ECSA QUALITY CONTROL REPORT

Quality assurance 344

TALITA HEYNS
Stellenbosch Student number: 23540141

Oct 2022

Table of Contents

Part 1: Data wrangling 2

Part 2: Descriptive Statistics 2

2.1 Data analysis 2

2.1.1 Analysis of product classes 2

2.1.2 Analysis of Price of sales 4

2.1.3 Analysis of reason for purchase 6

2.2 Process capability indices: 7

Part 3: Statistical process control 8

3.1 Control charts 8

3.2 All data samples averages and standard deviations 10

Part 4: Optimizing the delivery process 11

4.1 Lists and plots of out-of-control samples 11

4.1.1 For class type “Technology”: 12

4.1.2 For class type “Clothing”: 14

4.1.3 For class type “Gifts”: 15

4.2 Likelihood of making a Type I or Manufacturer’s error 16

4.3 Centring the delivery process for best profit 17

4.4 Type II error for A 19

Part 5: DOE and MANOVA 20

5.2 Hypotheses and MANOVA results 20

Part 6: Reliability of the service and products 22

6.1.1 Problem 6 p 363 22

6.1.2 Problem 7 p 363 22

6.2 Problem 27 Chapter 7 p 363 23

6.3 Given Question 24

6.3.1 Initial question 24

6.3.2 Increase the number of vehicles to 22 27

Conclusion: 28

References: 29

List of Tables:

Table 1: X-bar Control Chart ... 9
Table 2: S Control Chart ... 9
Table 3: Top section of full samples table in R ... 10
Table 4: Bottom section of full samples table in R ... 10
Table 5: Outlier positions per class group identified ... 11
Table 6: Outlier x-bar values per class group identified ... 11
Table 7: : Table of most consecutive s-bar instances per class group ... 12
Table 8: Section of total cost, cost of being overtime, cost of shifting the average per iteration of average hours shifted ... 17
Table 9: results of summary(manova_data) ... 20
Table 10: results of eta_squared(manova_data) ... 20
Table 11: The trial and error resulting p-values per number of vehicles unavailable ... 25
Table 12: The trial and error resulting p-values per number of employees unavailable ... 26

List of Figures

Figure 1: Number of purchases per class per year matrix ... 2
Figure 2: Boxplots of Age vs class group in sale entries ... 3
Figure 3: Price per class item over the years of sales ... 4
Figure 4: Scatter plot of prices per class over the years ... 4
Figure 5: Delivery time vs Price of sale per item class ... 5
Figure 6: The number of sale transactions per reason of purchase ... 6
Figure 7: instance densities of age per reason for purchase ... 6
Figure 8: Histogram of data vs its LSL and USL ... 8
Figure 9: X-chart for Technology class group ... 9
Figure 10: Plot section of sample averages for Technology vs control limits ... 13
Figure 11: Section of most consecutive Technology sample standard deviations ... 14
Figure 12: Plot section of sample averages for Clothing vs control limits ... 15
Figure 13: Section of most consecutive Gifts sample standard deviations ... 16
Figure 14: Total cost per shift in delivery times ... 18
Figure 15: Visualization of Type II error for Technology rule A ... 19
Figure 16: Boxplots of Price vs Class types ... 21
Figure 17: Boxplots of Delivery Times vs Class types ... 21
Figure 18: Pr 27 p 363 given figure ... 23

List of Equations:

Equation 1: binomial distribution ... 25

Introduction

Statistics and analysing data are one of the most important and common practises in the management of production and a company in general. One can only plan accurately for the future if one understands what has happened in the past and how exactly the parameters should be adjusted. This can only be effectively done by statistics and analysing previous data.

This report therefore focusses on some of the main principles of statistics and data analytics that are common in the industrial engineering world such as maintaining quality control and identifying and adjusting factors to ensure quality control may be attained in the future. This report will therefore focus on an example dataset of sales information and inspect and determine its quality issues and also inspect how quality control systems can be applied to ensure future improvement of quality performance.

Firstly, the report consists of wrangling the raw dataset into a usable data set, thereafter the data is analysed by descriptive statistics to get a general idea what the data is about and how the data features correlate to each other including some process capability indices. The data is then processed to do statistical process control using x and s tables and charts. Thereafter the delivery processes are optimized by determining type I and type II errors and error indicators. MANOVA is then used to analyse a hypothesis. The final section then determines the reliability of the service and products by inspecting how the change in number of total vehicles and employees (along with days they aren't will affect the number of days expected to provide a reliable service.

The information provided in this report is mainly inspired by the textbook *Managing for Quality and Performance Excellence* by J.R Evans and W.M. Lindsay (2019).

Part 1: Data wrangling

The data needed to be split into invalid and valid datasets to use the valid data set that is fully completed and accurate, containing no missing values or faulty values such as negative Price values and missing values named as "NA". The separate datasets were also adjusted to be logical to follow and use such as ordering the dates in increasing order.

Part 2: Descriptive Statistics

2.1 Data analysis

2.1.1 Analysis of product classes

For the first data analysis, the number of sales instances were counted per class, to visualize the mode class of sales:

Figure 1: Number of purchases per class per year matrix

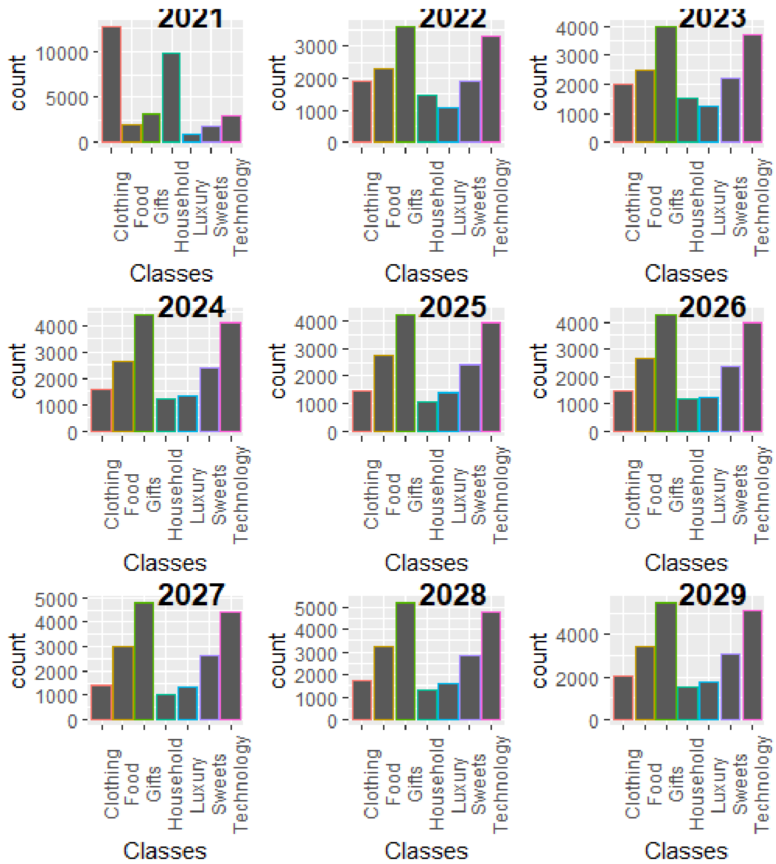
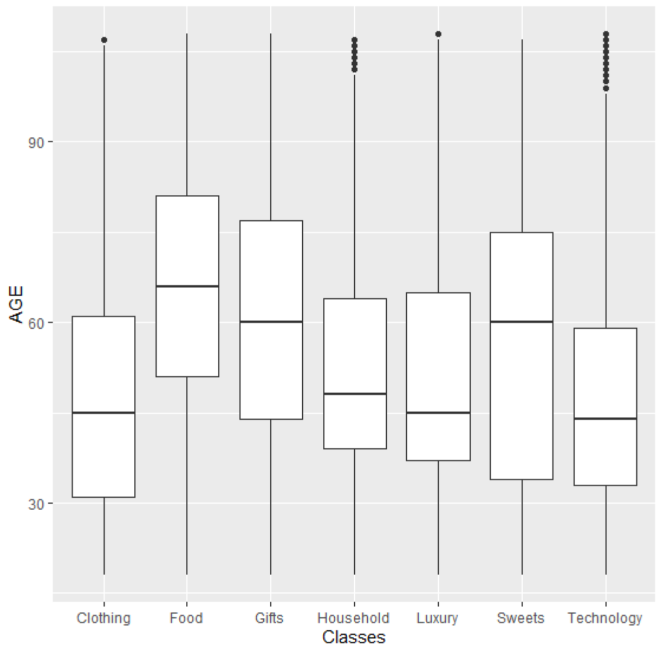


Figure 1 shows how the class groups variated randomly from year 2021 to 2022 but how the class group distribution remained approximately the same from 2022 to 2029. These bar plots show that the mode product class was "Gifts" for 2022 until 2029 but that it was "clothing" in 2021. The least bought product class was "Luxury" for 2021 until 2023 but the least bought product class changed to "Household" for 2024 until 2029. From 2022, the class "Technology" also increased greatly to remain the second most commonly sold class product for 2022 until 2029.

To further investigate the sales per class, the differences in age per class purchase is also inspected in this boxplot.

Figure 2: Boxplots of Age vs class group in sale entries



The box plots show the average age group mainly interested per class product as well as uncommon outliers that can be useful for future product designs to consider the main target audience.

Figure 2 for example shows the item class "Food", "Sweets" and also "Gifts" to be purchased by a mean age between 60 and 70, with the "Technology" class items to be purchased by the mean age between 40 and 50 but with a couple of outliers older than 90. The ages of the customers for Sweets also has the greatest distribution interval (having an inter quartal range from 35 to 75) than the rest of the class groups' ages.

2.1.2 Analysis of Price of sales

To compare the Price per class item and how that has changed over the years of sales, the following jitter plot matrix was created.

