of certain numerical features in the valid data set-

variables	min	Q1	mean	median	Q3	max	outlier	n_missing	card_con
X	1.00	45004.25	90002.54638	90004.50	134999.75	180000	0	0	179978
AGE	18.00	38.00	54.56552	53.00	70.00	108	0	0	91
Price	35.65	482.31	12294.09837	2259.63	15270.97	116619	22171	0	78832
Delivery.time	0.50	3.00	14.50031	10.00	18.50	75	17516	0	148

The 'Age' feature ranges between 18-108 with an average of 55, therefore, most sales are done by customers aged 55 ,which will define this age group as their target audience. The shortest delivery time is 0.5 hours and the longest is 75 hours, this equals to a 3 day difference which is something to look into if the retailers follow a responsive strategy. The average price of a sale is R12 294, with the cheapest price being R35.65 and the most expensive being R116 618.The big difference in prices can be due to the cost of different classes ,cost of production or willingness of a customer to pay more. The 'X' feature has very high cardinality (as high as there are instances), therefore, this feature does not contribute to giving information of the sales and therefore will not be used in further visualizations. There are outliers in the price and delivery time features, however, these values are necessary to conclude relevant information about these items, therefore, these outliers are used in the data.

Analysis of categorical features-

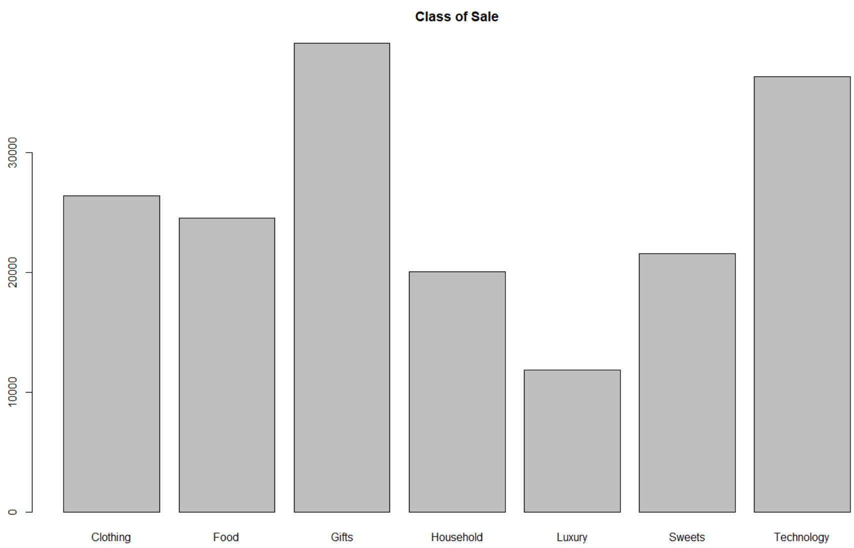
classname	countC	missing_values	cardinality	Modes	Mode_frequency	mode_percentage
ID	179978	0	15000	41842	27	0.0150018335574348
Class	179978	0	7	Gifts	39149	21.7521030348154
Year	179978	0	9	2021	33443	18.5817155430108
Month	179978	0	12	12	15225	8.45936725599796
Day	179978	0	30	17	6126	3.40374934714243
Why bought	179978	0	6	Recommended	106985	59.4433764126726

ID's count is not the same as the unique cardinality. This could be a mistake, since this implies that almost 30000 instances have the same ID. Gifts are the most popular sale. The most frequently bought is in the year 2021, 17th of December. This makes sense, as this is close to Christmas, when buying gifts are essential. The company received the majority of their sales due to recommendations from other clients.

2.2. Visualisation of categorical features

Class of Sale-

Most sales were gifts and technology related items. The least number of sales were for the luxurious items, but this can be due



to more expensive pricing

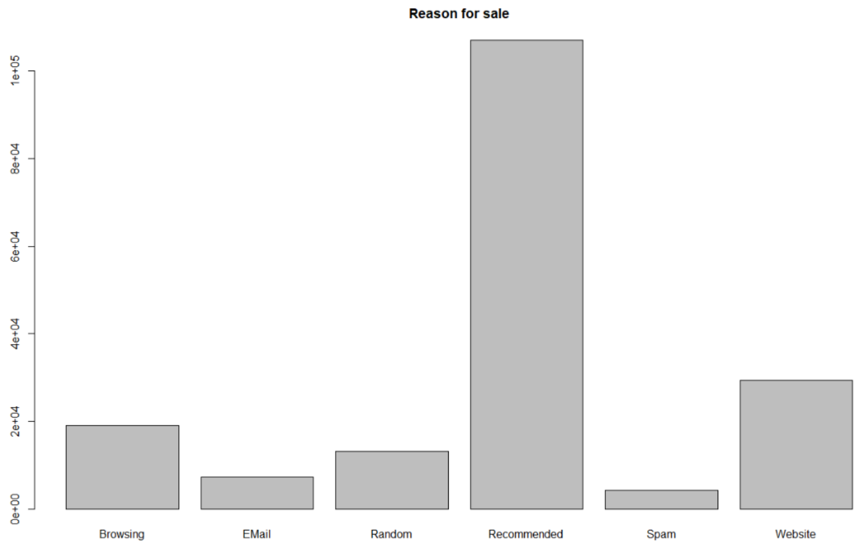
and the ‘not a need, but

nice to have’ quality of a

luxurious items. Technology can be high due to the increasing use in modern companies and households.

Reason for Sale-

Most customers were recommended to buy their products, which is an indication of good quality service or products,since customers are talking about it and in turn promoting their sales. The low email, spam and

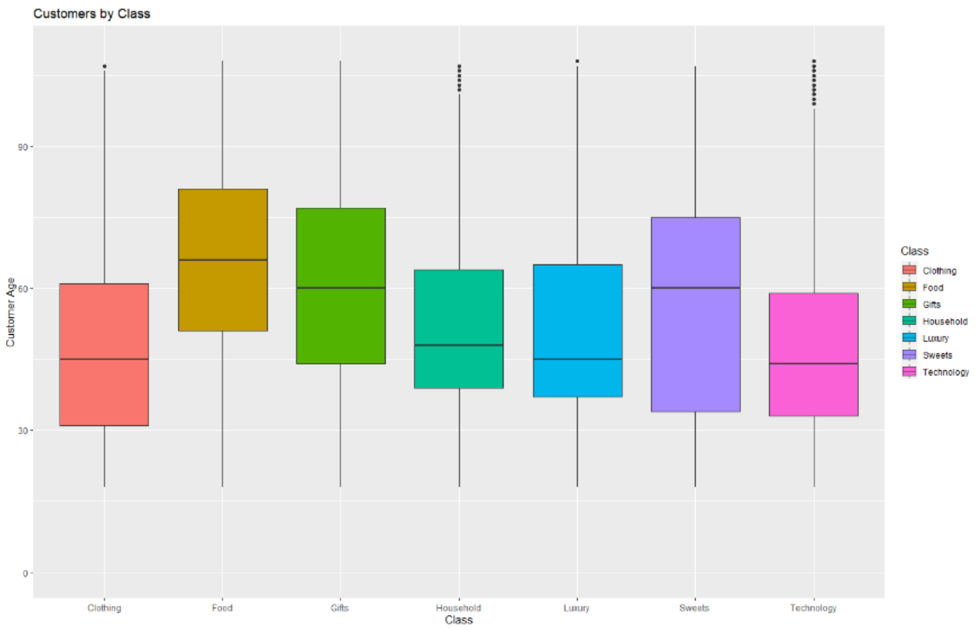


browsing frequencies give an indication that the business should improve their advertising resources, as the cost of advertising is large compared to the sales it brings.

2.3. Visualisation of continuous features

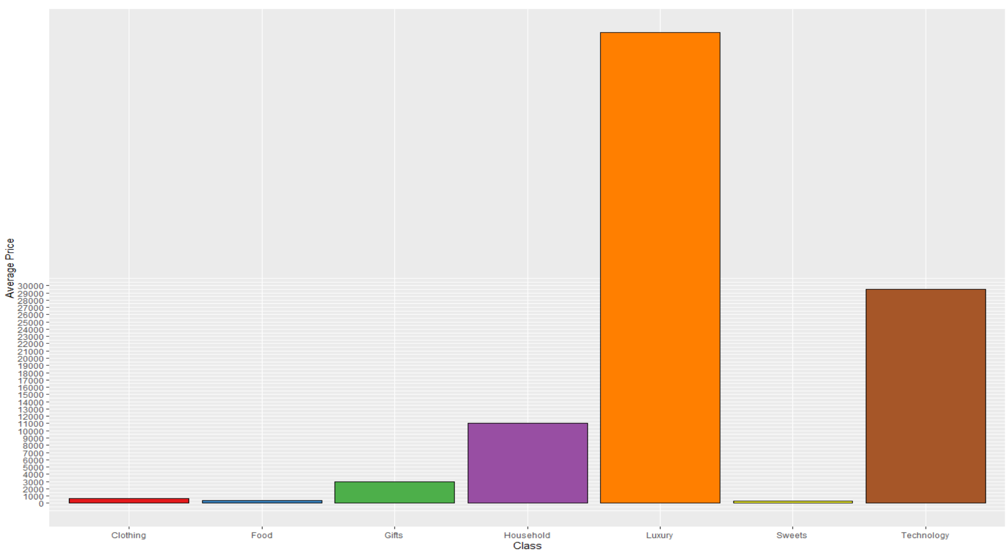
Age vs Class-

This plot gives an idea on which age range prefers which class of product. Clothing and Household products are mostly sold to ages of +-50 whereas Food, Sweets and Gifts are mostly sold to ages of +-65. Technology and Luxury products are mostly sold to the ages of 50 .It is clear that most products are bought by customers between the ages of 50 and 70. All the classes pass the age of 90-100, we can assume that retirement homes or elderly make use of the online platforms to buy these products , therefore it is not necessary to disregard these ages. The food,clothing and gifts are equally distributed, which makes sense as these are items that appeal to all age groups, however, food and gifts have a higher mean, which can also be due to more elderly ordering via online or delivery platforms. The household and luxury items have a interquartile range of +- 70-40, this makes sense as the majority of these age groups have fixed jobs and households to look after and the wish to improve their lifestyles.



