Quality Assurance 344

ECSA Project

Preamble to the Engineering Counsel of South Africa

(ECSA) report that proves graduate attribute 4

(ECSA GA4).

Author: 21205353 (Student number)

Jean – Pierre Higgo

2022

ABSTRACT

This report contains comments and findings about programming in R which has to do with Statistical Process Control (SPC). In the report there will be a brief discussion about how the given data set was wrangled into usable format. Statistical methods have been used to identify the critical trends in the data set. An application of Statistical Process control has been made in the form of X-Chart and S-Charts. The results of these graphs were displayed in the form of a table. Further analysis of the data has been done into the future to compare with the initial values. Techniques has been applied to ensure certain processes stays within certain acceptable quality boundaries.

ACKNOWLEDGEMENTS

The author thanks Dr. Theunis Dirkse van Schalkwyk for his guidance. Without the knowledge gained from studying Quality Assurance 344 this report would not have been possible.

The later editions include the format of a formal report which was contributed by the Language Centre at Stellenbosch University.

TABLE OF CONTENTS

List of figures

List of tables

INTRODUCTION

Through studying quality assurance, we were introduced to the pioneers of quality. Dr Walter Shewhart developed the Plan, Do, Check, Act also known as the (PDCA) cycle, he also had theories about process control. In this project we made use of statistical process control which was mainly developed by Shewhart. Deming said the following: “If I could reduce my message about quality to just a few words, I’d say it all has to do with reducing variation.” Variation can be introduced in so many ways examples of this would be temperature differences, different skills of workers, tool wear and poor quality of raw materials. Our main focus in this data set was the delivery time of goods and how these variable changes over time. The objective of this report is to implement the theoretical knowledge we have obtained about Statistical Process Control on a practical real-life data set. This data set is unique in the way that it forces us to analyse real time data. Thus, the process needs to be controlled as time varies. The delivery times needs to remain in check for items to reach their destinations in time. If these values fall outside certain criteria for example one or two standard deviations from the central line, then changes need to made to the process in order to control the delivery times. These changes would ensure quality standards are kept in check. The quality of a product can be seen a competitive advantage which would make one business stand out from its competitors.

1. Data Wrangling

The data set which was provided in csv format had to be imported into R programming language. This was done using the read function in R. Upon inspection of the data set it became clear that there were several data entries containing NA value which has to be discarded. We were also given instruction to separate the data into two categories namely valid data and incomplete data. This was performed by sub setting data with NA values into a variable known as Sale Table Incomplete. In order to obtain the valid data an operation was executed to subtract the incomplete instances from the complete or valid data. This will be done in a way which will still be applicable if there were any changes to the data set, regardless of the value of N, n or m. It is important to keep the index of the original data set intact. These processes have been done using basic R code, Tidyverse was not necessary. In this particular dataset, we have 180000 instances, in this case only 179978 instances are valid. Meaning that 17 instances, excluding the negative prices belong to the invalid dataset.

2. Descriptive statistics

Through the creation of a boxplot which shows both price and class, it becomes evident that the classes with the highest prices are luxury, technology and households. Gifts are also relatively large compared to the price of the other classes.

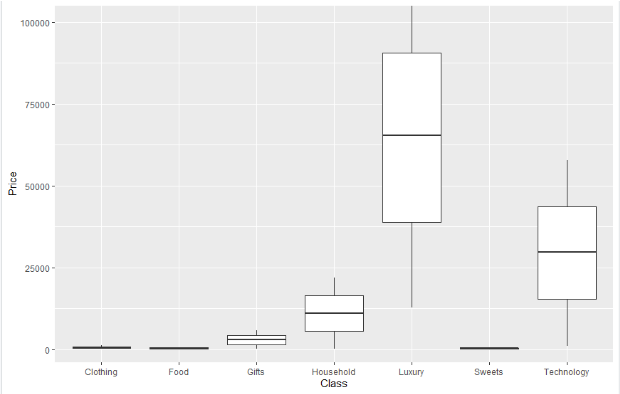

Figure 1 Boxplot of different classes

If we plot the Price vs the count of purchases, we see that lower prices have much larger amounts of sales, expensive items have significantly less item count.

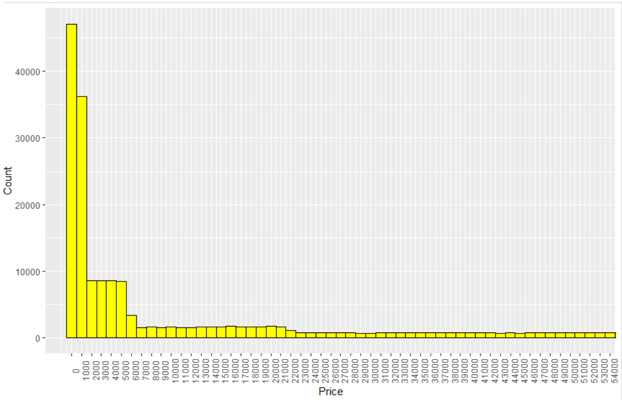

Figure 2 Price vs Count Graph

When plotting the age vs count graph, we see that the age closely resembles a skewed right distribution.

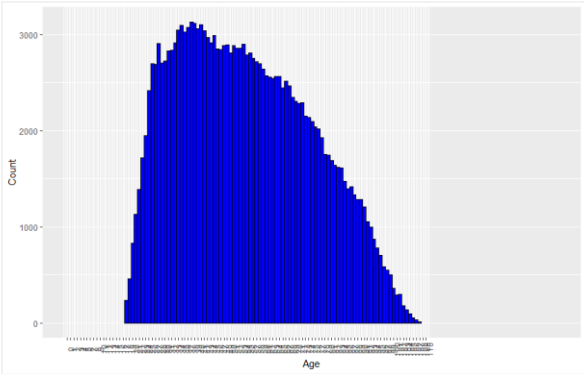

Figure 3 Age vs Count

The main reason for purchase is because of recommendations, in the case of a business the sales department would be able to make use of this information. This means that high emphasis needs to be placed on the quality of products and aspects such as customer service. Customers would only recommend their products if they are satisfied with the given products. Websites is second largest reason for purchase, which means

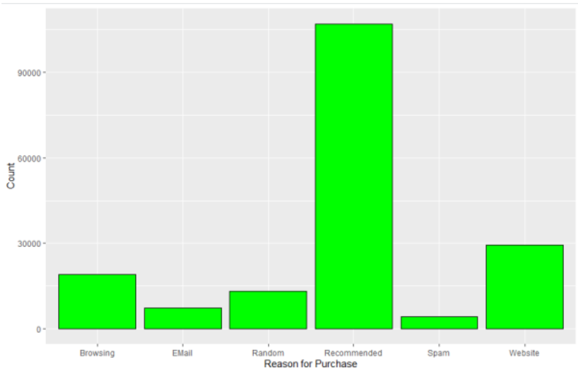

Figure 4 Reason for purchase

that the sales team can invest time and effort in improving the website or budget money for the given website.

The class which is occurs the most often in the data set would be the gift class, closely followed by technology. Even though a class may be popular doesn't mean that that class would bring in the most revenue. For example, sweets are relatively cheap in comparison with the technology class. So, no assumption regarding finances needs to be made from this graph.

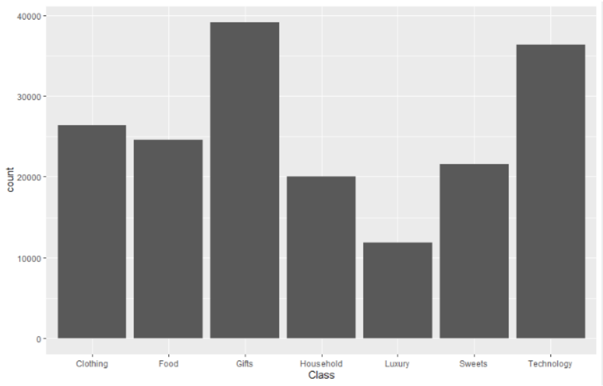

Figure 5 Classes occurrences

From the graph below we can see a distribution which can be classified as bimodal. This is only for one class. Which means the graph for multiple classes would be a multimodal graph.

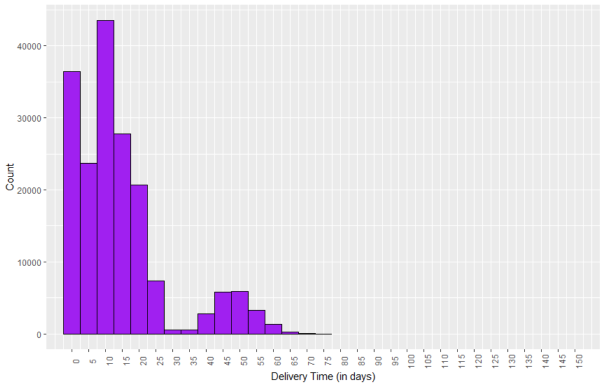

Figure 6 Delivery Time

Process Capability

We are asked to calculate the technology class process capability indices namely C_P, Cpu, Cpl and Cpk for the process delivery times of technology class items. We also have been asked to explain why an LSL of 0 would be logic. Any process that is executed isn't an exact science there will be some sort of variation within that process which can be viewed as a normal distribution. The normal distribution is in the shape of a bell-curve. The normal distribution has two cut off limits namely the LSL and the USL, any item or part outside the normal distribution will be reject. Items below the LCL will also be rejected.

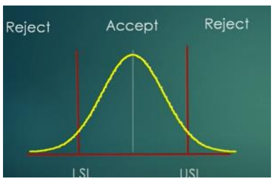

Figure 7 Accept and reject region of a normal distribution

Process Capability Ratio (C_P) is the specification width divided by the process width. A C_P value of one would indicate that the process will fall between +3 sigma and -3 sigma. If C_P is less than one it indicates that the process falls outside the specification limits. A C_P value of greater than one means that the process is operating well within the specification limits. The process is thus in control and does not touch the specification limits. A process which falls within the boundaries with the narrowest width would be the more capable process. Process capability index (C_{pk}) in the case where the process mean shifts it cause some of the parts or objects to fall into a rejection region. C_P can't calculate the shift in the graph and that is why C_{pk} is used. C_{pk} is the following equation $\min\left(\frac{USL-\mu}{3\sigma};\frac{\mu-LCL}{3\sigma}\right)$. C_P is always greater than C_{pk}. If the process is centred the C_P will equal C pk. If C_{pk} equal zero it means the process is overlapping with one of the specification limits. If the C_{pk} is less than 0 the process mean has gone beyond the specification limit. If C_{pk} is between 0 and 1 the process mean is some distance away from the specification limit. If C_{pk} is greater than one the process is well within the limits. A C_{pk} value of 1.33 would be considered good.