Figure 3: Price per class item over the years of sales

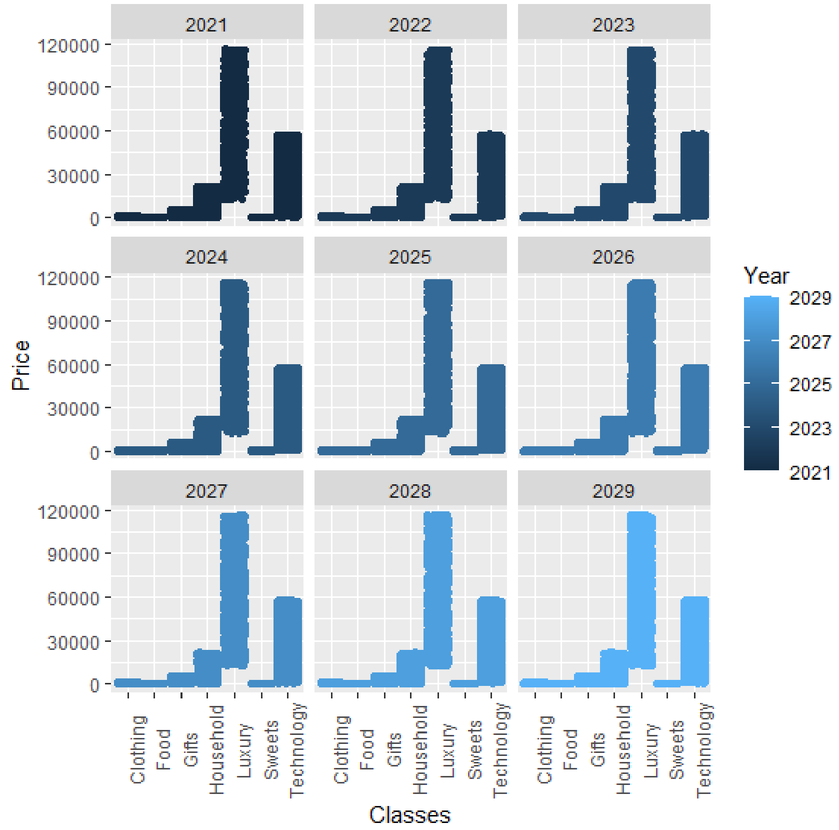


Figure 4: Scatter plot of prices per class over the years

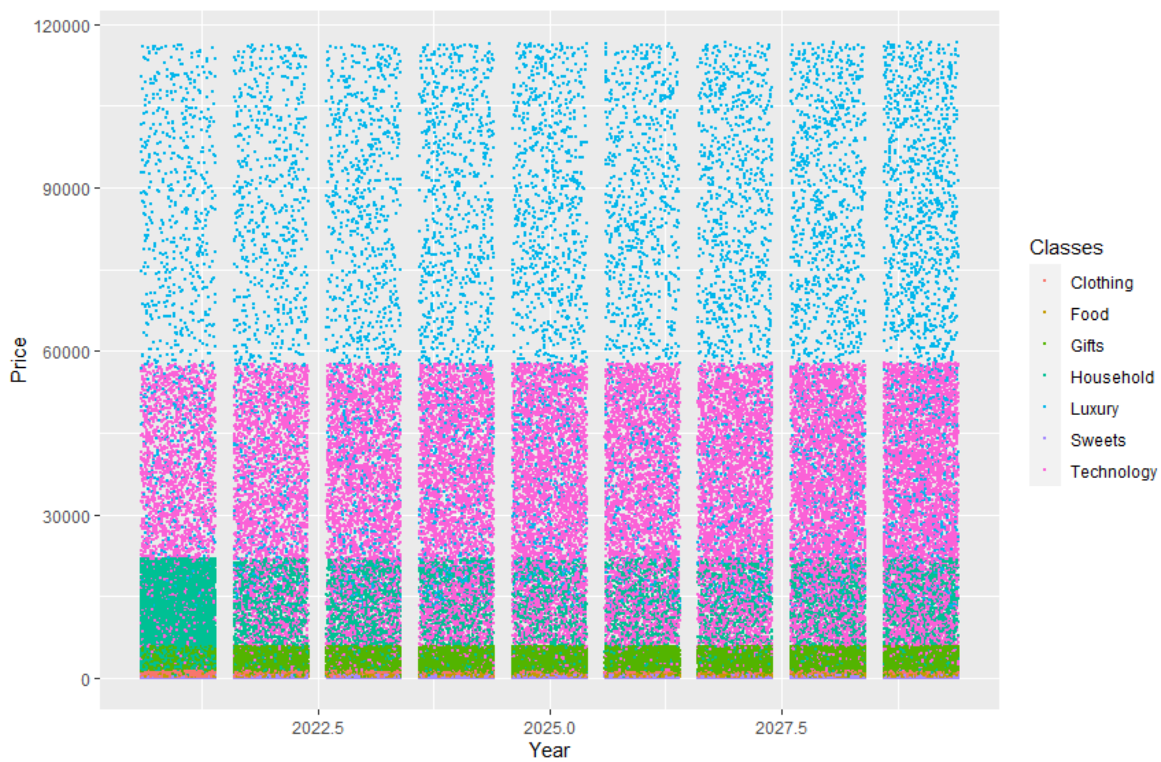
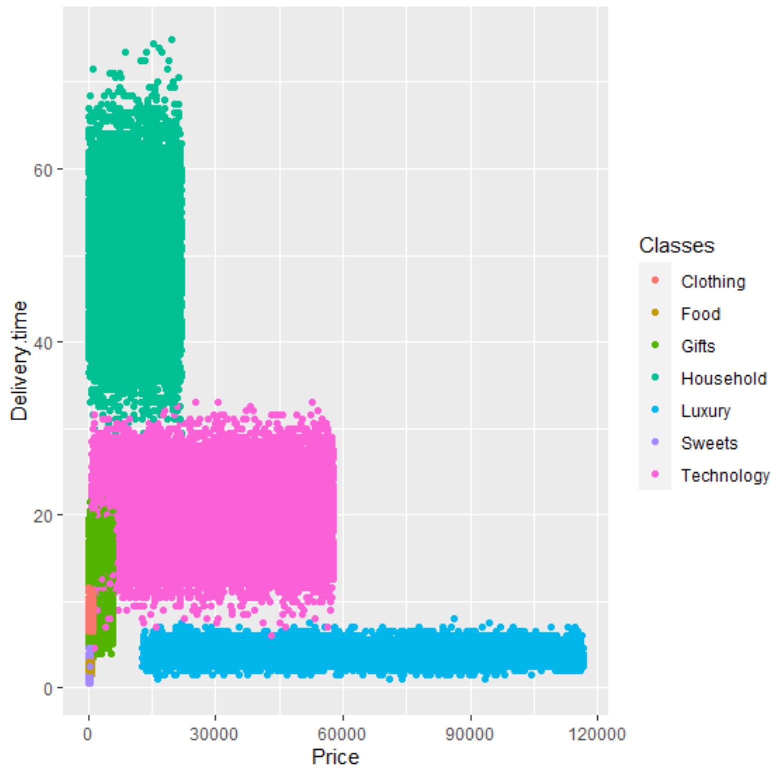


Figure 3 indicates that the prices per class item remained mostly the same over the sales years and as well as the difference in price per class. This is also confirmed by the steady levels of price ranges per class in Figure 4, proving that there is no relationship between the prices and year of sale per class type. For the given data, the prices per class groups therefore didn't change as the years progressed. Figure 4 therefore also shows that there is a uniform distribution of prices over the years of sales per class group.

The Luxury class items has by far the highest prices, price minimum as well as greatest price range, with Food, clothing and Sweets having the lowest prices and price ranges throughout all the years' instances.

To see if there might be a correlation between delivery time and price, the following colour grouped scatter plot was created.

Figure 5: Delivery time vs Price of sale per item class

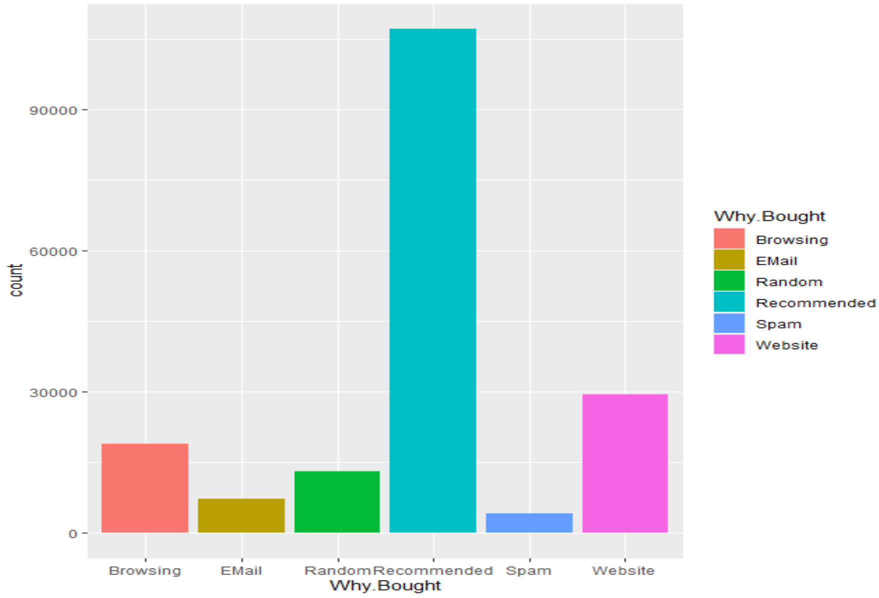


The graph indicates that the longest delivery time are that of household products, but does not have the highest prices. The long delivery times will therefore rather be due to factors such as product size and not necessarily the product price. Technology however also has the second highest delivery times as well as prices, even though they can be relatively small in size generally. This could suggest that the prices might be higher for technology due to also longer delivery times.