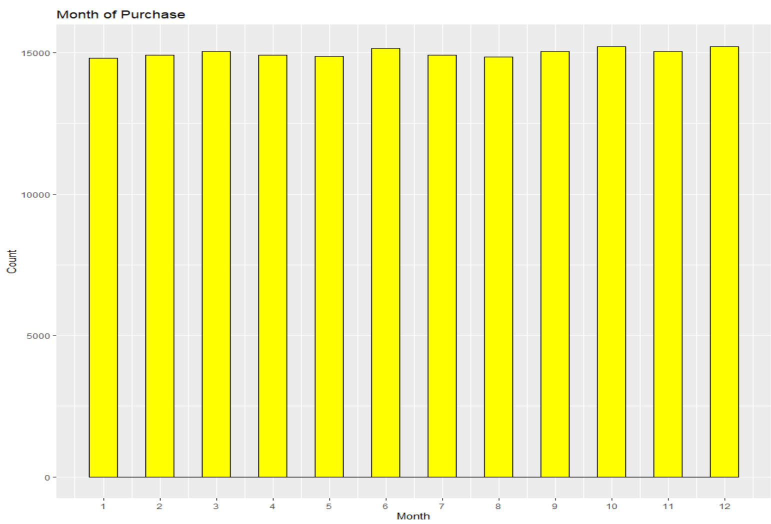
Average price per Class-

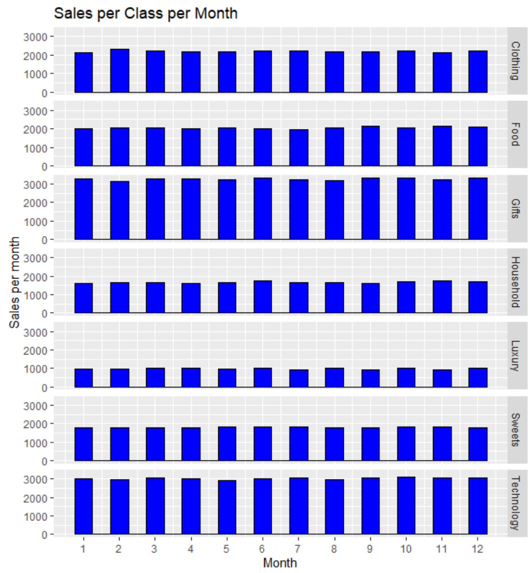
This graph shows which Class obtains the highest prices. From this the Luxury Class is by far the most expensive class with an average price much higher than R30 000, this is because luxurious items are items that are of high quality and one of a kind. Technology has an average of R20000, whereafter the other classes follow with much lower prices in the R1000-12000 range. Technology's higher price range is justified due to the increasing demand of higher modernization and specialized production. Food and clothing are items that all people need, therefore, either the government puts subsidies on these products or restricts suppliers to increase these prices by means of tax or floor prices, so that these products are accessible by most people.



Sales per Month per Class-

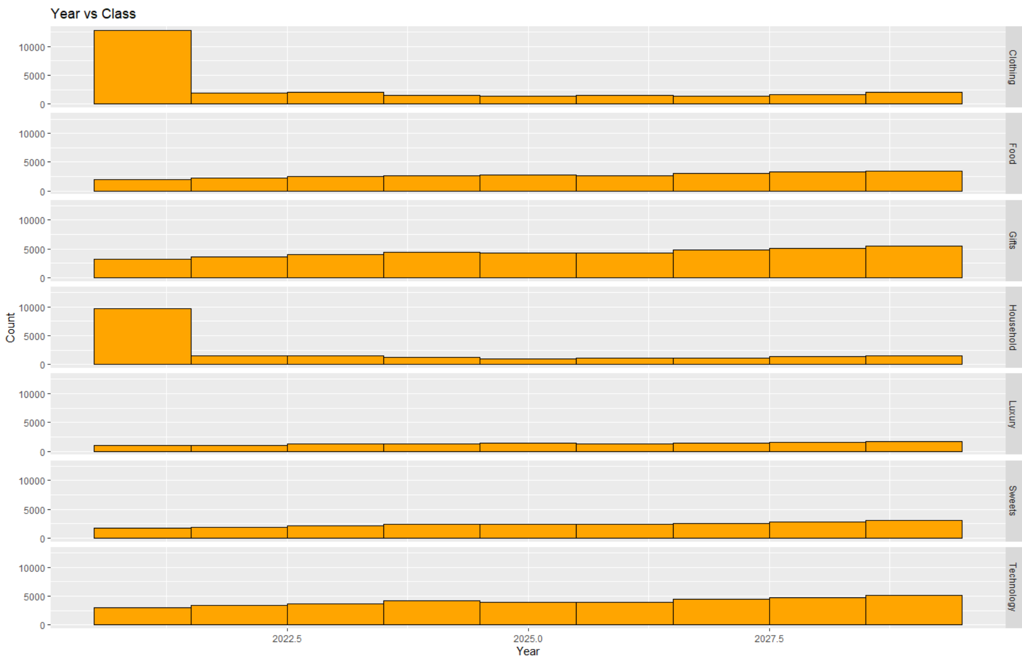
From these graphs we can see that the sales per month do not change much, this indicates that there was no seasonality or change in trends throughout the year, meaning the sales per class per month is uniformly distributed. There is a small spike in sales in months 6,10 and 11; this can be due to holiday seasons. The company can improve their sales per month by catering for specific trends, introducing promotions or improving their advertising.





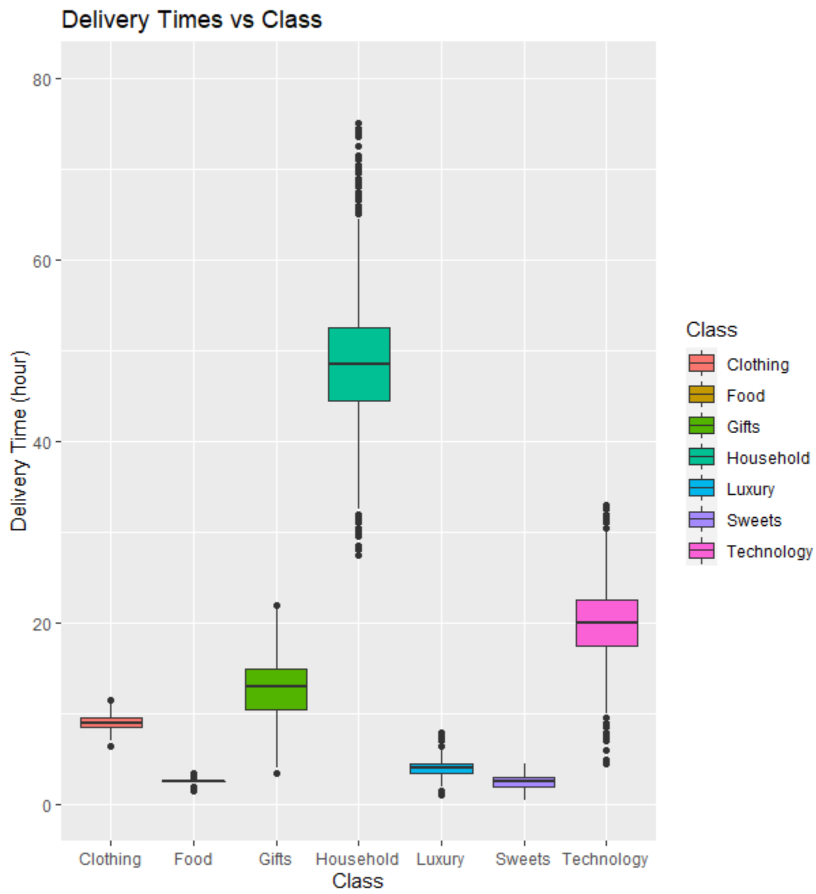
Yearly sales vs Class-

This graph can be used to determine trends and seasonality of items. Less clothing and household items were bought throughout the years, which makes sense as these are long lasting items. More technology and gifts were bought as the years increased, which agrees with the increase in modernization.



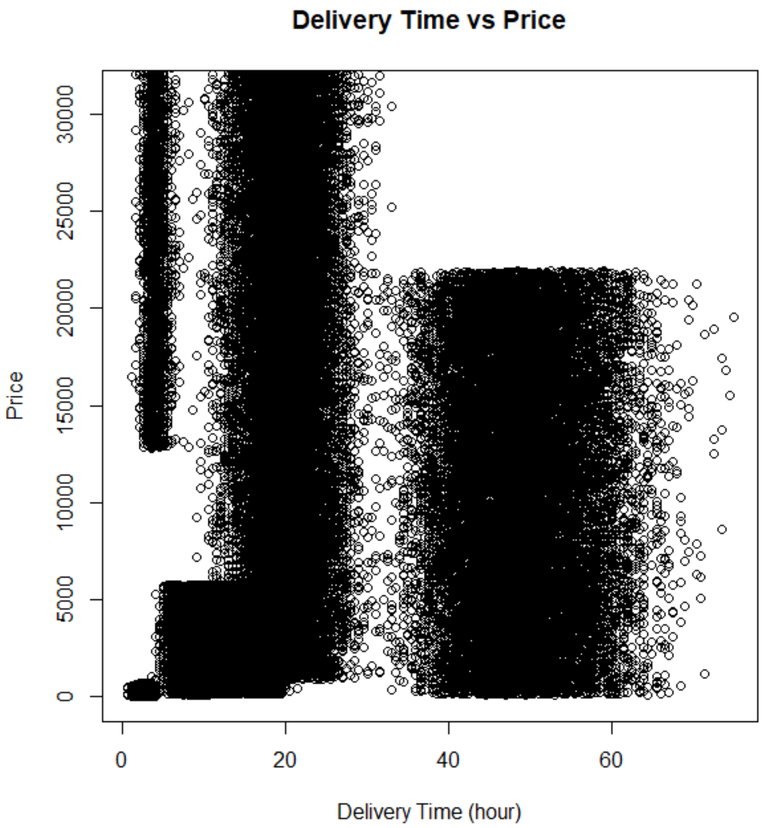
Delivery time per Class-

From this graph information is obtained on the delivery time of each Class. Households have the highest delivery times. Food and Sweets have the shortest delivery times since they are perishable items. Clothing also has short delivery times and this can be to accomplish competitive advantage, since clothing is a common item for retailers to sell.Customers that buy luxury goods prefer high quality and responsiveness, therefore, to cater for these customers the company delivers these products fast.