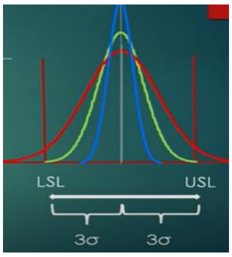

Figure 8 Different values of C_P

2.1 Data Observations

Values obtained from the given result $C_P = 1.142$ $C_{pk} = 0.38$ $C_{pl} = 1.905$ $C_{pu} = 0.38$ A C_P of 1.142 means that the process is operating well within the specification limits. A C_{pk} of 0.38 would indicate that the mean of the process has shifted from the centre. The process does not meet the requirements, but this is caused by shifting of the mean. The corrective action would be to try and move the process mean so that better C_{pk} can be achieved. C_{pl} and C_{pu} is the one-sided process potential, the C_{pu} is a smaller value which indicate that the process mean is closer to the upper specification limit.

Explain why a LSL of 0 is logical:

The variable in question is the delivery time of products. Negative values for the LSL would not be possible since time can't be negative. The delivery time will be a value between 0 and the required time to deliver the product. The lower control limit does need to be zero since this is the relative measure which values are compared to each other. Thus, LCL equal to 0 is logical.

3.1 Statistical process control

X-bar chart for delivery times [1:30] (X-bar with s)

	myClassNames	UCLXT	U2SIGMAT	U1SIGMAT	CLXT	L1SIGMAT	L2SIGMAT	LCLXT
1	Clothing	9.40493352386633	9.25995568257756	9.11497784128878	8.97	8.82502215871122	8.68004431742245	8.53506647613367
2	Household	50.2483278659662	49.0196259847182	47.7909241034702	46.5622222222222	45.3335203409742	44.1048184597263	42.8761165784783
3	Food	2.70580826490218	2.63414037583517	2.56247248676816	2.49080459770115	2.41913670863414	2.34746881956713	2.27580093050012
4	Technology	22.9746158797126	22.1078920679566	21.2411682562005	20.3744444444444	19.5077206326884	18.6409968209323	17.7742730091763
5	Sweets	2.89310560598847	2.75632327755553	2.61954094912259	2.48275862068966	2.34597629225672	2.20919396382378	2.07241163539084
6	Gifts	9.48856467334077	9.11274681926422	8.73692896518766	8.36111111111111	7.98529325703456	7.60947540295801	7.23365754888145
7	Luxury	5.49396512637278	5.24116193610037	4.98835874582796	4.73555555555556	4.48275236528315	4.22994917501074	3.97714598473833

Figure 9 X-Bar chart [1:30]

S-bar chart for delivery times [1:30]

	myClassNames	UCLXT	U2SIGMAT	U1SIGMAT	CLXT	L1SIGMAT	L2SIGMAT	LCLXT
1	Clothing	9.40493352386633	9.25995568257756	9.11497784128878	8.97	8.82502215871122	8.68004431742245	8.53506647613367
2	Household	50.2483278659662	49.0196259847182	47.7909241034702	46.5622222222222	45.3335203409742	44.1048184597263	42.8761165784783
3	Food	2.70580826490218	2.63414037583517	2.56247248676816	2.49080459770115	2.41913670863414	2.34746881956713	2.27580093050012
4	Technology	22.9746158797126	22.1078920679566	21.2411682562005	20.3744444444444	19.5077206326884	18.6409968209323	17.7742730091763
5	Sweets	2.89310560598847	2.75632327755553	2.61954094912259	2.48275862068966	2.34597629225672	2.20919396382378	2.07241163539084
6	Gifts	9.48856467334077	9.11274681926422	8.73692896518766	8.36111111111111	7.98529325703456	7.60947540295801	7.23365754888145
7	Luxury	5.49396512637278	5.24116193610037	4.98835874582796	4.73555555555556	4.48275236528315	4.22994917501074	3.97714598473833

Figure 10 S-Bar chart [1:30]

X-bar chart for delivery times [31: end of samples] (X-bar with s)

	myClassNames	UCLXT	U2SIGMAT	U1SIGMAT	CLXT	L1SIGMAT	L2SIGMAT	LCLXT
1	Clothing	9.47982749384551	9.31995568544138	9.16008387703725	9.00021206863312	8.84034026022899	8.68046845182486	8.52059664342073
2	Household	52.9375067890794	51.5465551508935	50.1556035127077	48.7646518745218	47.3737002363359	45.9827485981501	44.5917969599642
3	Food	2.72224676265653	2.64890340768498	2.57556005271342	2.50221669774187	2.42887334277031	2.35552998779876	2.28218663282721
4	Technology	22.7245787592099	21.8185077691756	20.9124367791414	20.0063657891071	19.1002947990729	18.1942238090386	17.2881528190043
5	Sweets	2.90565388433046	2.77095712639325	2.63626036845604	2.50156361051883	2.36686685258163	2.23217009464442	2.09747333670721
6	Gifts	14.1043648599277	13.7171229952403	13.3298811305528	12.9426392658653	12.5553974011779	12.1681555364904	11.7809136718029
7	Luxury	4.55630707843382	4.35124808295062	4.14618908746743	3.94113009198423	3.73607109650104	3.53101210101784	3.32595310553465

Figure 11 X-Bar chart [31:end of samples]

S-bar chart for delivery times [31: end of samples]

	myClassNames	UCLS	U2SIGMAS	U1SIGMAS	CLS	L1SIGMAS	L2SIGMAS	LCLS
1	Clothing	0.955583584833818	0.839681589387477	0.723779593941136	0.607877598494795	0.491975603048454	0.376073607602113	0.260171612155772
2	Household	8.31397709212241	7.30558122768347	6.29718536324454	5.2887894988056	4.28039363436667	3.27199776992773	2.2636019054888
3	Food	0.438386897396513	0.385215288976071	0.332043680555629	0.278872072135187	0.225700463714744	0.172528855294302	0.11935724687386
4	Technology	5.41575511914017	4.75888236083903	4.10200960253789	3.44513684423675	2.78826408593561	2.13139132763447	1.47451856933333
5	Sweets	0.805107617784373	0.707456736238007	0.609805854691641	0.512154973145275	0.41450409159891	0.316853210052544	0.219202328506178
6	Gifts	2.31461677296082	2.03387869108347	1.75314060920612	1.47240252732877	1.19166444545142	0.910926363574064	0.630188281696713
7	Luxury	1.22567582091096	1.07701454661896	0.928353272326961	0.779691998034961	0.631030723742962	0.482369449450963	0.333708175158964

Figure 12 S-Bar [31: end of samples]

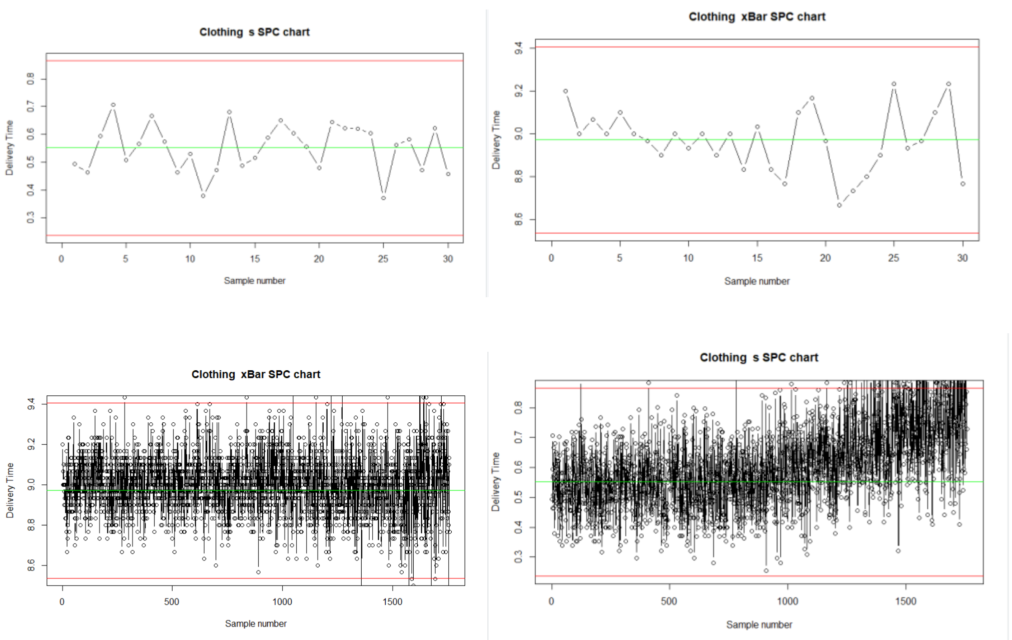

Figure 13 All graphs of clothing

When calculating the graphs, it is important to remove the outlier from the data in the first 30 samples to ensure that your limits are right. Since all of the following questions make use of these values in the calculations. The sweets and food class has certain outliers which makes it difficult to compute these values. Thus, these limits needs to be recalculated before using them again.