2.1.3 Analysis of reason for purchase

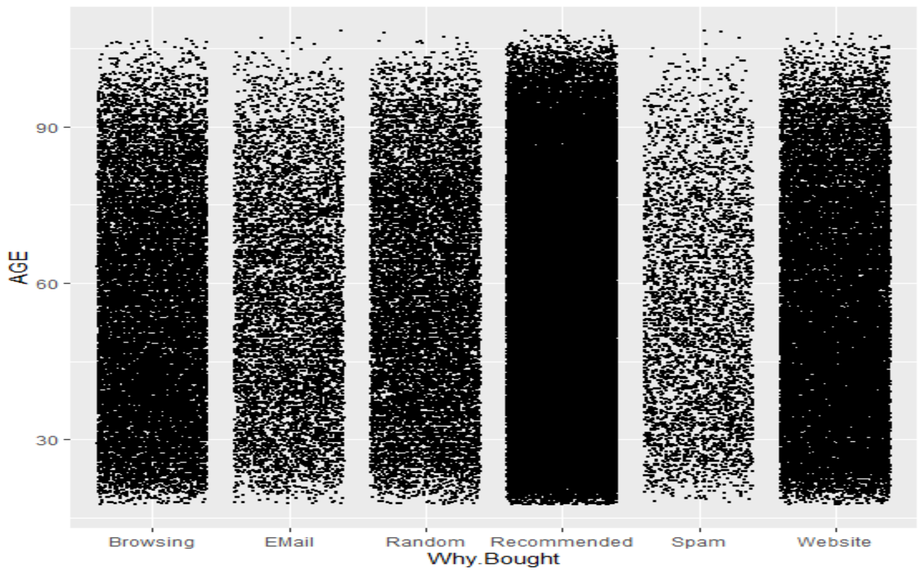
To inspect the most successful method of gaining more clients and sales, the following bar plot was created.

Figure 6: The number of sale transactions per reason of purchase



The graph shows how recommendations are the mode and greatest reason of purchases, indicating a good general customer satisfactions and support. It however also shows that emails and spam are the least common reasons for purchases. This can help sales know which advertising methods to invest in and even perhaps leave like emails. To understand the reasons for purchases of our customer spread better, the jitter plot of age per reason for purchase was created.

Figure 7: instance densities of age per reason for purchase



By looking at the areas of highest densities per reason of purchase, this graph shows that the reason of purchase “recommendations” is spread equally across all ages of clients. The reason of purchase of “browsing” as well as “random” is more frequent among the age group between 30 and 60 but the other internet related reason of “website” is frequent in a greater range of ages from 25 to 70.

2.2 Process capability indices:

The process capability indices of the process delivery times for the technology class items were determined as:

for LSL = 0 and USL = 24 hours

Cp = (UCL – LCL)/(6*standard deviation)

Cp = 1.142

Cpl = (average – LCL)/(3*standard deviation)

Cpl = 1.905

Cpu = (UCL – average)/(3*standard deviation)

Cpu = 0.379

Cpk = min(Cpu, Cpl)

Cpk = 0.379

The assigned LSL is logical to be 0, since the process delivery times cannot be a negative number of hours or therefore less than 0, so the logical minimum will be 0

The potential capability measures are used to indicate the specification spread, the spread of the process and the process' spread relative to its control limits (Potential (within) capability for Normal Capability Analysis*,* [S.a.]).

The resulting potential capability of the process (Cp) is found to be 1.142 for the valid dataset. This fairly medium (not low <1) value indicates that the data is spread between the Upper and lower specification limits (USL and LSL) so that the USL and LSL interval width is smaller than that of the dataset. The production spread can therefore fit inside the limits 1.142 times, which is barely more than 1, meaning that the production spread has little freedom to move and can therefore be improved.

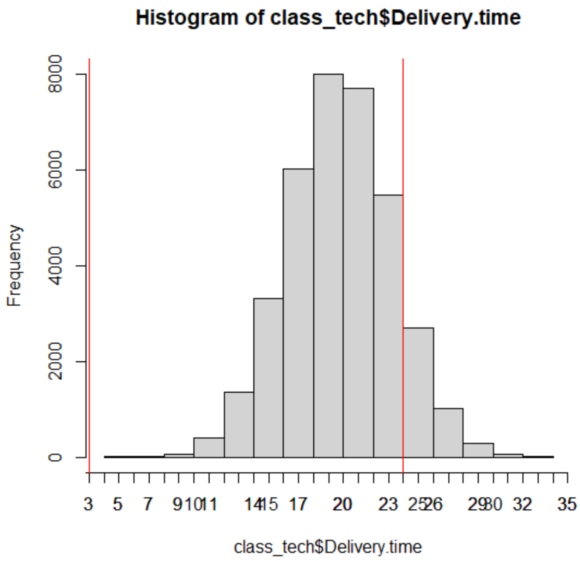
The Cpl and capability measure compared to the lower specification limit is found to be 1.905, indicating a high capacity of the dataset to the lower specification limit.

The Cpu capability measure compared to the higher specification limit on the other hand is found to be 0.379 and quite low, indicating a low capacity towards the upper specification limit.

With the great difference and values of Cpu and Cpk, the data is therefore not centred but towards the right-hand side in the specification limits interval as seen in figure 8.

The Cpk and measure of the potential capability of the entire process is therefore the same as the Cpk which is 0.379. The ideal system that is centred in the middle of the control limits, have the same Cp and Cpk values. The current system can therefore be improved and is not centred neither ideally spread within the limits.

Figure 8: Histogram of data vs its LSL and USL



This figure proves that the data is indeed more centred towards the right side of the control limits.

The dataset centred towards the right-hand side of the control limits and having a number of data instances above the USL, suggests that there are (for these control limits) a number of 'bad' products that lie above the upper control limit**. The 'good' products are considered to be inside the lower and upper control limits, as the voice of the customer of USL and LSL implies. Any products outside these limits are considered as bad products and only products inside the limits are considered good.**

Part 3: Statistical process control

3.1 Control charts

The s-chart uses the sample standard deviations to measure the variation within a subgroup whereas the x-chart uses the sample means. (Xbar-s Control Charts: Part 1 | BPI Consulting - SPC for Excel. [S.a.])

The following x- and s- charts were created for delivery times per class type for the first 30 samples of size = 15 instances to use for process control.