Delivery time vs Price-

This graph shows the correlation between delivery time and how it affects the price of an item. If it takes longer to produce & deliver a product, the price will be reduced to compensate for the waiting time. Whereas, if the delivery time is shorter, the price will be higher due to the responsive nature of the production. Products' prices in



between these delivery times will depend on the nature of the class.

2.4. Process Capabilities Calculations

```
USL <- 24
LSL <- 0
#determine the cp, cpu, cpl, cpk
technology <- validData[which(validData$Class == "Technology"),1:10]
sDevTechnology <- sd(technology$Delivery.time)
avgTechnology <- mean(as.integer(technology$Delivery.time))
Cp <- (USL-LSL)/(6*sDevTechnology)
Cpu <- (USL-avgTechnology)/(3*sDevTechnology)
Cpl <- (avgTechnology-LSL)/(3*sDevTechnology)
Cpk <- min(Cpu,Cpl)
CpVal<- cbind(Cp,Cpu,Cpl,Cpk)
print(CpVal)
```

```
           Cp       Cpu      Cpl       Cpk
[1,] 1.142207 0.4035293 1.880884 0.4035293
```

Why is an LSL of 0 logic?

An LSL of 0 is logical because the lower specification limit cannot go lower than 0, i.e., it is not possible to deliver an item in less than zero days and therefore we set the LSL to zero.

About the indices-

If the Cp values are less than 1 it means that the process can't meet the specifications. The Cp of 1.142207 means that the delivery of the technology class can deliver the products within the specifications required. A CpK value of 1.33 is desirable if the process is within targets and considered capable. (www.itl.nist.gov, n.d.)

3.Statistical Process Control

3.1. X&S- Charts and limits

3.1.1.X Chart Limits-

X- chart is a type of control chart that is used to monitor the mathematical means of successive samples of constant size, in this case 30 samples of 15 each. This type of control chart is used for characteristics that can be measured on a continuous scale, in this instance, the delivery process times. The data was sorted from oldest to newest before determining these limits.

CLASS	UCL	UCL2	UCL1	CL	LCL1	LCL2	LCL
CLOTHING	9.4047	9.2598	9.1149	8.97	8.8251	8.6802	8.5353
HOUSEHOLD	50.2462	49.0182	47.7902	46.5622	45.3342	44.1062	42.8783
FOOD	2.7119	2.6383	2.5647	2.4911	2.4175	2.3439	2.2704
TECHNOLOGY	22.9745	22.1138	21.253	20.3922	19.5315	18.6707	17.8099
SWEETS	2.9007	2.7597	2.6188	2.4778	2.3368	2.1958	2.0548
GIFTS	9.4879	9.1123	8.7367	8.3611	7.9855	7.6099	7.2343

LUXURY	5.4847	5.2335	4.9823	4.7311	4.4799	4.2287	3.9776

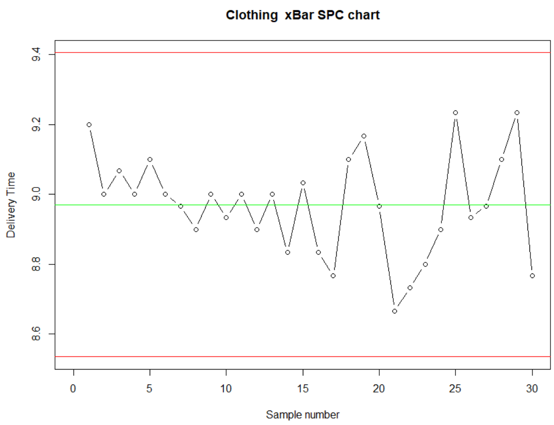
3.1.2. S Chart Limits-

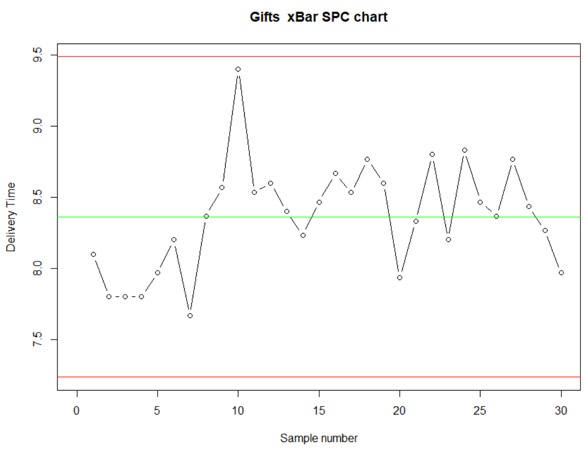
An S-chart is a type of control chart used to display the process variabilities,i.e.,standard deviation,when measuring subgroups at regular intervals from a process.

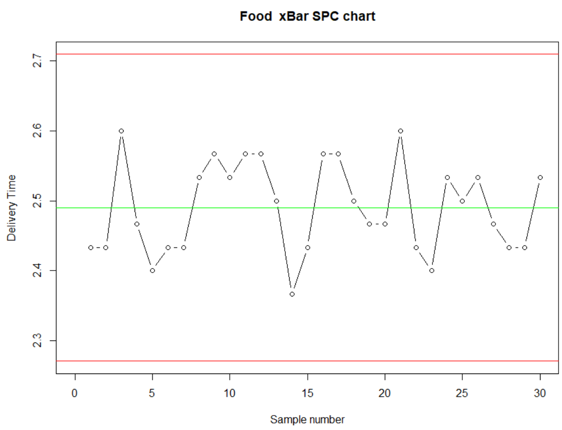
CLASS	**UCL**	**UCL2**	**UCL1**	**CL**	**LCL1**	**LCL2**	**LCL**
CLOTHING	0.8664	0.7614	0.6563	0.5512	0.4462	0.3411	0.236
HOUSEHOLD	7.3432	6.4528	5.5623	4.6719	3.7814	2.891	2.0005
FOOD	0.44	0.3867	0.3333	0.2799	0.2266	0.1732	0.1199
TECHNOLOGY	5.1473	4.5231	3.899	3.2748	2.6506	2.0264	1.4023
SWEETS	0.843	0.7408	0.6386	0.5363	0.4341	0.3319	0.2297
GIFTS	2.246	1.9737	1.7013	1.429	1.1566	0.8842	0.6119
LUXURY	1.5021	1.3199	1.1378	0.9556	0.7735	0.5913	0.4092

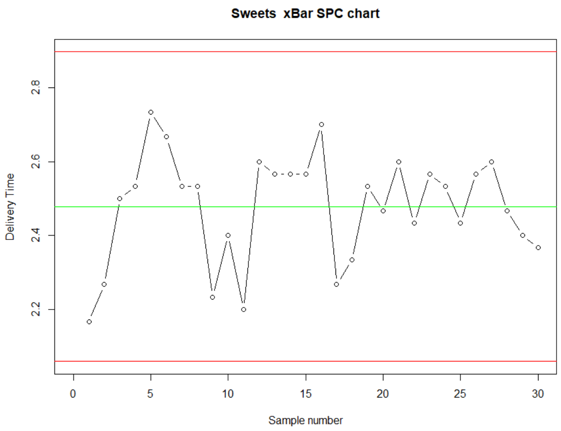
3.1.3. Chart Analysis (First 30 samples)

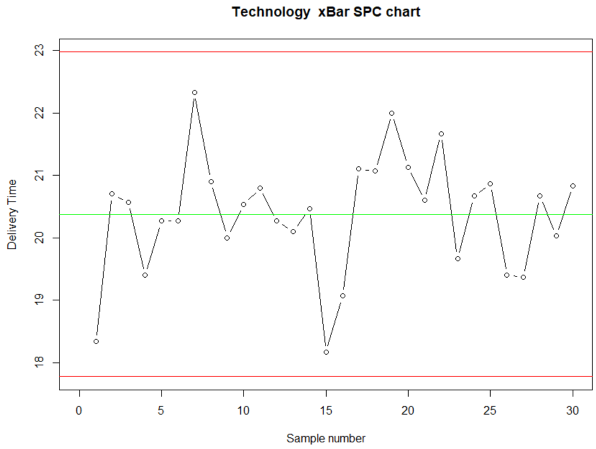
The first 30 samples (subgroups of 15 per sample) were used to approximate the mean and standard deviation of the delivery process time of each class. The control limits are then calculated for the mean of each subgroup. The first 30 samples should be in control and to display whether it is, the samples of each class are plotted on X-charts.