4.1. A. Optimizing the delivery process

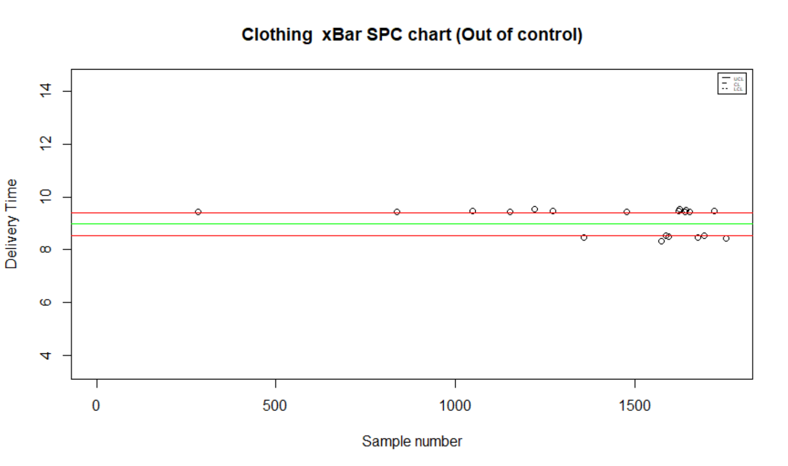

Figure 14 Clothing X-Bar Out of Control

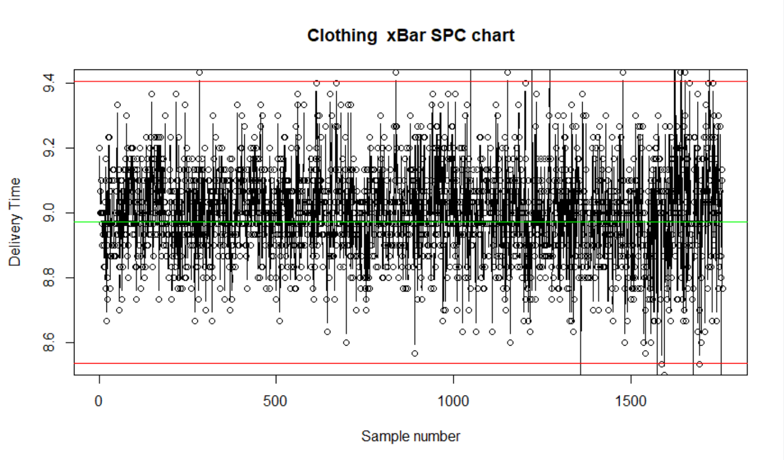

Figure 15 Clothing graph (All values)

Total found	FIRST	SECOND	THIRD	LAST 3	LAST 2	LAST
20	282	837	1048	1695	1723	1756

Table 1 Results

Points beyond the upper and lower control limit. From the figure above 20 samples were identified which fell beyond the control limits. There are normally a special cause present in such case. Employees need to find the origin of the problem and attempt to remove this issue. Only a small number of samples were identified, which means it might be difficult to determine the cause of the problem. This could be for example that truck delivering might be stuck in traffic. This class is in statistical process control to a certain extent. Control samples needs to be investigated to find the cause of outliers.

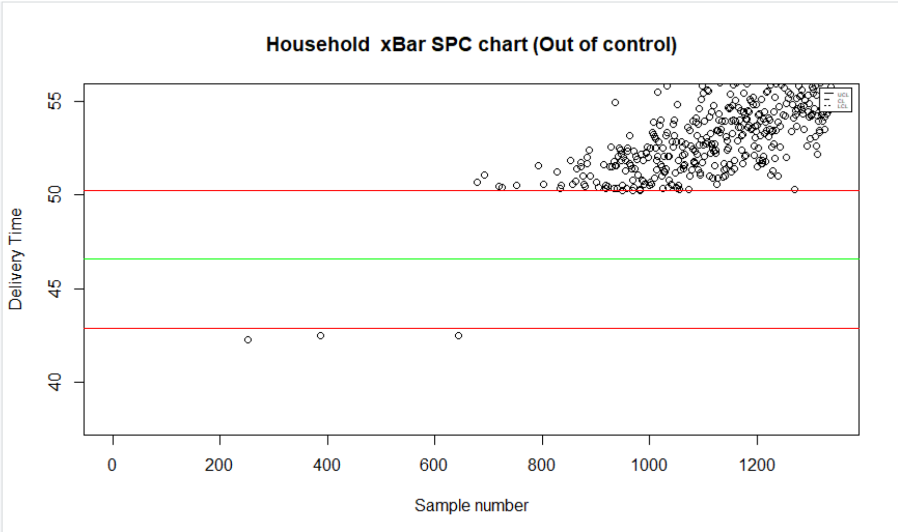

Figure 16 Household X-Bar (Out of control)

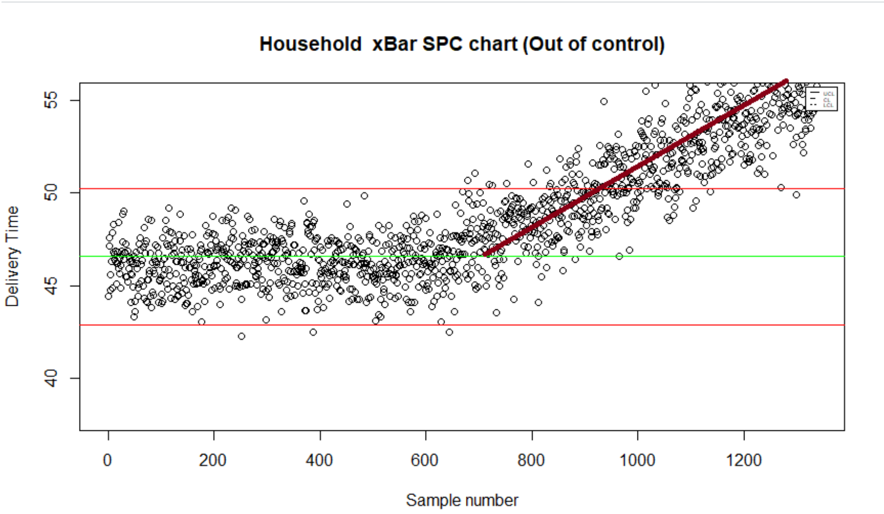

Figure 17 Household X-Bar (All values)

Total found	FIRST	SECOND	THIRD	LAST 3	LAST 2	LAST
395	252	387	643	1335	1336	1337

Table 2 Household Results

The household class is not in statistical control. A whopping 395 delivery times fell beyond the upper and lower control limit. Here employees that must interpret the graph would see that more than 7 points in a row trend upwards. This rule clearly would force employees to investigate the problem. Major issues will be identified. Examples of this would be that large items takes more handling time. The process was in control for the first 600 sample, where after standards have been lowered on a gradual rate. Control samples needs to be investigated to find the cause of outliers.

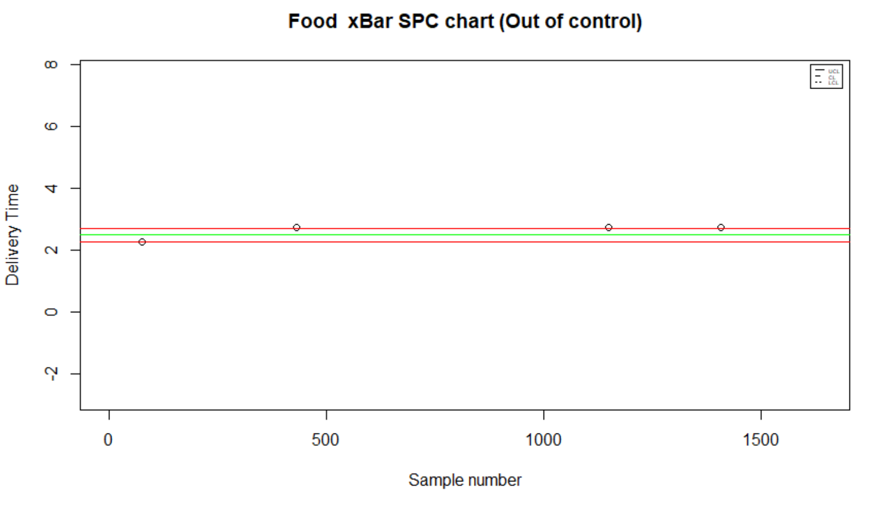

Figure 18 Food X-Bar Out of Control

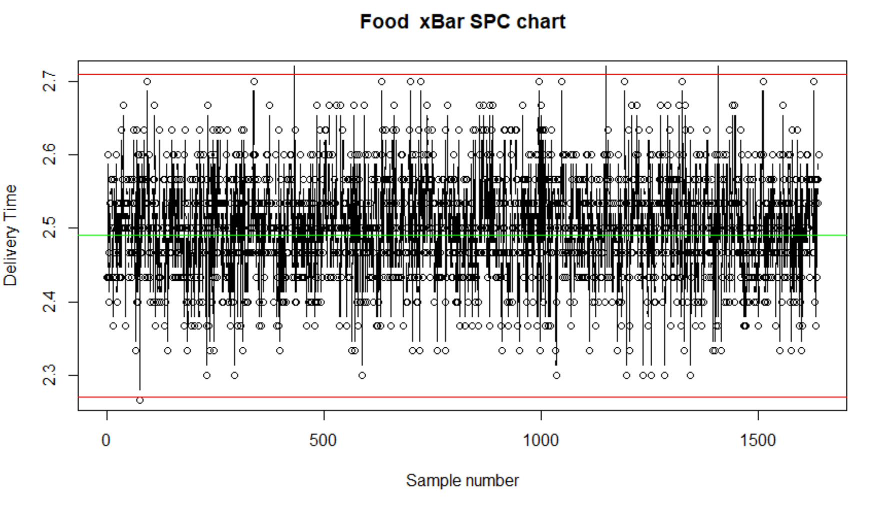

Figure 19 Food X-Bar (All Values)

Total found	FIRST	SECOND	THIRD	LAST 3	LAST 2	LAST
4	75	432	1149	NA	NA	1408

Table 3 Food Results

The food class have points beyond and below the upper and lower control limits. The rest of the samples fall within the given limits. A large number of samples fall close to the limits. Which means that the variability within the system is quite high. The out-of-control samples that were identified are spread far away from each other. Steps need to be taken to eliminate the possible cause of the problem. A reduction in variability in would stabilize the whole process. Control samples needs to be investigated to find the cause of outliers.