The control limits were created by using the statistics formula sheet and general information such as in the textbook of Evans, J.R, Lindsay W.M. 2019 (*Managing for Quality and Performance Excellence).*

Table 1: X-bar Control Chart

Classes	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
Technology	22.97461588	22.10789207	21.24116826	20.3744444	19.5077206	18.6409968	17.774273
Clothing	9.404933524	9.259955683	9.114977841	8.97	8.82502216	8.68004432	8.5350665
Household	50.24832787	49.01962598	47.7909241	46.5622222	45.3335203	44.1048185	42.876117
Luxury	5.493965126	5.241161936	4.988358746	4.73555556	4.48275237	4.22994918	3.977146
Food	2.709457732	2.636305155	2.563152577	2.49	2.41684742	2.34369485	2.2705423
Gifts	9.488564673	9.112746819	8.736928965	8.36111111	7.98529326	7.6094754	7.2336575
Sweets	2.89704151	2.757286932	2.617532355	2.47777778	2.3380232	2.19826862	2.058514

Table 2: S Control Chart

Classes	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
Technology	5.180569704	4.552222403	3.923875102	3.295527801	2.667180501	2.0388332	1.410485899
Clothing	0.866559568	0.761455227	0.656350886	0.551246545	0.446142204	0.341037862	0.235933521
Household	7.344180066	6.453410134	5.562640203	4.671870271	3.781100339	2.890330408	1.999560476
Luxury	1.511051768	1.327777466	1.144503163	0.96122886	0.777954558	0.594680255	0.411405952
Food	0.437246584	0.384213283	0.331179982	0.278146682	0.225113381	0.17208008	0.11904678
Gifts	2.246333333	1.973877297	1.701421261	1.428965225	1.156509188	0.884053152	0.611597116
Sweets	0.835339146	0.734021506	0.632703866	0.531386225	0.430068585	0.328750945	0.227433304

Figure 9: X-chart for Technology class group

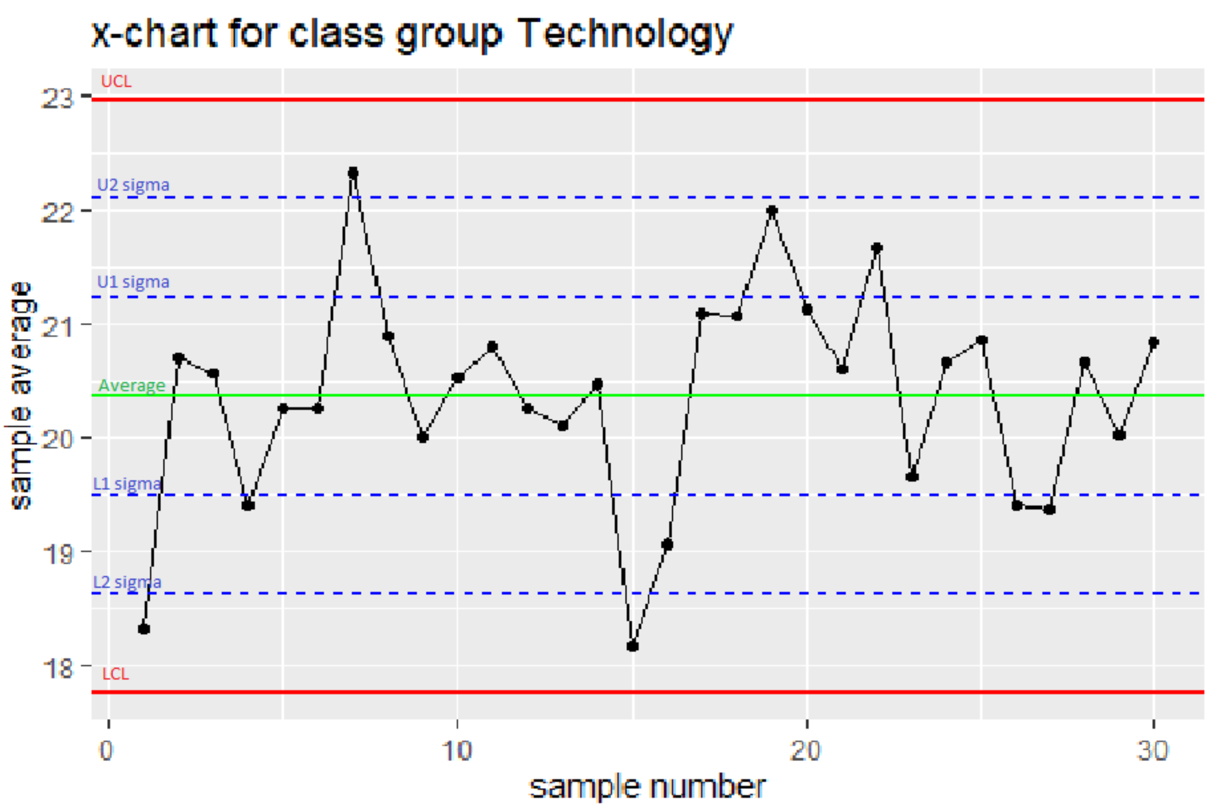


Figure 9 is an example of an x-chart for the class group Technology, where the dataset lies between the UCL and LCL and the dataset is divided into sigma sections.

3.2 All data samples averages and standard deviations

The following shows the top and bottom section of the full table of averages and standard deviations of all further possible samples per class type generated. **This was generated only after the x-chart and s-chart outliers of the first 30 samples were removed to ensure that the initial 30 sample data the control limits are based on, are error-free.**

Table 3: Top section of full samples table in R

	Key	Class_sample	sample_avg	sample_std
2	1	Technology	18.33333	3.967127
3	2	Technology	20.70000	2.986876
4	3	Technology	20.56667	2.411777
5	4	Technology	19.40000	3.501020
6	5	Technology	20.26667	1.971825
7	6	Technology	20.26667	3.288870
8	7	Technology	22.33333	3.618734
9	8	Technology	20.90000	4.089184
10	9	Technology	20.00000	4.153312
11	10	Technology	20.53333	3.592983
12	11	Technology	20.80000	3.663527
13	12	Technology	20.26667	3.011091
14	13	Technology	20.10000	3.413210
15	14	Technology	20.46667	3.308359
16	15	Technology	18.16667	4.620091
17	16	Technology	19.06667	3.565042
18	17	Technology	21.10000	2.785165
19	18	Technology	21.06667	3.580037
20	19	Technology	22.00000	3.630230
21	20	Technology	21.13333	1.912988
22	21	Technology	20.60000	2.428992
23	22	Technology	21.66667	3.171675
24	23	Technology	19.66667	3.643324
25	24	Technology	20.66667	3.265986
26	25	Technology	20.86667	4.393448
27	26	Technology	19.40000	2.414243
28	27	Technology	19.36667	3.313752
29	28	Technology	20.66667	3.436499
30	29	Technology	20.03333	3.907258
31	30	Technology	20.83333	1.819210