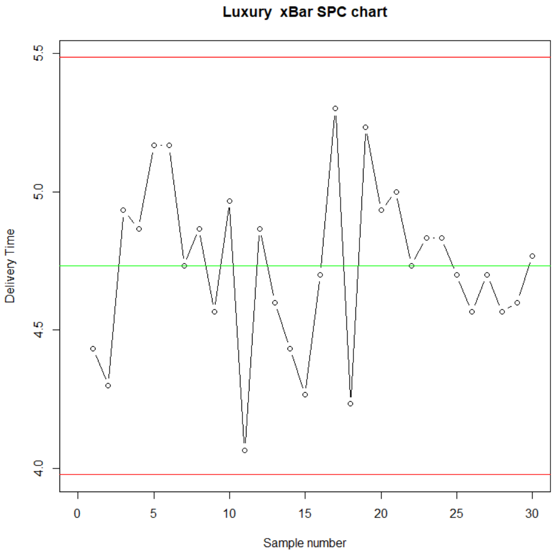










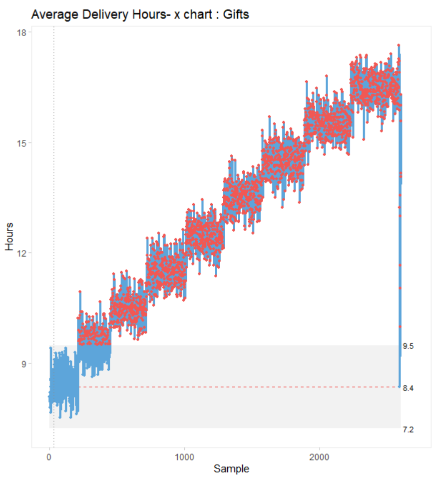


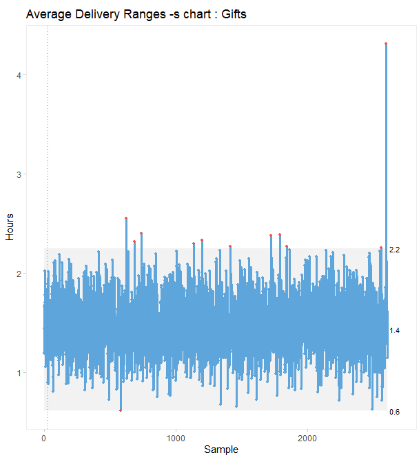
As seen by the first 30 samples of each class above, they are in control, since all the sample means lie in between their UCL and LCL (red lines). The variation of the mean associated with processes can therefore be considered as in control, if referring to the first 30 samples as an estimate for the population.

3.2. Process control

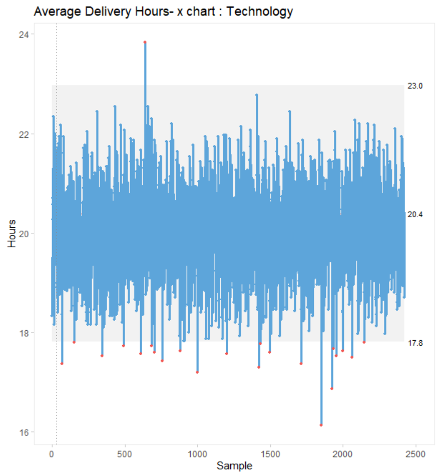
In the charts below the black dotted (vertical) line represents the 30th sample and the red dotted (horizontal) line represents the mean of the first 30 samples. If the sample is outside of the control limits, it is represented with red dots.

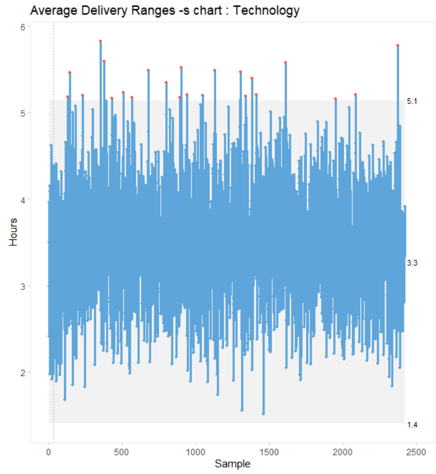
GIFTS



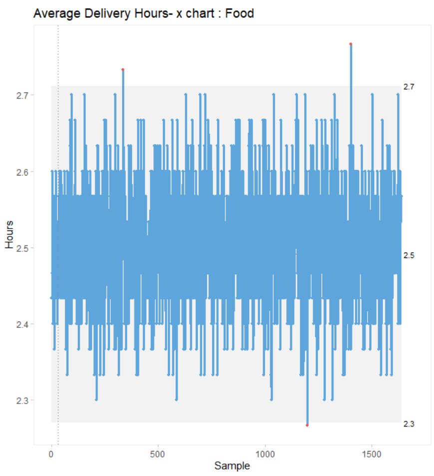


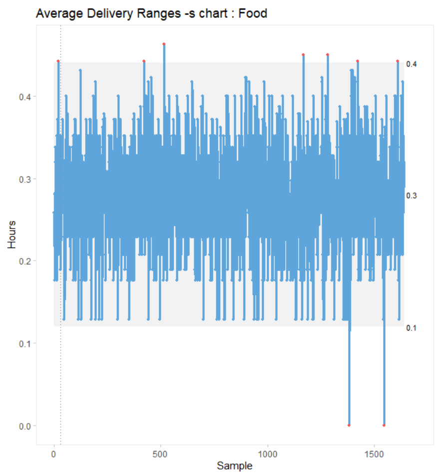
TECHNOLOGY



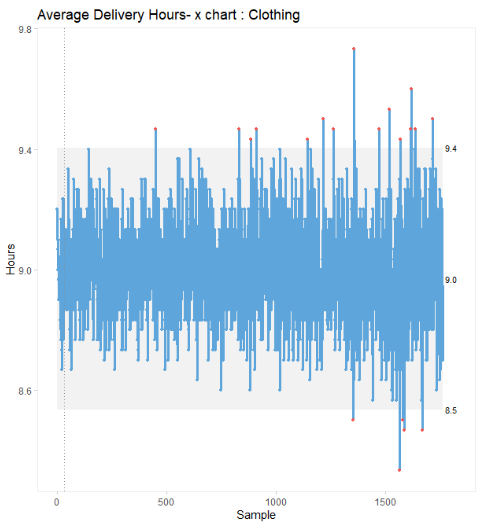


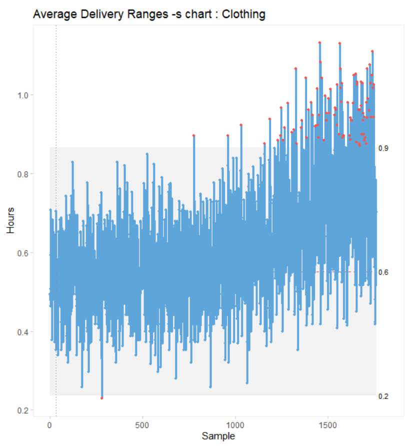
FOOD





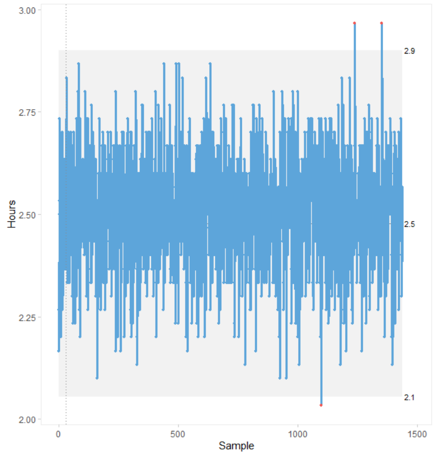
CLOTHING



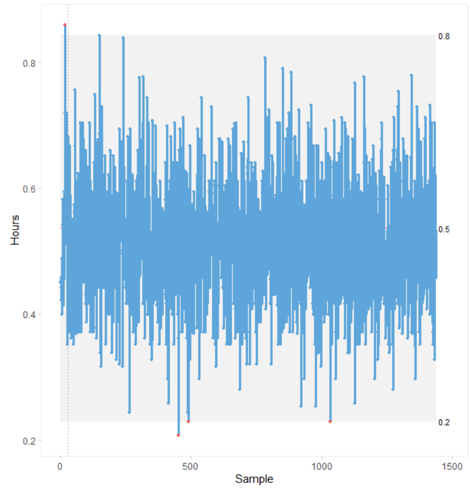


SWEETS

Average Delivery Hours- x chart : Sweets

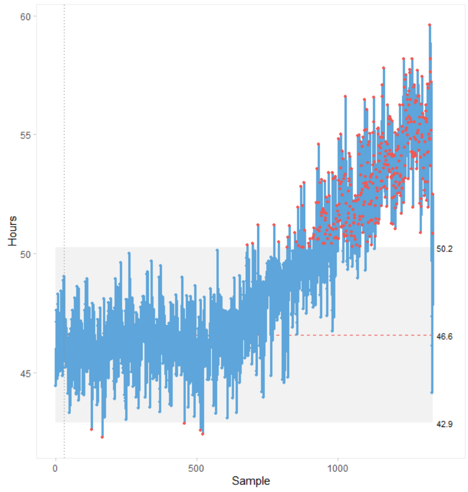


Average Delivery Ranges -s chart : Sweets

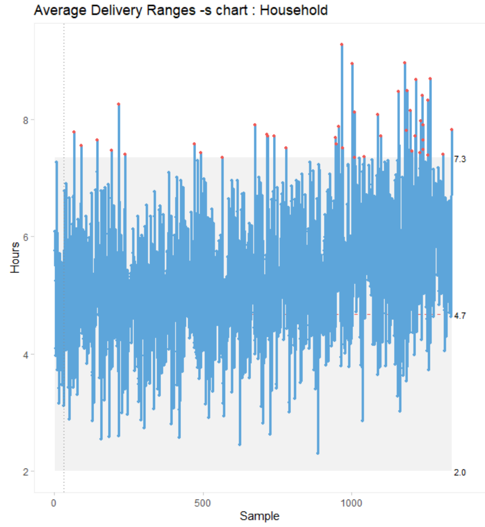


HOUSEHOLD

Average Delivery Hours- x chart : Household

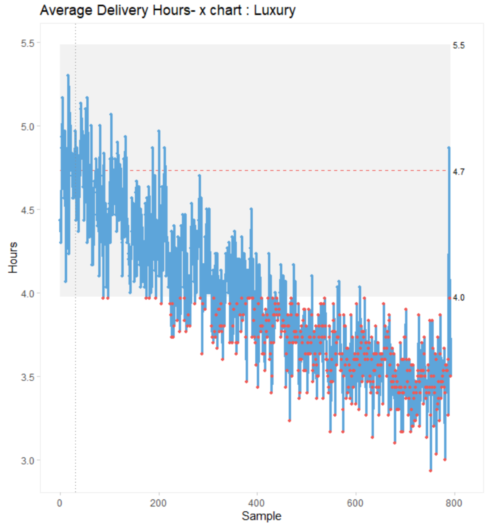


Average Delivery Ranges -s chart : Household

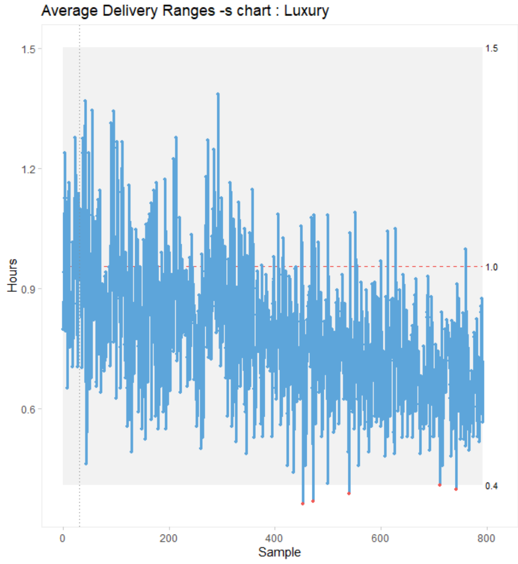


LUXURY

Average Delivery Hours- x chart : Luxury



Average Delivery Ranges -s chart : Luxury



When looking at the graphs, the mean of the process is monitored as the process continues (samples increase). The graph above can provide useful information with regards to which processes are stable or are out of control.