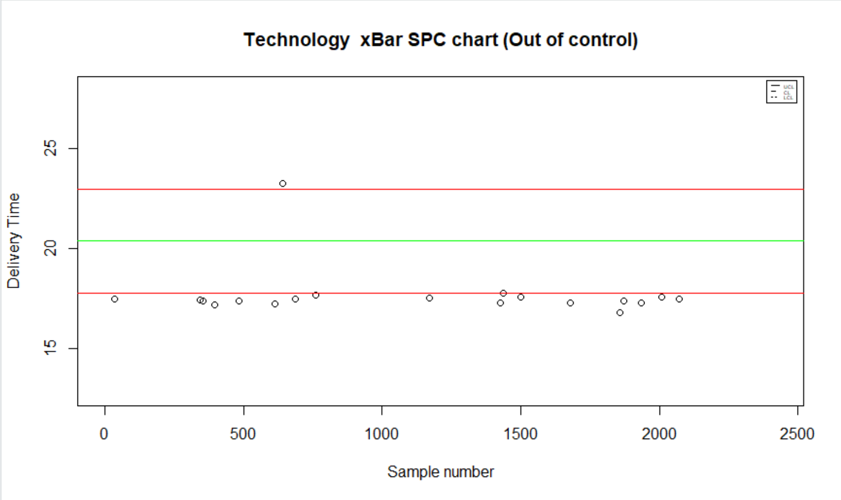

Figure 20 Technology X-Bar (Out of Control)

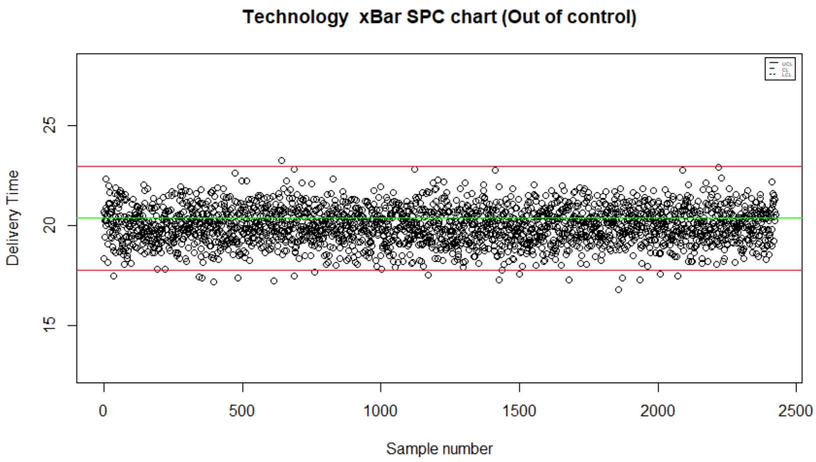

Figure 21 Technology X-Bar (All values)

Total found	FIRST	SECOND	THIRD	LAST 3	LAST 2	LAST
19	37	345	353	1933	2009	2071

Table 4: Technology Results

The technology class have 19 sample beyond the upper and lower control limits. Here these points are more than 3 standard deviations away from the mean. There would be a special cause behind the shifts or outliers in the graph. The further the points are from the limits the stronger the cause for special cause variation. In this case some points are well beyond the limits. A special cause has to be identified. An example would be that technology items are fragile and takes longer to transport. The rest of the sample values actually lie close to the mean.

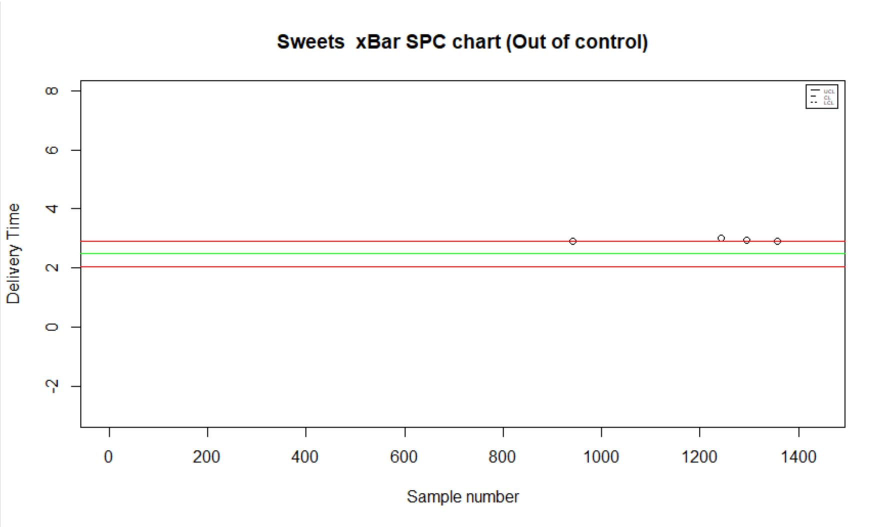

Figure 22 Sweets X-Bar (Out of control)

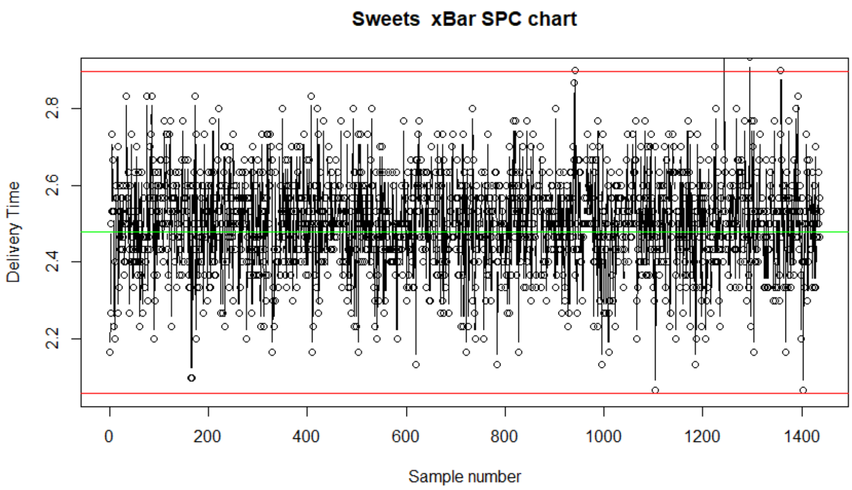

Figure 23 Sweets X-Bar (All values)

Total found	FIRST	SECOND	THIRD	LAST 3	LAST 2	LAST
4	942	1243	1294	NA	NA	1358

Table 5 Sweets Results

Sweets class is in statistical process control, for almost 900 samples. There after points are found that fall beyond the control limits. Which mean the process is not in statistical control. The values that were identified fall close to the limits, which does not give a strong cause for special cause variation. There is a possibility that there actually might be a reason behind these values that fall outside the limits. It would be difficult to identify the problem, but there must be a reason for these values shown in the graphs. Control samples needs to be investigated to find the cause of outliers.

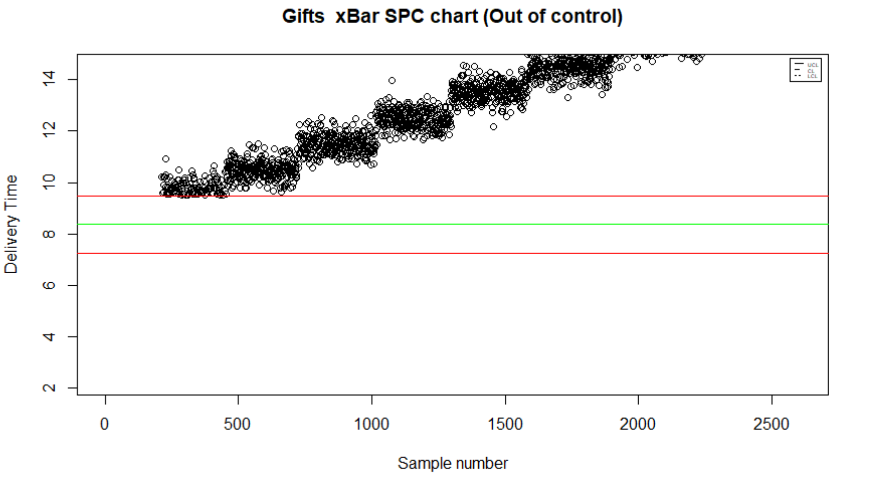

Figure 24 Gifts X-Bar (Out of control)

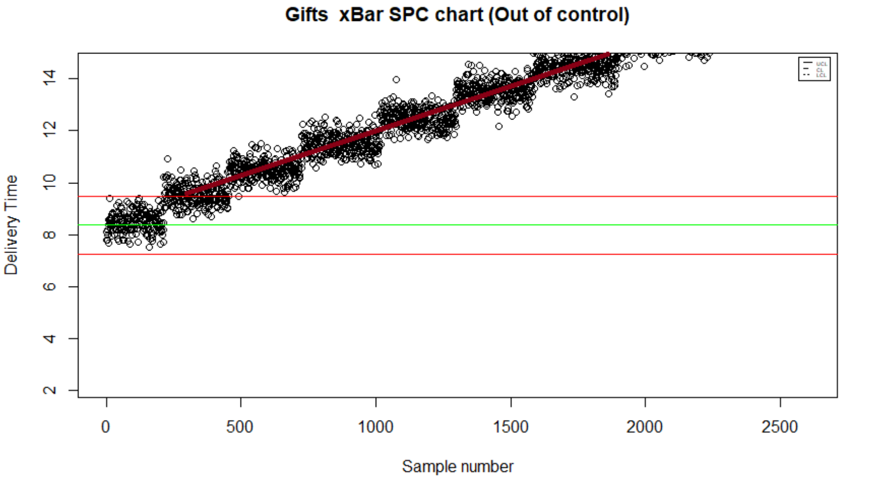

Figure 25 Gifts X-Bar (All values)

Total found	FIRST	SECOND	THIRD	LAST 3	LAST 2	LAST
2287	213	216	218	2607	2608	2609

Table 6 Gifts Results

Gifts is not in statistical process control, there is an obvious increasing trend. This trend extends beyond the upper control limit. Beyond the standard 3 sigma limit. It keeps on increasing gradually. This class needs urgent attention to ensure that the company does not lose customers, due to poor customer satisfaction. The data seems to be grouped in small groups, the mean of each groups increasing every few hundred samples.