Table 4: Bottom section of full samples table in R

	Key	Class_sample	sample_avg	sample_std
11969	11968	Sweets	2.633333	0.3994043
11970	11969	Sweets	2.666667	0.5563486
11971	11970	Sweets	2.300000	0.5606119
11972	11971	Sweets	2.566667	0.4577377
11973	11972	Sweets	2.400000	0.5411628
11974	11973	Sweets	2.666667	0.4498677
11975	11974	Sweets	2.666667	0.4879500
11976	11975	Sweets	2.433333	0.6229729
11977	11976	Sweets	2.433333	0.3716117
11978	11977	Sweets	2.433333	0.5300494
11979	11978	Sweets	2.500000	0.5976143
11980	11979	Sweets	2.633333	0.4805751
11981	11980	Sweets	2.333333	0.4879500
11982	11981	Sweets	2.533333	0.7187953
11983	11982	Sweets	2.500000	0.5669467
11984	11983	Sweets	2.600000	0.4705620
11985	11984	Sweets	2.333333	0.5232681
11986	11985	Sweets	2.700000	0.7973169
11987	11986	Sweets	2.333333	0.4082483
11988	11987	Sweets	2.466667	0.6935073
11989	11988	Sweets	2.533333	0.3518658
11990	11989	Sweets	2.466667	0.4418576
11991	11990	Sweets	2.466667	0.3994043
11992	11991	Sweets	2.633333	0.3518658
11993	11992	Sweets	2.500000	0.5669467
11994	11993	Sweets	2.466667	0.3994043
11995	11994	Sweets	2.566667	0.6229729
11996	11995	Sweets	2.533333	0.6113996

Part 4: Optimizing the delivery process

4.1 Lists and plots of out-of-control samples

To make the data understandable, the out-of-control samples will be indicated per class type. Plots will also be created to visualize the out-of-control samples for the first class-type "Technology" to prove and visualize the method used to identify also the rest of the class types' out-of-control samples.

Rule A: The sample means outside the outer control limits

Rule B: The most consecutive samples of the standard deviations between -0.3 and +0.4 sigma control limits

The tabular summary of outliers found per class group for the x-bar values is seen in Table 5 and 6 (rule A summary). Further descriptions of certain class group's correlating outliers can be found in the next section.

Table 5: Outlier positions per class group identified

	class_group	total_outliers	X1st_outlier_p	X2nd_outlier_p	X3rd_outlier_p	X3rd_final_outlier_p	X2nd_final_outlier_p	final_outlier_p
1	Technology	19	37	345	353	1933	2009	2071
2	Clothing	20	282	837	1048	1695	1723	1756
3	Household	395	252	387	643	1335	1336	1337
4	Luxury	440	142	171	184	789	790	791
5	Food	4	75	432	1149	432	1149	1408
6	Gifts	2287	213	216	218	2607	2608	2609
7	Sweets	4	942	1243	1294	1243	1294	1358

Table 6: Outlier x-bar values per class group identified

	class_group	total_outliers	X1st_outlier	X2nd_outlier	X3rd_outlier	X3rd_final_outlier	X2nd_final_outlier	final_outlier
1	Technology	19	17.500000	17.433333	17.366667	17.300000	17.600000	17.466667
2	Clothing	20	9.433333	9.433333	9.466667	8.533333	9.466667	8.433333
3	Household	395	42.233333	42.466667	42.500000	57.366667	54.566667	55.800000
4	Luxury	440	3.966667	3.966667	3.866667	3.400000	3.266667	3.600000
5	Food	4	2.266667	2.733333	2.733333	2.733333	2.733333	2.733333
6	Gifts	2287	10.233333	9.600000	9.900000	16.033333	16.933333	15.700000
7	Sweets	4	2.900000	3.000000	2.933333	3.000000	2.933333	2.900000

In real life, if there are many outliers and especially if they are close to each other or greatly outside the limits, they will need to be inspected and the process might need to stop to first inspect and correct potential mistakes. Only some outliers occasionally however are natural and may only be taken notice of and carefully kept track of but then it is not yet necessary to stop the entire process for inspection, since stopping the process and doing inspection takes time and causes hours of production lost. Unnecessary stops and inspection are therefore avoided.

Table 7 is a summary of the findings of most consecutive samples in the range of -0.3 and 0.4 sigma in each class type. (Summary of rule B findings). A more in-depth description per class group can be found below.

Table 7: : Table of most consecutive s-bar instances per class group

	Classes	last_num	last_std	consec_count
1	Clothing	223	0.5498918	4
2	Household	45	4.5638330	3
3	Food	756	0.2968084	5
4	Technology	1776	3.2026775	6
5	Sweets	94	0.5070926	4
6	Gifts	2477	1.5259657	7
7	Luxury	63	0.9536896	4

The process should be concerned when 7+ consecutives are found. Table 7 therefore indicates that class group 'Gifts' are to be inspected and a red flag is indicated with its high maximum number of consecutives of 7. The rest of the class groups that have high maximum of consecutives such as "Technology" should be kept an eye on but the rest with lower maximum consecutives does not indicate an error yet.

4.1.1 For class type "Technology":

4.1.1.1 For Rule A:

The control limits were calculated to be LCL = 17.77 and UCL = 22.97

The sample numbers that were out of control regarding outliers were:

37 345 353 398 483 613 643 688 760 1171 **1428 1438 1501 1677 1858 1872 1933** 2009 2071

And these sample numbers' correlating x-bar values are:

17.50 17.43 17.37 17.2 17.40 17.23 23.23 17.50 17.67 17.53 **17.30 17.76 17.60 17.30 16.80 17.40 17.3** 17.60 17.46667

These sample averages prove to correctly be outside the control limits.