Household, Luxury and Gifts are out of control processes, as their points are increasingly deviating from the control limits with an increase in samples and therefore, these classes are considered not stable. Luxury's delivery times are reduced with an increase in samples, which can be a marketing strategy as the company shortens the waiting time for customers that pay the highest.This can be a possible reason for the increase in household and gift delivery times, as they dont have the capacity to deliver these items in time, due to making luxury items their priority. These processes would have to be improved and once improved, measurement can be taken again.

The processes for the Technology, Clothing, Food and Sweets classes are mostly stable with some instances out of the control limits, therefore mostly in control. The special causes of variation for these 'out of control instances' should be documented and eliminated to prevent it from happening again.

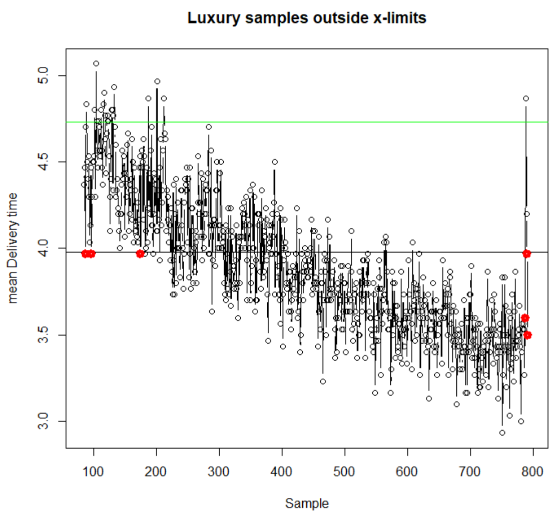
4. Optimising the delivery process

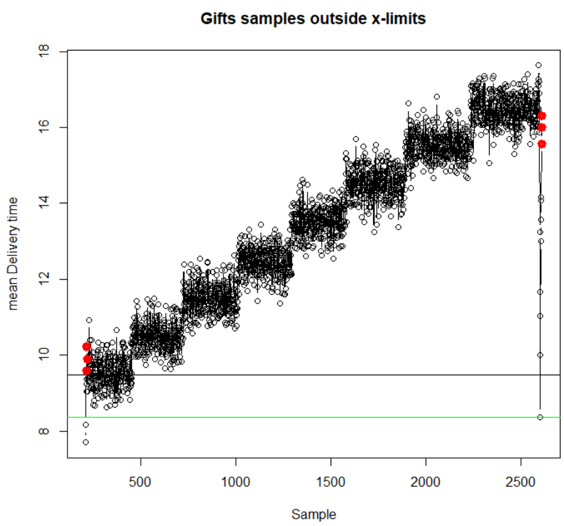
4.1. Inspection of X-Charts

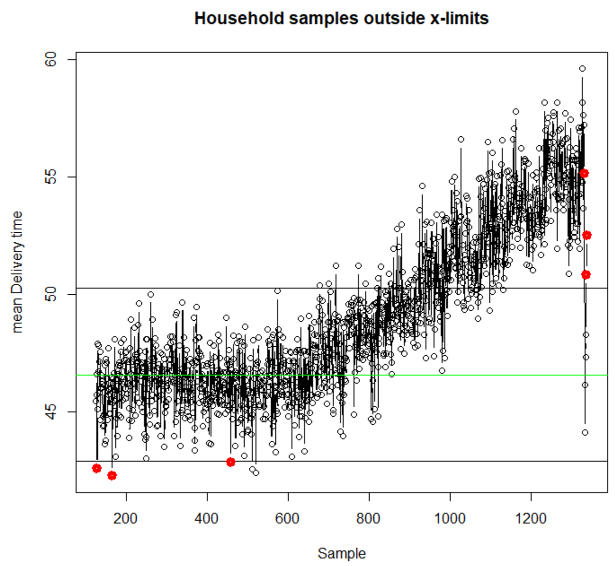
A - Sample means outside Outer Control Limits

As mentioned above Gifts, Household and Luxury are out of control processes and it can be seen in the table below as well, as it has the most sample means that are outside of the control limits. Delivery times can be affected by substandard delivery reliability, which is a common attribute in ineffective outbound warehouses. This can also be a result of improvements in delivery times for Luxury and a degradation in delivery times for Gifts and Household items. If delivery is too much for the warehouse , it is possible that the company doesn't have access to a mode of transport that can accommodate all the products from the different classes, which in turn will increase delivery times for certain classes.

Class	Total found	1st	2nd	3rd	3rd Last	2nd Last	Last
Clothing	20	450	832	885	1635	1667	1713
Household	393	128	165	457	1331	1336	1337
Food	3	336	1197	1401	NA	NA	NA
Technology	23	67	152	344	2000	2062	2147
Sweets	3	1099	1238	1351	NA	NA	NA
Gifts	2288	212	215	217	2607	2608	2609
Luxury	442	87	97	175	787	790	791





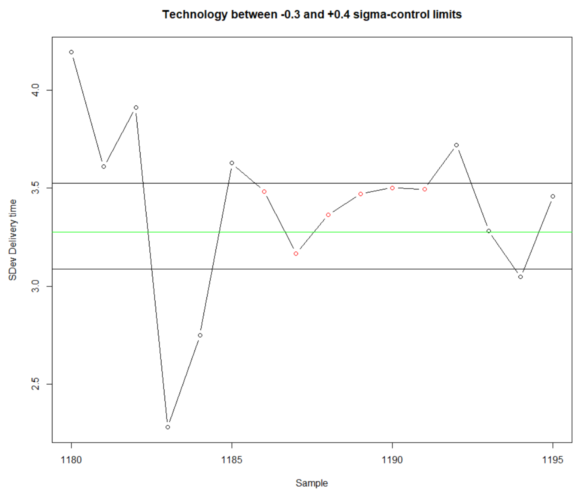


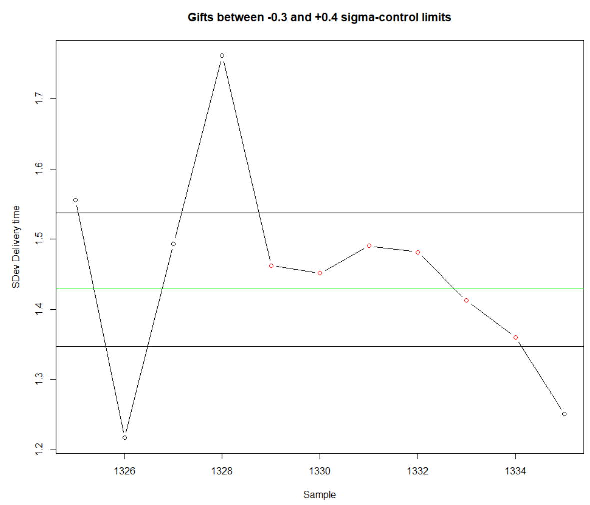
Above are the plotted graphs of the classes with the most consecutive samples outside the x-means limits, with the green line representing the mean and the black lines the upper/lower limits. The red dots are the first and last 3 samples outside the control means.

B- Most consecutive samples of s-bar between -0.3 and +0.4 sigma control limits

Class	maximum between -0.3 & 0.4 sigma	Position of first	Last Sample position
Clothing	5	665	665
Household	4	253	761
Food	5	752	905
Technology	6	1191	1191
Sweets	5	692	692
Gifts	6	1334	1334
Luxury	3	230	230

Gifts and Technology has the most consecutive samples between the -0.3 and +0.4 sigma control limits, therefore they will be the most stable within those limits out of the other classes. The red dots represent the samples within the given boundaries and the green line represents the s-chart mean.





4.2. Type 1 Error

A type 1 error occurs when it is believed that a process is out of control, when it is actually in control,i.e, we reject the Ho hypotheses, when actually it should have been accepted.By looking at the control charts, values are identified that are out of control which may indicate that the process is out of control when in reality, it is not out of control.

The null hypothesis assumption is used to calculate a type 1 error:
H0 = mean of the classes process within the control limits.
H1 = mean of the classes process is not within the control limits.