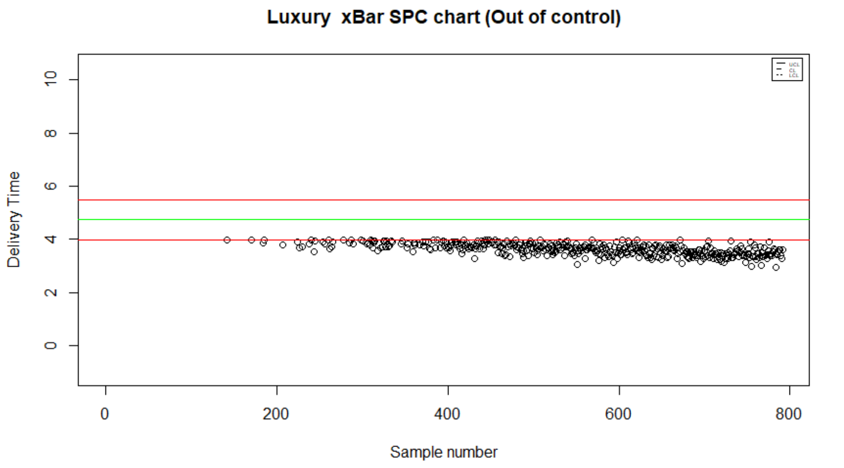

Figure 26 Luxury X-Bar (Out of control)

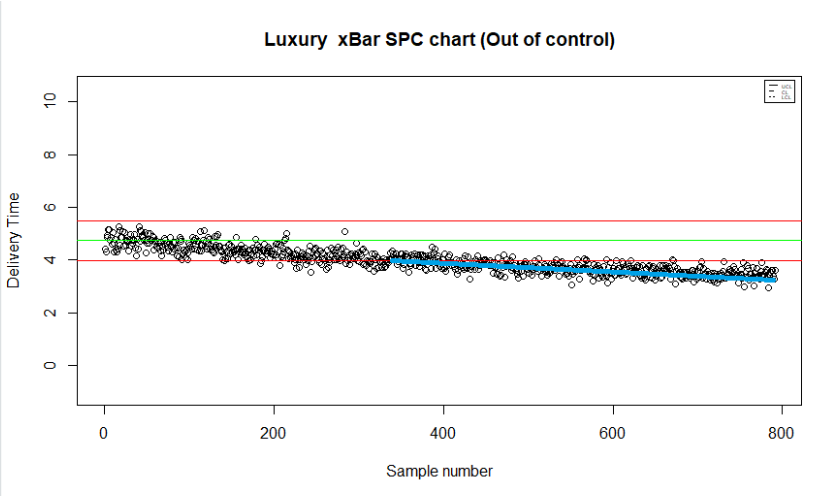

Figure 27Luxury X-Bar (All Values)

Total found	FIRST	SECOND	THIRD	LAST 3	LAST 2	LAST
440	142	171	184	789	790	791

Table 7Luxury Results

In the luxury class we can clearly see a decreasing tend, thus the process is not in statistical process control. This means that delivery times seems to decrease, this is not necessarily a bad thing. A clear rule can be identified. Here 7 continuous decreasing samples will be identified by the trained employees. In this specific case it is actually beneficial to see delivery times decreasing. A reason for this decrease could be that new transportation techniques have been used to speed up the delivery times. The mean of the process must also be lowered. Control samples needs to be investigated to find the cause of outliers.

	V1	V2	V3	V4	V5	V6	V7	V8
1	Class	Total found	First	Second	Third	Last3	Last2	Last1
2	Clothing	20	282	837	1048	1695	1723	1756
3	Household	395	252	387	643	1335	1336	1337
4	Food	4	75	432	1149	432	1149	1408
5	Technology	19	37	345	353	1933	2009	2071
6	Sweets	4	942	1243	1294	1243	1294	1358
7	Gifts	2287	213	216	218	2607	2608	2609
8	Luxury	440	142	171	184	789	790	791

Figure 28 Total Values beyond limits

4.1 B. Consecutive samples of "sample standard deviations "

Find the most consecutive samples of "s-bar or sample standard deviations" between -0.3 and +0.4 sigma-control limits and the ending sample number (last sample to be in the given range)

Class	**Most consecutive samples**	**Ending Sample Number**
Clothing	4	**223**,1121
Household	3	**45**,588,647,766,843
Food	4	85,223,640,879
Technology	6	**1776**
Sweets	5	**316**
Gifts	7	**2477**
Luxury	4	**63**

Table 8 Most Consecutive Samples

	V1	V2	V3
1	Class	Most Consecutive	Last element
2	Clothing	4	1121
3	Household	3	843
4	Food	4	879
5	Technology	6	1776
6	Sweets	5	316
7	Gifts	7	2477
8	Luxury	4	63

Figure 29 Most consecutive elements

In some cases, more than one occurrence of consecutive values has been identified. In the case of clothing 3 instances were identified with 4 consecutive samples. In the household class there were 9 instances of 3 consecutive samples. In the food class a rare case of only one occurrence happened with 5 consecutive samples. The technology class also had only one occurrence of 6 consecutive samples. The sweets class had 3 occurrences with 4 consecutive samples. The gift class only had one occurrence with 7 consecutive samples. The luxury class had one occurrence of 4 consecutive samples. A trend can be identified in that, when the consecutive amount is high there are a lower number of occurrences. Values might differ a bit due to small differences in the 0.4 and -0.3 sigma limits.

4.2 Type 1 (Manufacturer's Error)

This means rejecting the null hypothesis when it is actually true. This also known as the alpha value.

```
##4.2
#4.2
'#The probability of making a type 1 error
in a general sense UCL and LCL cover 99.74% of the data
its P(Xbar < LCL) + P(xbar  >UCL) for Xbar chart
its P(Xbar < LCL) + P(xbar  >UCL) for S chart values
And then that becomes
P(Z<. -3)+p(z>3) or
this 0.00135*2 or 0.27%'
pnorm(-3)*2 #0.27%
```

Figure 30 R Code for 4.2

As calculated in R the probability of making a type 1 error would be the probability of (XBar< LCL) + probability of (XBar > UCL), thus P(z<-3) + P(z>3) = 0.00135*2=0.0027

Yielding an answer of **0.27%**

4.3 Delivery times for technology items

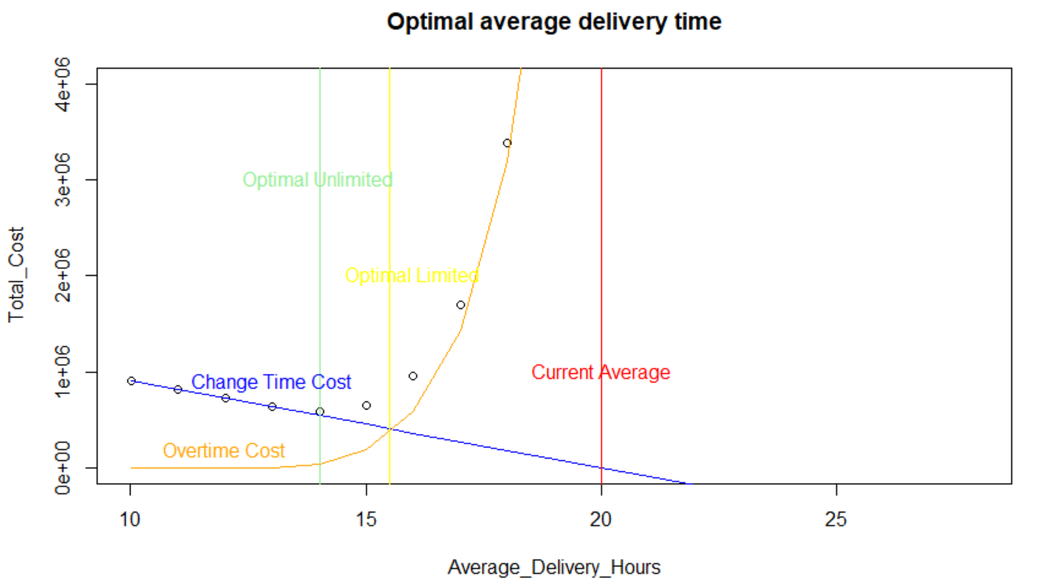

Figure 31 Delivery times for technology items

It becomes evident in the graph above that the total cost increase as the average delivery days increase. Thus, indicating a direct relationship between these 2 variables. The cost for faster deliveries and lost sales cross each other at the 16-hours mark. Currently the delivery time is at 26 hours. Thus, if we decrease our hours by 4, we would have optimal hours at hour 16. Yielding the lowest cost. At a total cost of almost 2600000. Thus meaning 4 days less than the current Average.

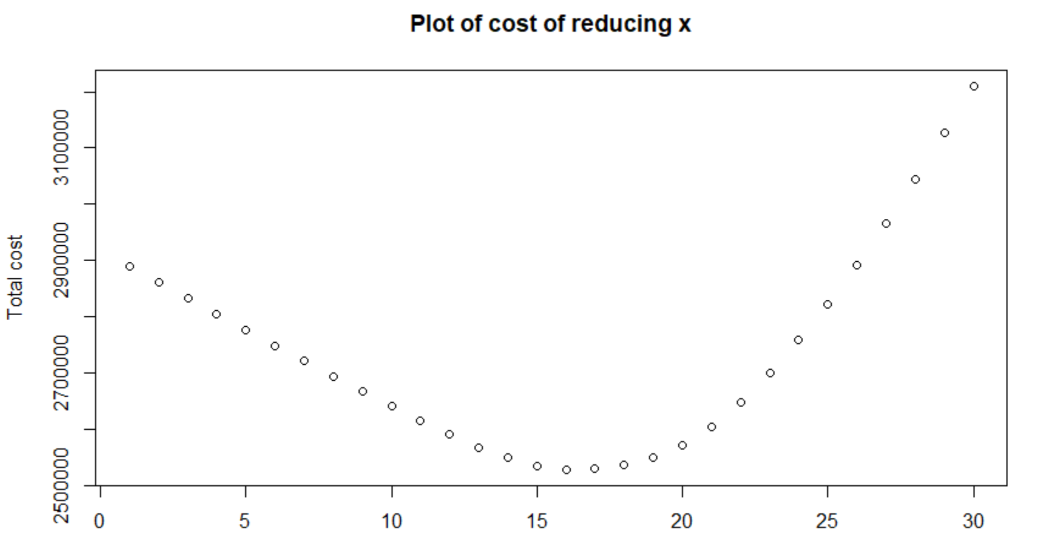

Figure 32 Graph Showing Total Cost vs Hours

Here is another graph with the same information displayed above. This graph shows the relationship between Total Cost and the hours. A minimum is reached in the 16-hour region, which means that the total cost is at a minimum.