By demonstration of them being out-of-control samples, the following section of the plot of sample means compared to the control limits for Technology was created to include the sample numbers listed that are bold and underlined:

Figure 10: Plot section of sample averages for Technology vs control limits

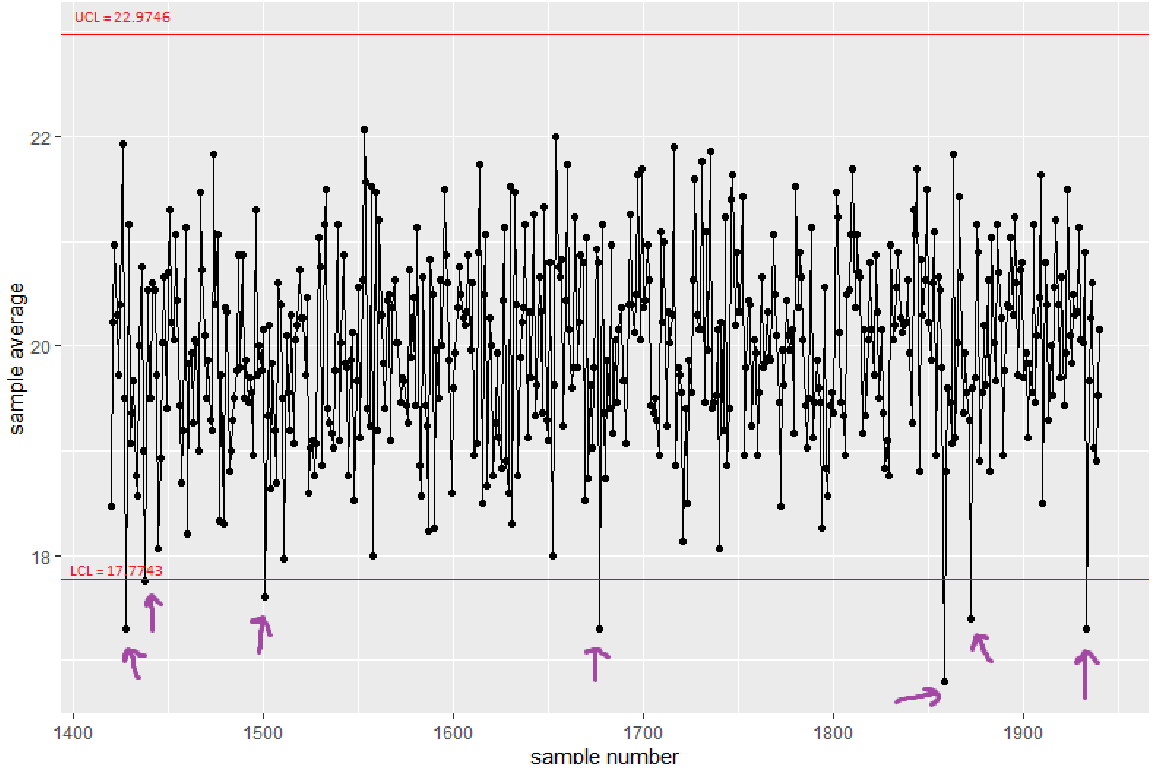


Figure 10 clearly indicates the outliers of sample numbers listed in bold and underlined, all having sample averages outside the control limits. Some of them (like the 2nd purple arrow indicated sample from the left) are borderlines but are still considered outliers since all instances outside the limits are considered outliers and errors.

Clear or frequent outliers as the 3 purple arrow indicated outliers on the right-hand side as well as the left-hand side (that seems to be repeating a pattern) will for example raise great concern and will need to be inspected to solve the problem and try to prevent or reduce future outliers and bad products in the system.

4.1.1.2 For rule B:

The most consecutive sample deviations in the range of between -0.3 and +0.4 sigma control limits for the class group were found to be:

last sample number of the maximum consecutives: 1776

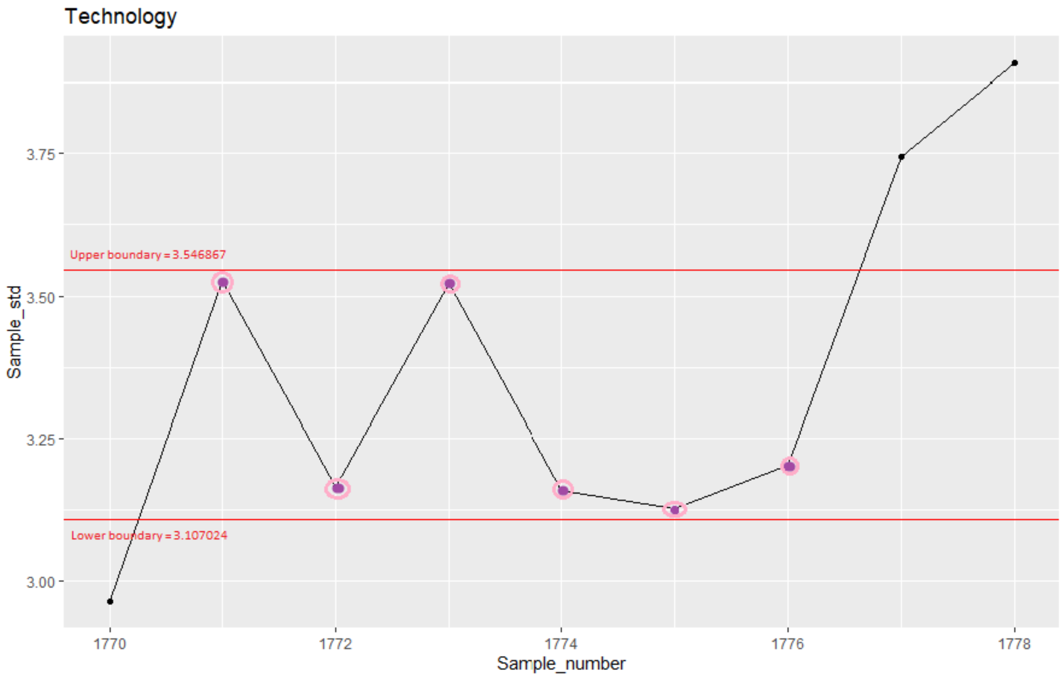
last sample standard deviation of the maximum consecutives: 3.2026775

Number of maximum consecutive standard deviations: 6

The number of consecutives in a system that raises alarm is 7+ but since 6 consecutives are pretty high as well and close to 7, we will use it as example as well.

The following graph illustrates how these 6 sample standard deviations of Technology indeed are consecutive in the given range.

Figure 11: Section of most consecutive Technology sample standard deviations



It is abnormal for production to have as many consecutive sample standard deviations and is quite a good indication of some error. The system will therefore need to be inspected and perhaps adjusted.

4.1.2 For class type "Clothing":

To demonstrate another example, rule A for clothes is also further described.

4.1.2.1 For Rule A:

The control limits were calculated to be LCL = 8.535 and UCL = 9.405

Sample numbers of the outliers outside the control limits are:

282 837 1048 1152 **1222 1271 1359 1479 1574 1587 1596** 1624 1626 1640

1644 1653 1677 1695 1723 1756

And their correlating sample averages are:

9.43 9.43 9.467 9.43 **9.53 9.47 8.467 9.43 8.3 8.53 8.50** 9.467 9.53 9.43

9.50 9.43 8.467 8.533 9.47 8.43

These sample averages prove to correctly be outside the control limits

By demonstration of them being out-of-control samples, the following section of the plot of sample means compared to the control limits for Clothing was created to include the sample numbers listed that are bold and underlined.