SPC indicated control	Process is fine	Process is not fine
SPC-Process is not fine	Type I Error or Manufacturer's Error	Correct to fix process
SPC-Process is fine	Correct to do nothing	Type II Error or Consumer's Error

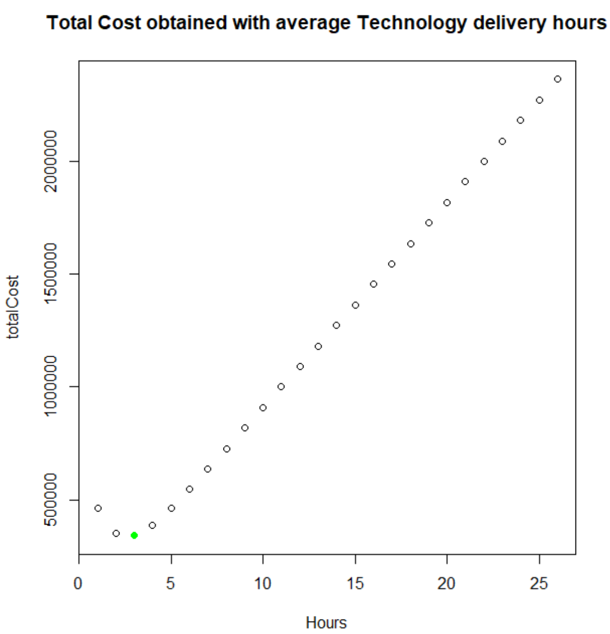
The table shows the probabilities of making type 1 errors for the different classes for 4.1. A and B. For A the probability of a sample falling outside +-3 (six sigma) is used and for B the probability of a point falling outside 0.4 and -0.3 is used. (X-bar and s Charts, n.d.)

Rule	Probabilities	Probability %
A	0.00269979606326019	0.269979606326019
B	0.726666836200723	72.6666836200723

4.3.Minimizing technology delivery costs

" If you lose R329/item-late-hour in lost sales if you deliver technology items slower than 26 hours, and it costs you R2.5/item/hour to reduce the average time by one hour, on how many hours should you centre the delivery process for best profit?"

To determine the minimum delivery (optimal) cost for Technology, it is necessary to compare the costs of all of the different hours and find the hour that has the lowest cost associated with it. This comparison can be obtained by means of a loop through all of the possible delivery times and its associated costs.



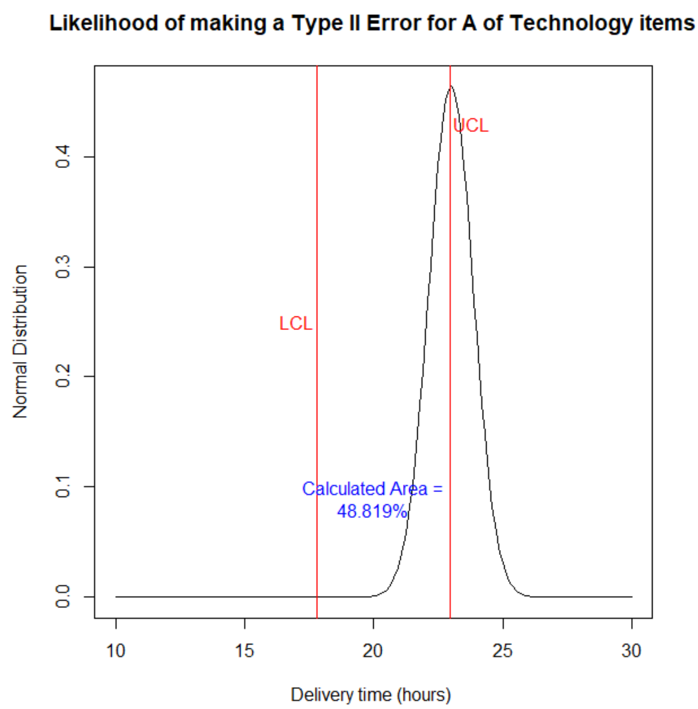
It can be seen that the best centre for delivery times for Technology is 3 hours, as seen in the plot above. This implies that by reducing the delivery times of each delivery by 3hours, and so also the weighted average from 20 to 17 hours. The cost associated with this average is then R340 870, which is then the minimum cost associated with this class. This means that the costs of reducing delivery times will be less than the costs accrued for delivery times over 26 hours.

4.4. Type 2 Error

A type ll error is the error that takes place when the company fails to reject the null hypothesis that is in reality, actually false. Therefore, a type ll error produces a false negative.

SPC indicated control	Process is fine	Process is not fine
SPC-Process is not fine	Type I Error or Manufacturer's Error	Correct to fix process
SPC-Process is fine	Correct to do nothing	Type II Error or Consumer's Error

The probability of making a type 2 error for the Technology class if the average change from 20.01095 to 23 is 0.4881974, thus having a 48.81% possibility of making an error. This error will result in a higher variation of delivery times for Technology,i.e,the average delivery times of the process will be longer than actually stated, which can cause customers to be unhappy.



5. DOE & MANOVA

The MANOVA is an analysis of variance with two or more continuous outcome variables. MANOVA is used to test the following hypotheses:

5.1 First Hypotheses

H0: The price and delivery times of an item **do not depend** on the class of an item.

H1: The price and delivery times **depend** on the class of an item.

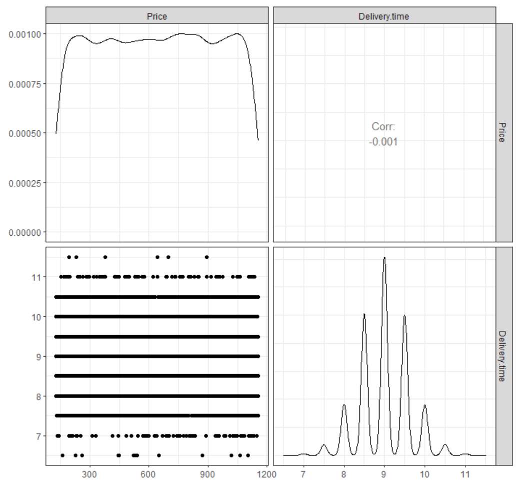
```
              Df Pillai approx F num Df den Df    Pr(>F)
Class          6 1.6797   157291     12 359942 < 2.2e-16
Residuals 179971
```

A p value of 0.05 is chosen to do the MANOVA test,since 0.05 is the most popular and universal p value. The test is not significant as the p-values actually obtained are less than 0.05. Therefore, the **H0 hypothesis is rejected** and it is clear that the price and delivery times depend on the type of class.

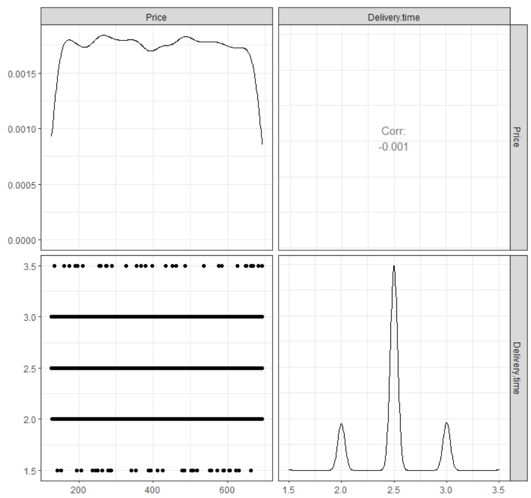
Dependant variable	P-value	Analyses
Delivery time	2.2e-16	P<0.05. Delivery time differs depending on the class of the product.
Price	2.2e-16	P<0.05.Price differs depending on the class of the product.

Firstly, the pairwise relationship between the price and delivery times was reviewed below:

Clothing

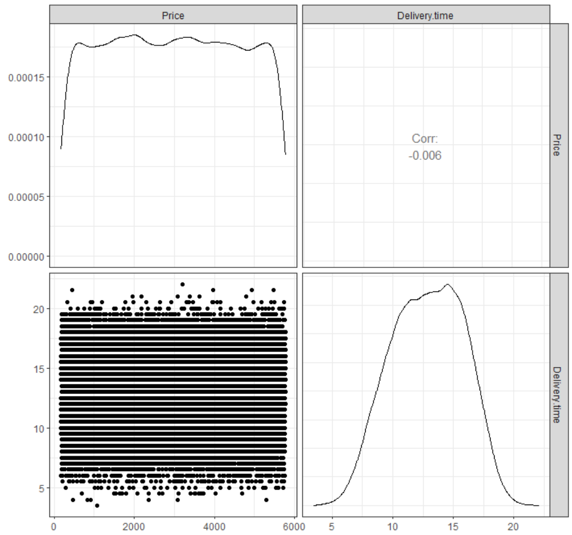


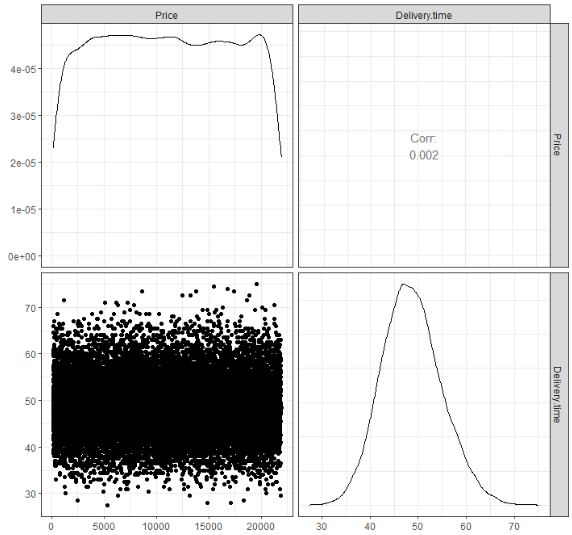
Food



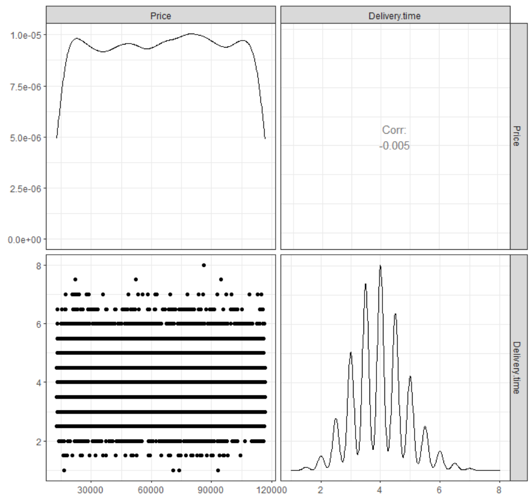
Gifts

Household

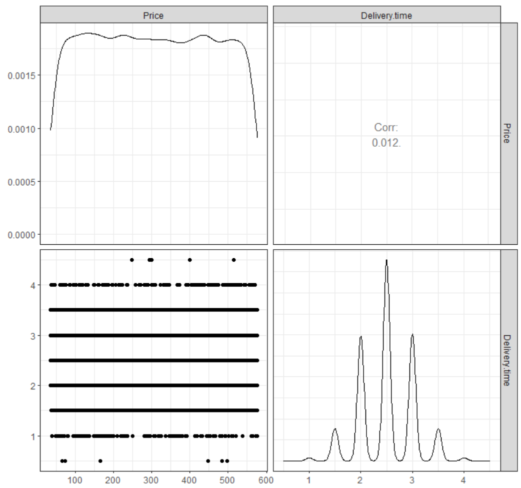




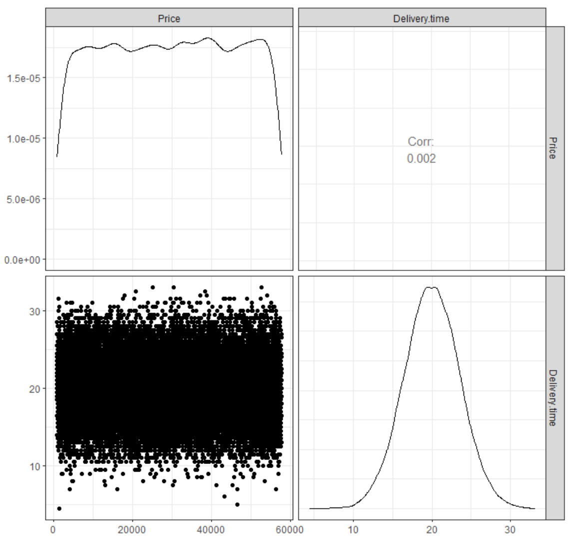
Luxury



Sweets

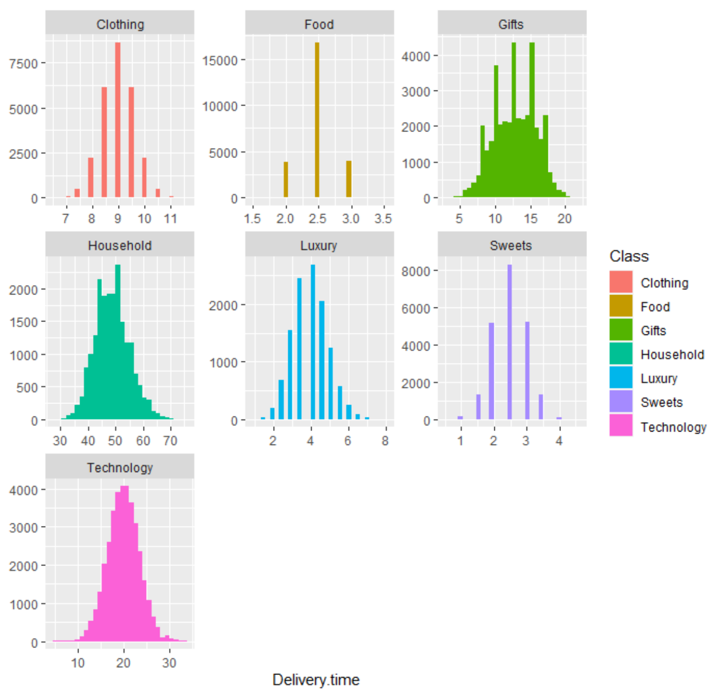


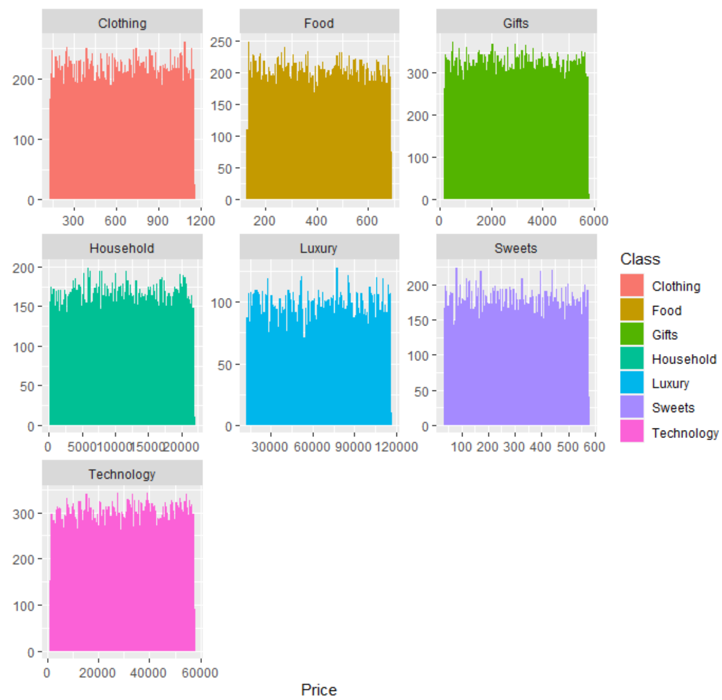
Technology



From the graphs above, it is clear that the delivery times follow a normal distribution and that the prices do not. It is also clear that for some constant price, the delivery times are not consistent and vary between a range of values for one type of class .This implies that there is not a strong linear relationship between the dependant variables, as it is not necessarily so that with the increase of price, delivery time will increase/decrease. The company should employ a more efficient logistics partner or find a more consistent system for delivering items of the same price range at the same times, especially if the items are from the same class. This will improve revenue.

The graphs below represent the different distributions between classes and how the range differs between classes. The prices are not densely packed, which indicates that there are fixed prices for items of different classes with different quality or production methods. The range of delivery times are also specific per class, indicating the direct relationship between these variables. The delivery times per class however, have a wide range, which indicates that there are no fixed delivery times per class but rather a range of delivery times for a class.





5.2. Second hypotheses

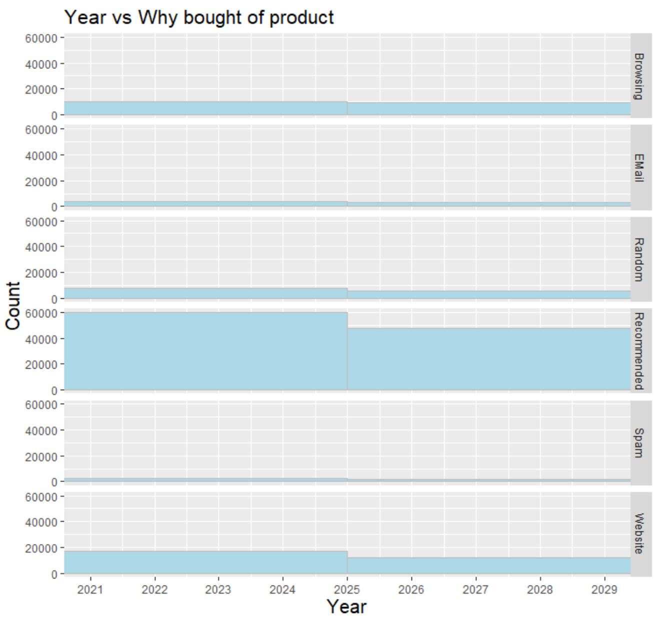
H0: The day,month and year of a sale **does not depend** on the reason for buying the item.

H1: At least the day,month or year **depends** on the reason for buying an item.

A p value of 0.05 is chosen to do the MANOVA test,since 0.05 is the most popular and universal p value. The test is not significant as the p-value of 2.2e-16 <0.05 ,therefore, the **H0 hypothesis is rejected** and it is clear that at least one of the features has an influence on the buying pattern for the reason why the product is bought.

```
               Df    Pillai approx F num Df den Df    Pr(>F)
Why.Bought      5 0.0023241   27.907     15 539916 < 2.2e-16 ***
Residuals  179972
---
```

Dependant Variable	P-value	Analyses
Day	0.5495	P> 0.05. Day has no influence on why a product is bought.
Month	0.7884	P> 0.05. Month has no influence on the reason why a product is bought.
Year	2.2e-16	P< 0.05. Year differs depending on the reason the product is bought.



The class for the products has no effect on the day and month a sale takes place. However, the year influences when a product is bought,as seen in the graph. The reason why a customer bought a product decreased in 2025. This is something that the company should look into, as the sales reduced and so also the revenues received from it. They will have to increase advertising or improve quality so that they remain competitive advantage.

5.3. Third hypotheses

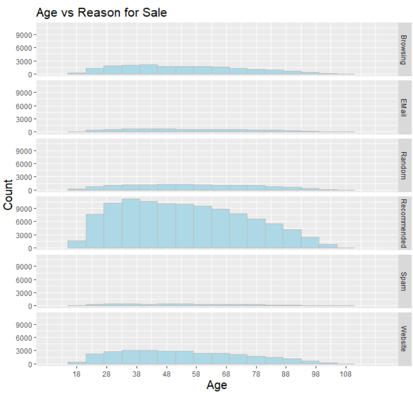
H0: The price, delivery time and age of a sale **does not depend** on the reason for buying the item.

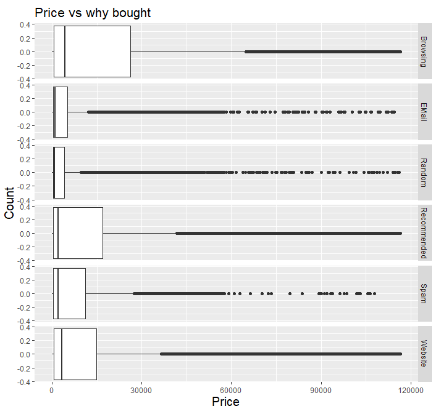
H1: At least the price, delivery time or age **depends** on the reason for buying an item.