4.4 Type 2 (Consumer's) Error

Not rejecting the null hypothesis when it is false. Meaning we assume that the mean we are working with H_0 are right. But in this case the H_0 is false. Meaning we might reject products or good which actually are correct and we might accept product or goods which actually needs to be rejected. Visually it means that H_0 is the graph on the left which have deviated from the H_1 which we assume to be correct. We will reject H_0 to the left of a certain value and accept H_0 to the right of this value. The probability of making a type 2 error would be the area to the right of this value. That is the blue area. (Not rejecting the null hypothesis when it is wrong) This is for a one tailed example.

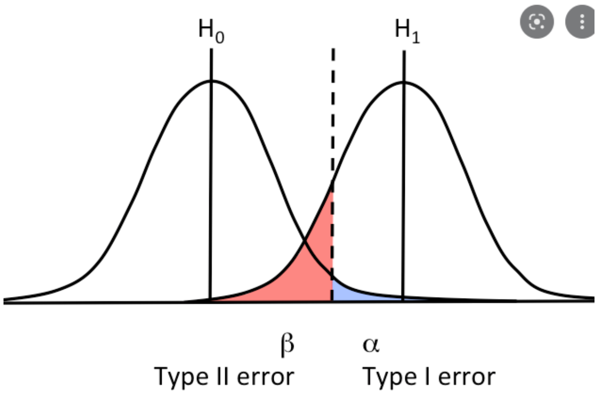

Figure 33 Type 2 Error

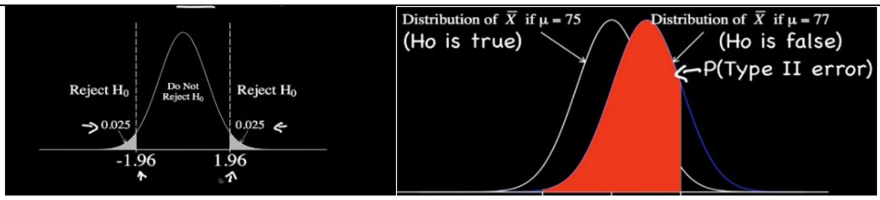

Figure 34 Graphs displaying two tailed examples

We are actually dealing with a two tailed example. Where there is a region where we do not reject H_0 and 2 areas where we reject H_0. This graph shifts to a side due to the H_0 that is false. And causes the wrong parts the be accepted and rejected. The probability of making a type 2 error would be the probability of the new graph (blue), having values between the old (white graph) H_0's. The type 2 error is the area indicated in red.

The same rules apply for the given question.

```
## 4.4
TechUCL<-XBarbarTotal[[4]]+0.789*sbarTotal[[4]]
TechLCL<-XBarbarTotal[[4]]-0.789*sbarTotal[[4]]
Standarddev<-((TechUCL-TechLCL)/6)
answer<-(pnorm(TechUCL,mean=23,sd=Standarddev)-pnorm(TechLCL,mean=23,sd=Standarddev))
## TYPE 2 error likelihood= 0.48831767
```

Figure 35 Screenshot of R-Code

We firstly work out the standard deviation. We get this value by calculating the Upper control limit minus the lower control limit. Thus, area would cover 6 sigma, 3 sigma above and 3 sigma below. Thus, we need to divide the answer by 6 to get the standard deviation for one sigma. The probability of making a type 2 error would be the probability of x bar being less than the upper control limit given a mean of 23 minus x bar being greater than the lower control limit given a mean of 23.

The likelihood of making a type 2 error is 0.48831767, meaning **48,832%** probability. Which makes sense because graphically it would be similar.

5. DOE and MANOVA

Here we are exploring the relationship between classes and why bought. This was not an arbitrarily decision, the decision was made on the hope that a clear relationship between these values could be established, given the information in previous questions.

Null hypothesis: Variables (Price, Age, Delivery Time) has no influence on the reason why a certain item is bought

Alternative hypothesis: Variables (Price, Age, Delivery Time) has a significant influence on why a certain item is bought.

Your instincts tell you that the price would for sure have an effect on the reason why an item would be bought for example expensive items would sell less than relatively inexpensive items.

The MONOVA, is calculated. Based on the results of this MONOVA. If there is enough information then the null hypothesis may be accepted or rejected.

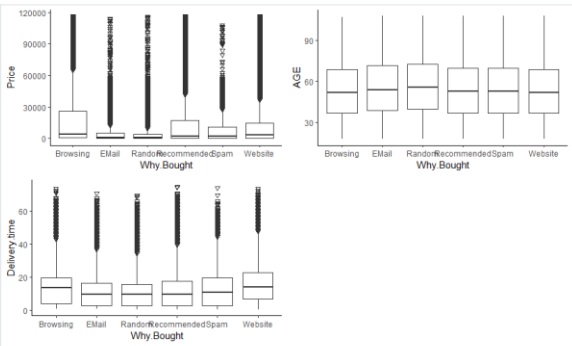

Figure 36 MANOVA Results

```
Descriptive:
   Why.Bought      n    AGE     Price  Delivery.time
1    Browsing  18990 53.845 16133.660         14.740
2       EMail   7224 55.757  6661.580         14.422
3      Random  13121 56.963  4288.261         14.179
4 Recommended 106975 54.482 13441.292         13.226
5        Spam   4208 54.659  9360.900         15.235
6     Website  29443 53.965 11017.647         19.033

Wald-Type Statistic (WTS):
           Test statistic df   p-value
Why.Bought "12950.513"    "15" "<0.001"

modified ANOVA-Type Statistic (MATS):
           Test statistic
Why.Bought       12412.12

p-values resampling:
           paramBS (WTS) paramBS (MATS)
Why.Bought "<0.001"      "<0.001"
```

Figure 37 MANOVA p-values

From the result above it is clear that the p-value of 0.001 is much smaller than the alpha value of 0.05. (p=0.001 < α=0.05) Thus, the null hypothesis needs to be rejected. In the case were on hypothesis is rejected the other needs to be accepted. Meaning that one or more of these variables actually have a significant impact on the reason why an item will be bought.

It is difficult o determine which one of these values actually have a significant impact. That is why we need to compare the boxplots of these values to ensure we can figure out which of the variables has the most significant impact on the reason why an item would be bought.

Starting with the variable which makes the most sense to have a large impact on the reason why a certain item would be bought, Price.

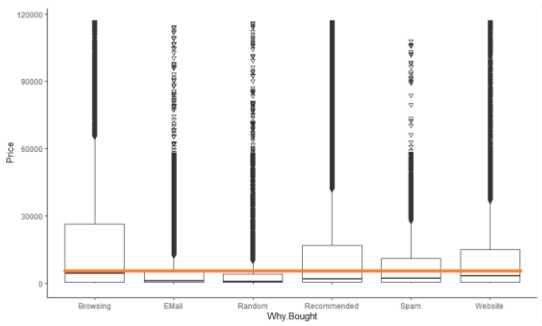

Figure 38 Price vs Why Bought

The mean for all the given classes differs, meaning there is a lot of variability within this variable. Large numbers of outliers can also be identified when drawing the boxplots of the given classes.

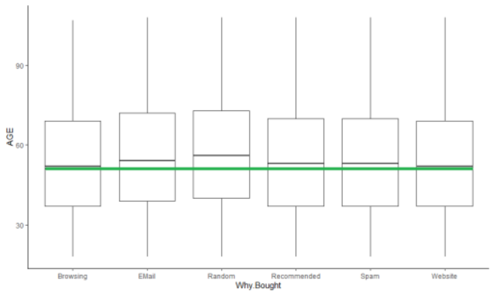

Figure 39 Age vs Why Bought

The age variable has a lot of means which are similar to each other. This means that the variable has a lot smaller impact on the reason why a certain item would be bought. Age’s contribution would be much less than that of price.

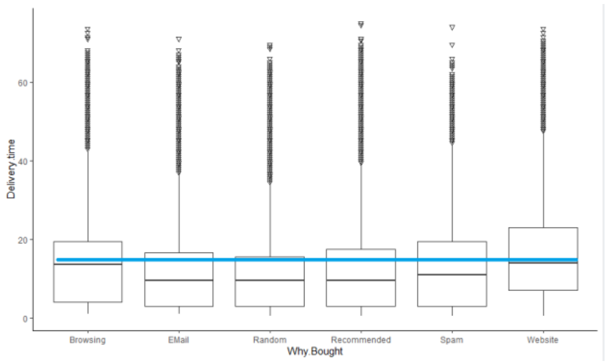

Figure 40 Delivery time vs Why Bought

The upper limit of the delivery time variable has less variability within it. Email, random, recommended and spam are close on the means. Browsing and website have means which differ but it will still be less variability than the Price

6. Reliability of the service and products.

6.1 Food deliveries must be kept cool during transit. The company has a subsidiary, **Lafrideradora**, who makes components for their units. Do Problem 6 and 7 of chapter 7 (page 363) and report the results to your management.

Problem 6

Refrigeration part 0.06 $\pm$ 0.04cm.

$45 to scrap part outside of specifications.

Find the Taguchi loss function for this situation.

$$L(x) = k(x - T)^2$$

$$45 = k(0.04)^2$$

$$k = 28125$$

Taguchi loss function:

$$\mathbf{L(x) = 28125(x - T)^2}$$

$$\mathbf{L = 28125(x - 0.06)^2}$$

Problem 7

7.a)

$$L(x) = k(x - T)^2$$

$$35 = k(x - T)^2$$

$$35 = k(0.04)^2$$

$$k = 21875$$

Taguchi loss function:

$$\mathbf{L(x) = 21875(x - T)^2}$$

$$\mathbf{L = 21875(0.027)^2}$$

$$\mathbf{L = 15.95}$$

7.b)

$$35 = \mathrm{k}(\mathrm{x} - \mathrm{T})^2$$

$$35 = k(0.027)^2$$

$$k = 48010.974$$

Taguchi loss function:

$$\mathbf{L(x) = 48010.974(x - T)^2}$$

6.2 Another subsidiary, **Magna plex**, manufactures some your technology items. Management thinks their current process is wasteful due to the identical machines used as backup in case of failures. Do Problem 27 of chapter 7 and report the results to management.