A p value of 0.05 is chosen to do the MANOVA test,since 0.05 is the most popular and universal p value. The test is not significant as the p-value of 2.2e-16 <0.05 ,therefore, the **H0 hypothesis is rejected** and it is clear that at least one of the features has an influence on the buying pattern for the reason why the product is bought.

```
                Df    Sum Sq    Mean Sq F value    Pr(>F)    
Why.Bought       5 1.5742e+12 3.1484e+11  736.26 < 2.2e-16 ***
Residuals   179972 7.6960e+13 4.2762e+08                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Dependant Variable	P-value	Analyses
Delivery time	2.2e-16	P<0.05. Delivery time differs depending on the reason the product is bought.
Price	2.2e-16	P<0.05.Price differs depending on the reason the product is bought.
Age	2.2e-16	P< 0.05. Age differs depending on the reason the product is bought.





The age is almost normally distributed among the reasons for sales. The delivery times are longer if a product is bought due to browsing or websites, this can possibly be due to a lack in IT-systems or outdated/slow technology. The price is the most expensive if the product is bought because of browsing. (Lani, 2013)

6. Reliability of service & products

6.1. Taguchi Loss

PROBLEM 6

"A blueprint specification for the thickness of a refrigerator part at Cool Food, Inc. is 0.06 +- 0.04 centimeters (cm). It costs $45 to scrap a part that is outside the specifications. Determine the Taguchi loss function for this situation."

$Loss(x) \ = \ k(x \ - \ T)^2$

$Loss(x)$ = 45

T =0.06

$(x \ - \ T)$=0.04

$k \ = \ 45/0.04^2 \ = \ 28\,125$

Thus Taguchi Loss function is ⟶ $Loss(x) \ = \ 28125(x \ - \ T)^2$

PROBLEM 7

"A team was formed to study the refrigerator part at Cool Food, Inc. described in Problem 6. While continuing to work to find the root cause of scrap, they found a way to reduce the scrap cost to $35 per part.

a. Determine the Taguchi loss function for this situation.

b. If the process deviation from the target can be reduced to 0.027 cm, what is the Taguchi loss?"

a.) $Loss(x) \ = \ k(x \ - \ T)^2$

$Loss(x)$ = 35

T =0.06

$(x \ - \ T)$=0.04

$k \ = \ 35/0.04^2 \ = \ 21875$

Thus Taguchi Loss function is ⟶ $Loss(x) \ = \ 21875(x \ - \ T)^2$

b.)$Loss(0.027) \ = \ 21875(x \ - \ T)^2$

$Loss(0.027) \ = \ 21875(0.027 \ - \ 0.06)^2$

= $23.82

6.2. Reliability of machines

PROBLEM 27

"Machine Reliability: A - 0.85 |B -0.92 |C -0.90

…a. Analyze the system reliability, assuming only one machine at each stage (all the backup machines are out of operation).

b. How much is the reliability improved by having two machines at each stage?"

a.) $Series\ reliability\ =\ Ra\ x\ Rb\ x\ ...\ Rn$

$System\ reliability\ for\ series\ =\ 0.85\ x\ 0.92\ x\ 0.9$

$Reliability\ =\ 70.38\%$

b.) $Parallel\ reliability\ =\sum_{i=A}^{n}\ (1\ -\ P(not\ breaking\ parallel))_i$

$System\ reliability\ for\ parallel\ =\ (1\ -\ (1\ -\ 0.85)^2)(1\ -\ (1\ -\ 0.92)^2)(1\ -\ (1\ -\ 0.9)^2)$

$Reliability\ =\ 96.15\%$

Running two identical machines simultaneously would increase the reliability with 25.77316%. Therefore, it would be advised to choose the parallel option, as the system will be both more reliable and it will reduce breakdowns and waiting time.

6.3. Expecting reliable delivery

6.3.1. Reliable deliveries

The probability of having reliable delivery times will be dependent on both the drivers and vehicles, as the drivers will have to drive the cars, and therefore, the number of vehicles available will force the same number of drivers (or more) to be available.

$P(reliable)\ =\ P(20\ or\ more\ cars\ available)\ *\ P(20\ or\ more\ drivers)$

<u>Reliable vehicles-</u>

Using the binomial equation, a weighted probability was found of having reliable vehicles:

$P(X\ <\ 1)\ =\ [21C0]\ *\ (p^0(1\ -\ p)^{*21-0})\ *\ 1560\ =\ 1344$ {Reliability of having 21 cars}

$P(X\ <\ 2)\ =\ [21C1]\ *\ (p^1(1\ -\ p)^{*21-1})\ *\ 1560\ =\ 190$ {Reliability of having 20 cars}

$P(X\ <\ 3)\ =\ [21C2]\ *\ (p^2(1\ -\ p)^{*21-2})\ *\ 1560\ =\ 22$ {Reliability of having 19 cars}

$P(X\ <\ 4)\ =\ [21C3]\ *\ (p^3(1\ -\ p)^{*21-3})\ *\ 1560\ =\ 3$ {Reliability of having 18 cars}

$P(X < 5) = [21C4] * (p^4(1 - p)^{*21-4}) * 1560 = 1$ {Reliability of having 17 cars}

$Weighted\ average = 1/1560 \sum_{i=i}^{5} [P(X < x) * (days\ x\ cars\ available)] = 0.007060816$

The number of total days having reliable vehicles is then:

$P(X < 1) = [21C0] * (0.007^0(1 - 0.007)^{*21-0}) * 1560$ {Days having 21 cars}

$P(X < 2) = [21C1] * (0.007^1(1 - 0.007)^{*21-1}) * 1560$ {Days having 20 cars}

Expected reliable delivery days for vehicles in a year is then:

$((P(X < 1) + P(X < 2))/1560) * 365$ = **361.5046 days**

Reliable drivers-

Using the binomial equation, a weighted probability was found of having reliable drivers:

$P(X < 1) = [21C0] * (p^0(1 - p)^{*21-0}) * 1560 = 1458$ {Reliability of having 21 drivers}

$P(X < 2) = [21C1] * (p^1(1 - p)^{*21-1}) * 1560 = 95$ {Reliability of having 20 drivers}

$P(X < 3) = [21C2] * (p^2(1 - p)^{*21-2}) * 1560 = 6$ {Reliability of having 19 drivers}

$P(X < 4) = [21C3] * (p^3(1 - p)^{*21-3}) * 1560 = 1$ {Reliability of having 18 drivers}

$Weighted\ average = 1/1560 \sum_{i=i}^{5} [P(X < x) * (days\ x\ drivers\ available)] = 0.003223914$

The number of total days having reliable drivers is then:

$P(X < 1) = [21C0] * (0.003^0(1 - 0.003)^{*21-0}) * 1560$ {Days having 21 drivers}

$P(X < 2) = [21C1] * (0.003^1(1 - 0.003)^{*21-1}) * 1560$ {Days having 20 drivers}

Expected reliable delivery days for drivers in a year is then:

$((P(X < 1) + P(X < 2))/1560) * 365$ = **364.2352 days**

Total reliable days in a year due to enough drivers and vehicles:

$P(total) = P(drivers) * P(vehicles) = 0.003 * 0.007 = 0.9883482$

0.9883482*365 = **360.7471 days**

This means out of 365 days in a year, the company can expect 361 days of reliable deliveries.

6.3.2. Increase number of vehicle by 1 to 22

If we increase the vehicle by one until 22 and if there needs to be 20 or more vehicles in order to be reliable, the new probability for reliable delivery would be-

$P(22\ cars)\ +\ P(21\ cars)\ +\ P(20\ cars)$ and by using the weighted average obtained above for vehicles the probability of each is calculated by:

$P(X < 1)\ =\ [22C0]\ *\ (0.007^{0}(1\ -\ 0.007)*^{21-0})$=0.8556542 {Probability of having 22 cars}

$P(X < 2)\ =\ [22C1]\ *\ (0.007^{1}(1\ -\ 0.007)*^{21-1})$ =0.1338607 {Probability of having 21 cars}

$P(X < 3)\ =\ [22C2]\ *\ (0.007^{2}(1\ -\ 0.007)*^{21-2})$=0.009994816 {Probability of having 20 cars}

New delivery reliability:

= 0.8556542+0.1338607 +0.009994816 = 0.9995098

New expected number of reliable delivery days in a year

=0.9995098*365 = **364.8211 days**

This means there will be 365 reliable days in the year. Increasing the total vehicles from 21 to 22 improves the probability of having more reliable cars from 362 days to 365 days (3 days).

Conclusion

After analysis of the sales data multiple conclusions can be drawn. The delivery times of luxury items are of great importance to their customers, however, this has a negative impact on the available delivery capacity of other items. The company should decide between a trade off between making higher revenue due to more luxury item sales, which are priced higher, or losing potential customers (such as for household/ gift customers) due to waiting longer for their products and losing the revenue that would have been received for these items. Gifts are one of the most sold items for this company, therefore, they would have to determine how long these customers are willing to wait before they find another customer and if putting luxurious items as their top delivery priority is worth the loss of customers.

The main form of income will result from sales of technology, as this is the second highest sold item and it has the second highest average price range. This is also a product with an increasing demand rate due to modernization in workplaces and in households, therefore, the company should focus on continuously improving their marketing and production for this specific class of items. The company can then also opt for a more reliable delivery system or negotiate with more efficient logistic companies, since a reduction in the delivery times from +-20 hours to 3 hours will minimize the cost associated with this class and in turn increase the profit even more.

References

X-bar and s Charts. (n.d.). [online] Available at: https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/X-bar_and_s_Charts.pdf.

Dirkse van Schalkwyk, T. (2022b). *QA344 Statistics.PDF*. *Sunlearn*, Stellenbosch, pp.1–28.

www.itl.nist.gov. (n.d.). *3.4.6. Assessing Process Capability*. [online] Available at: https://www.itl.nist.gov/div898/handbook/ppc/section4/ppc46.htm [Accessed 18 Oct. 2022].

Lani, J. (2013). *ANOVAs and MANOVAs*. [online] Statistics Solutions. Available at: https://www.statisticssolutions.com/anovas-and-manovas/ [Accessed 18 Oct. 2022].