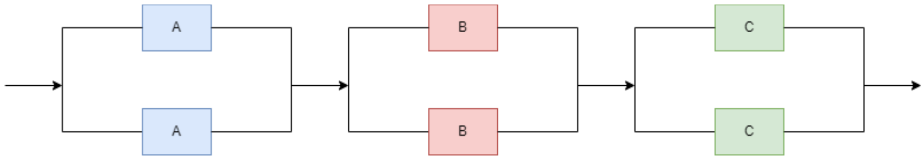

Figure 41Diagram of machine A,B,C

Machine	Reliability
A	0.85
B	0.92
C	0.90

Table 9 Machine Reliability

Use only one machine at each stage:

$$R_s = R_a * R_b * R_c$$

$$R_s = 0.85 * 0.92 * 0.90$$

$$R_s = 0.7038$$

Use two machines at each stage (Backup machine):

Failure rate= 1-Reliability

Machine	Reliability
A, A (one backup)	1-(0.15*0.15) =**0.9775**
B, B (one backup)	1-(0.08*0.08) =**0.9936**
C, C (one backup)	1-(0.1*0.1) =**0.99**

Table 10 Machine Reliability

$$R_s = R_{aa} * R_{bb} * R_{cc}$$

$$R_s = 0.9775 * 0.9936 * 0.99$$

$$R_s = 0.962$$

Observation:

Making use of a backup machine would drastically increase the reliability of a system. If a system is implemented it does not necessarily mean that it would be reliable. A system in series is dependent on the previous machine which mean one machine has the ability to switch a whole operation down. If there are penalties which the company can suffer then it would be unwise to have a low reliability rating. Penalties will accumulate and make a large dent in the companies ROI.

6.3 For the delivery process, there are **20 delivery vehicles available**, of which 19 is required to be operating at any time to give reliable service. During the past **1560 days**, the number of days that there were **only 20 vehicles available** was **190** days, only **19 vehicles** available was **22** days, only **18 vehicles availabl**e was **3** days and **17 vehicles available** only **once**.

There are also **21 drivers**, who each work an **8-hour shift per day**. During the **past 1560** days, the number of days that there were only **20 drivers available** was **95** days, only **19 drivers available** was **6 day**s and only **18 drivers available**, once only.

a.) Estimate on how many days per year we should expect reliable delivery times, given the information above.

b.) If we increased our number of vehicles by one to 21, how many days per year we should expect reliable delivery times? Hints: This example is based on Binomial probabilities. See chapter 6 for formulae. Part 2, when increasing the number of vehicles to 21, we must assume the probability of failure per vehicle stays constant.

Table summarising all the information

Vehicles Available	**Days**	**Drivers Available**	**Days**
20	190	20	95
19	22	19	6
18	3	18	1
17	1		

Table 11 Vehicle Availability

Using the binomial distribution formula:

$$f(x) = \binom{n}{x} * p^x * (1-p)^{n-x}$$

6.3. a) **Only considering vehicles:**

21 Vehicles

- P(21 vehicles) = $\frac{1560-190-22-3-1}{1560} = \binom{21}{0} * p^0 * (1-p)^{21-0}$

p-value (P(21 vehicles)) =0.0070718085

- P(20 vehicles) = $\frac{190}{1560} = \binom{21}{1} * p^1 * (1-p)^{21-1}$

p-value (P(20 vehicles)) =0.0066242837

19 Vehicles

- P(19 vehicles) = $\frac{22}{1560} = \binom{21}{2} * p^2 * (1-p)^{21-2}$

p-value (P(19 vehicles)) =0.0089231847

18 Vehicles

- P(18 vehicles) = $\frac{3}{1560} = \binom{21}{3} * p^3 * (1-p)^{21-3}$

p-value (P(18 vehicles)) =0.0121699342

17 Vehicles

- P(17 vehicles) = $\frac{1}{1560} = \binom{21}{4} * p^4 * (1-p)^{21-4}$

p-value (P(17 vehicles)) =0.0196856635

Weighted p:

$$\frac{1344 * (0.007071) + 190 * (0.00662) + 22 * (0.008923) + 3 * (0.0121699) + 1 * (0.019685)}{1560}$$
$$= 0.007061301318$$

P (0 failures) = $\binom{21}{0} * (0.007061301318)^0 * (1 - 0.007061301318)^{21-0}$

P (1 failures) = $\binom{21}{1} * (0.007061301318)^1 * (1 - 0.007061301318)^{21-1}$

P (2 failures) = $\binom{21}{2} * (0.007061301318)^2 * (1 - 0.007061301318)^{21-2}$

P (3 failures) = $\binom{21}{3} * (0.007061301318)^3 * (1 - 0.007061301318)^{21-3}$

P (4 failures) = $\binom{21}{4} * (0.007061301318)^4 * (1 - 0.007061301318)^{21-4}$

Days with 0 failures = 0.862060468*1560 =1344.81

Days with 1 failures = 0.1284087843*1560 =200.76

Days with 2 failures = 0.009108197601*1560 =14.2087

Days with 3 failures = 0.0004091687901*1560 =0.6383

Days with 4 failures = 0.00001306028708*1560 =0.023

Days with Reliable delivery= $\frac{1344+200}{1560}$*365 =**364.85 days**

21 Drivers:

- P(21 drivers) = $\frac{1560-95-6-1}{1560} = \binom{21}{0} * p^0 * (1-p)^{21-0}$

p-value (P(21 drivers)) =0.0032148302

- P(20 drivers) = $\frac{95}{1560} = \binom{21}{1} * p^1 * (1-p)^{21-1}$

p-value (P(20 drivers)) =0.0030847117

- P(19 drivers) = $\frac{6}{1560} = \binom{21}{2} * p^2 * (1-p)^{21-2}$

p-value (P(19 drivers)) =0.0044654848

- P(18 drivers) = $\frac{1}{1560} = \binom{21}{3} * p^3 * (1-p)^{21-3}$

p-value (P(18 drivers)) =0.008239491

Weighted p:

$$\frac{1458 * (0.0032148302) + 95 * (0.0030847117) + 6 * (0.0044654848) + 1 * (0.008239491)}{1560} = 0.003214937463$$

P (0 failures) = $\binom{21}{0} * (0.003214937463)^0 * (1 - 0.003214937463)^{21-0}$

P (1 failures) = $\binom{21}{1} * (0.003214937463)^1 * (1 - 0.003214937463)^{21-1}$

P (2 failures) = $\binom{21}{2} * (0.003214937463)^2 * (1 - 0.003214937463)^{21-2}$

P (3 failures) = $\binom{21}{3} * (0.003214937463)^3 * (1 - 0.003214937463)^{21-3}$

Days with 0 failures = 0.9346132739*1560 =1458

Days with 1 failures = 0.06330270201*1560 =98.75

Days with 2 failures = 0.00204170624*1560 =3.18506

Days with 3 failures = 0.00004170581482*1560 =0.06506107

Days with Reliable delivery= $\frac{1458+99+3}{1560}$*365 =**365 days**

Vehicle Reliability: $\sum_{x=0}^{1}\binom{21}{x} * (0.00706130)^x * (1 - 0.0070613)^{21-x}$=0.9999877326

Driver Reliability: $\sum_{x=0}^{1}\binom{21}{x} * (0.003214937)^x * (1 - 0.003214937)^{21-x}$=0.999599

Total Reliability: 0.9999877326*0.9999599= 0.999947707

0.999947707*365 =**364.82 days / 365 days**

For 22 Vehicles and 21 Drivers:

Vehicle Reliability: $\sum_{x=0}^{2}\binom{22}{x} * (0.00706130)^{x} * (1 - 0.0070613)^{22-x}$=0.9995096574

Driver Reliability: $\sum_{x=0}^{2}\binom{22}{x} * (0.003214937463)^{x} * (1 - 0.003214937463)^{22-x}$=0.9999511182

Total Reliability: 0.9995096574 *0.9999511182= 0.9994607996

0.999947707*365 =**364.98 days / 365 days**

Would not make a significant difference when adding another vehicle, reliability are already really high.

Conclusions

Through the completion of this project significant improvement were made to our programming skills as well as our ability to do data analytics on large data sets. We learned a lot about the reliability of a system and how it can be improved using systems which are in series or parallel. Computations of the Taguchi Loss functions, made us aware of the significance of these equations. Valuable lessons were learned about the implementation of statistical process control on a specific product. The statistical process control methodology can be applied to any product or service to determine whether this product or service will meet the high expectations of the customer. This implementation of statistical process control can give a company a competitive advantage, allowing the company to gain more profit in the long run.

References

van Schalkwyk, T., 2022. *QA344 Statistics*. [Accessed 19 October 2022].

"Control Chart Rules and Interpretation." *BPI Consulting*, https://www.spcforexcel.com/knowledge/control-chart-basics/control-chart-rules-interpretation. [Accessed 19 October 2022].

Youtube.com. 2022. [online] Available at: <https://www.youtube.com/watch?v=NbeHZp23ubs> [Accessed 19 October 2022].

Appendix A

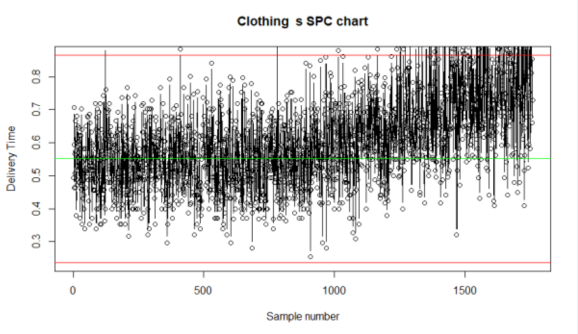

Figure 42 Clothing S-Chart [1: end]

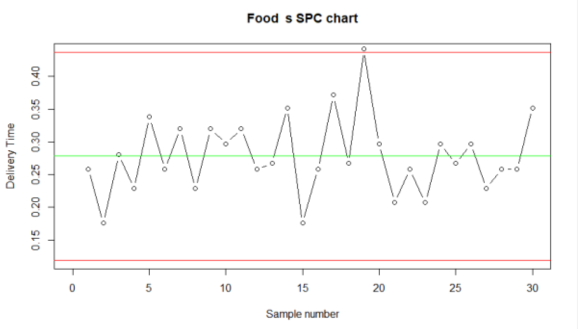

Figure 43 Food S-Chart [1:30]

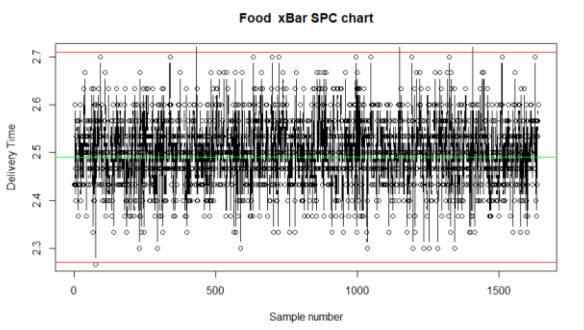

Figure 44 Food X-Chart [1: end]

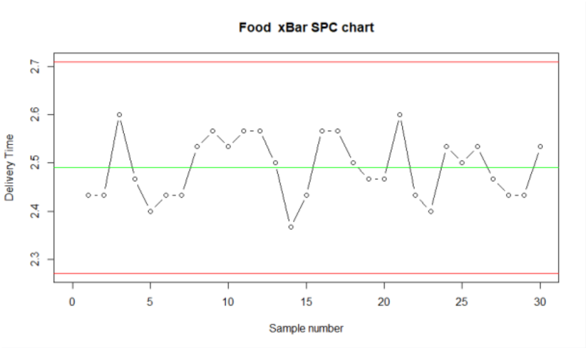

Figure 45 Food X-Bar Chart [1:30]

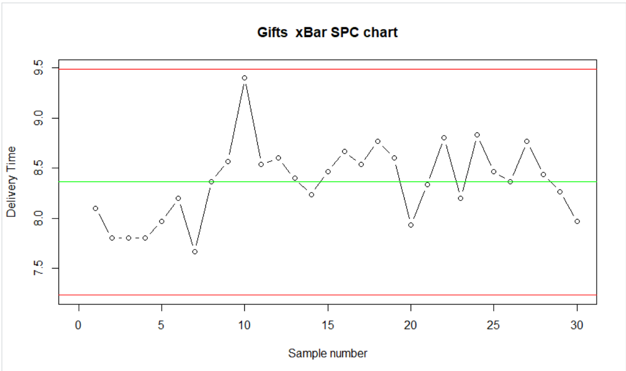

Figure 46 Gifts X-Chart [1:30]

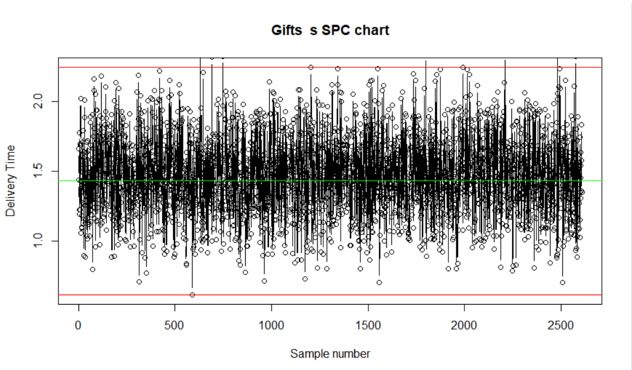

Figure 47Gifts S Chart [1: end]

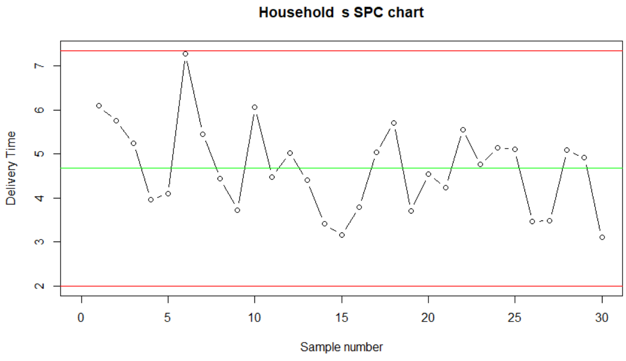

Figure 48 Household [1:30]

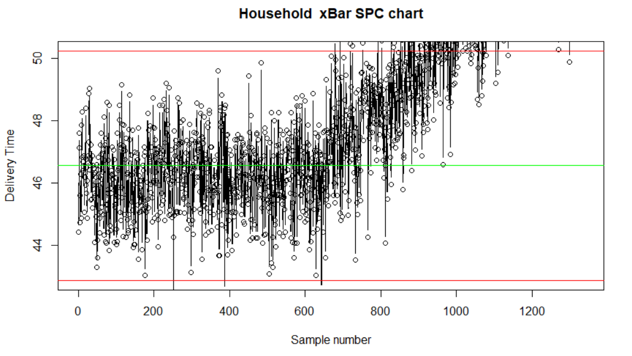

Figure 49 Household X Chart [1: end]

Figure 50 Household [1:30]

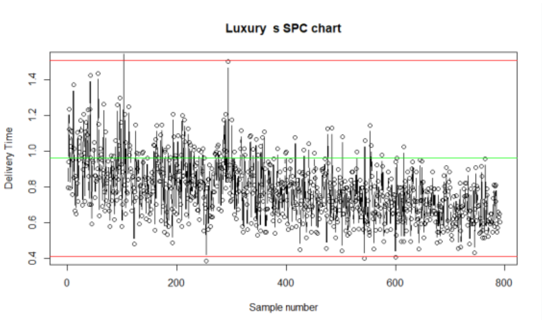

Figure 51 Luxury S Chart [1: end]

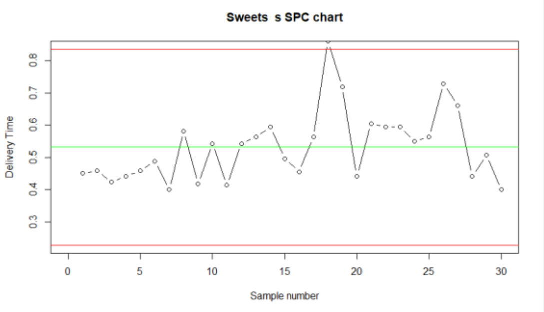

Figure 52 Sweets S Chart [1:30]

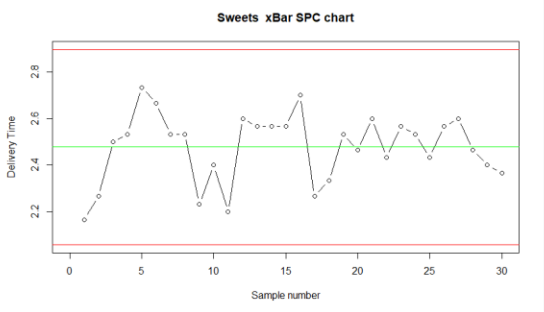

Figure 53 Sweets X Chart [1:30]

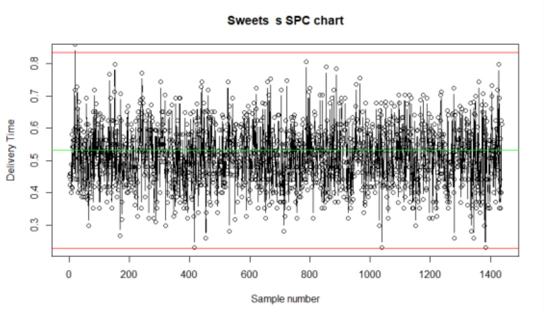

Figure 54 Sweets S Chart [1: end]

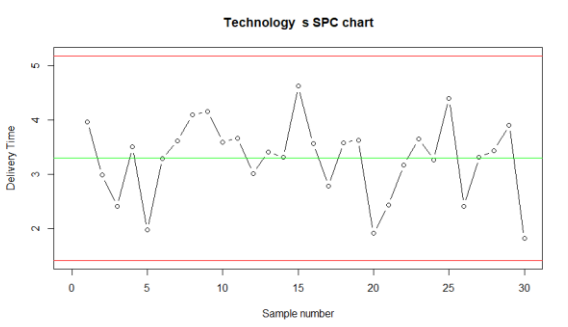

Figure 55 Technology S Chart [1:30]

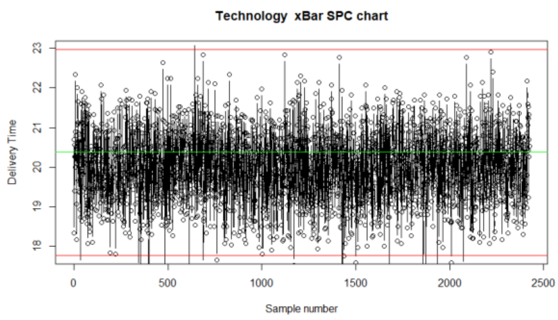

Figure 56 Technology X-Chart [1: end]

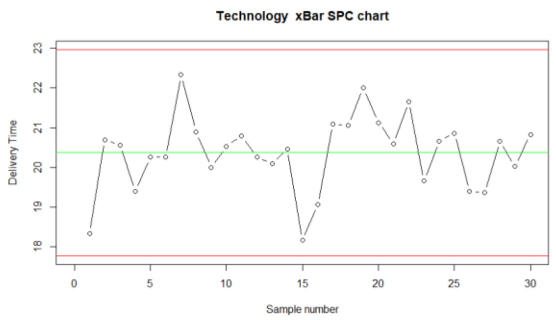

Figure 57 Technology X Chart [1:30]

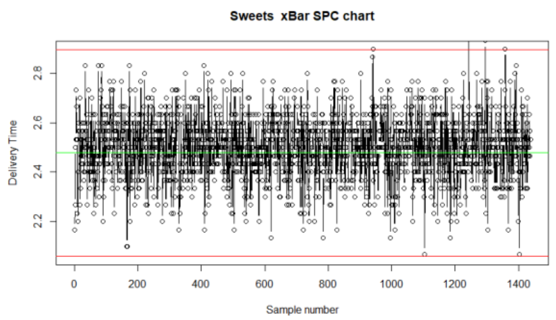

Figure 58 Sweets X- Chart [1: end]