Figure 12: Plot section of sample averages for Clothing vs control limits

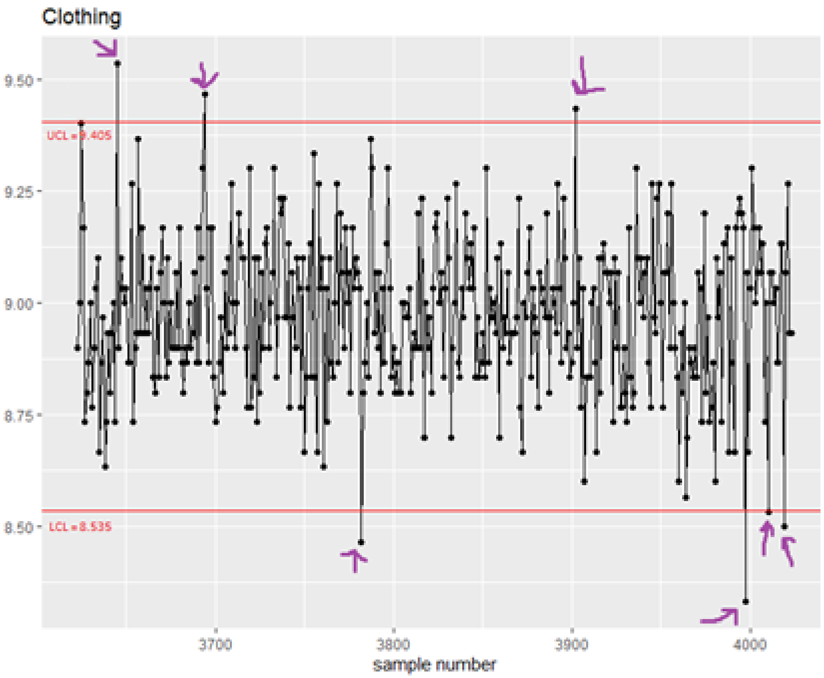


Figure 12 clearly indicates the outliers of sample numbers listed in bold and underlined, all having sample averages outside the control limits. These outlier raises concern for errors.

4.1.3 For class type "Gifts":

To demonstrate another example, rule B for gifts is also further described.

4.1.3.1 For Rule B:

The most consecutive sample deviations in the range of between -0.3 and +0.4 sigma control limits for the class group were found to be:

last sample number of the maximum consecutives:	2477
last sample standard deviation of the maximum consecutives:	1.5259657
Number of maximum consecutive standard deviations:	7

This is a lot of consecutive standard deviations and does indicate type II errors that will need to be inspected.

The following graph illustrates how these 7 sample standard deviations of Technology indeed are consecutive in the given range.

Figure 13: Section of most consecutive Gifts sample standard deviations

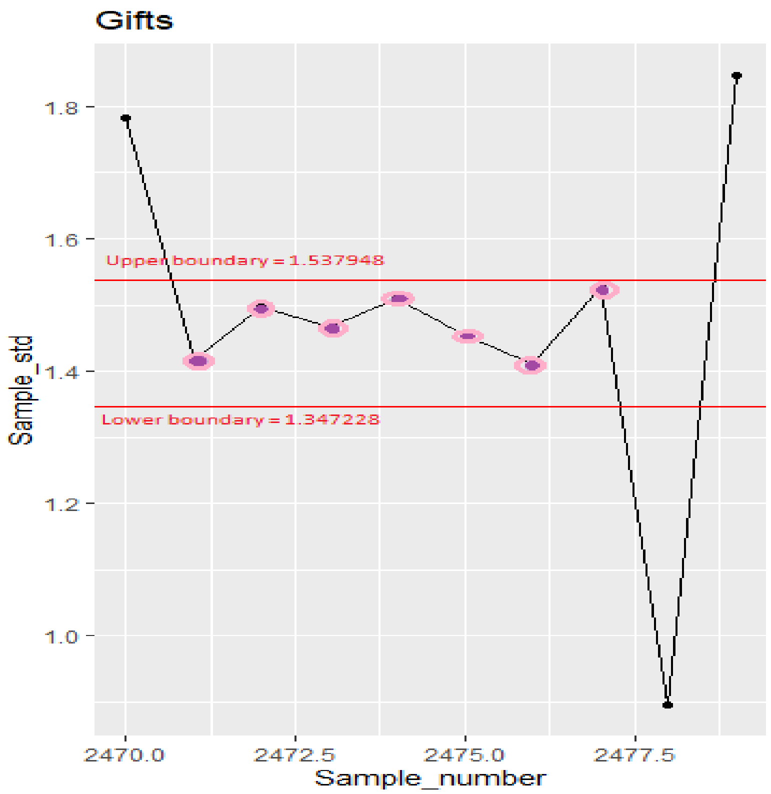


Figure 13 clearly indicates how these 7 consecutive sample standard deviations are an indication of problems in the system. Especially since we see the rest of the instances before and after the most consecutive few are widely spread, these great number of consecutive samples all in the given range shows suspicious production behaviour and is a clear enough indication that the production should be inspected and perhaps adjusted to prevent bad products.

4.2 Likelihood of making a Type I or Manufacturer's error

4.2.1 For rule A:

The chance of a sample to be outside the control limits will be 1 – the chance of being inside the control limits.

Therefore, as in our situation, the outer control limits are at 3 sigma and -3 sigma. The likelihood of making a Type I error for A is therefore:

= 1 – [pnorm(3) – pnorm(-3)]

= 1 – 0.9975

= 0.25%

4.2.2 For rule B:

The chance of having consecutive samples will therefore be the chance of having one such sample to the power of the number of consecutive samples.

For example, if the consecutive samples have a sample value of 12, with the chance of a sample having a value of 12 being 25%, and having 5 such consecutive examples, the likelihood of making a Type I error for B such as this will be:

= 0.25^5

= 0.09%

4.3 Centring the delivery process for best profit

For technology items' delivery times, the scenario of paying R329/item-late-hour for items with longer than 26-hour delivery times vs paying R2.5/item/hour to reduce the average time with one hour was solved by comparing the change in total cost per iteration of average hours reduced to find the number of hours to reduce the average time with to ensure the greatest profit and least cost.

Assuming the average delivery time can be reduced also with fractions of an hour.

The following was calculated per reduced delivery times set:

- The delivery times greater than 26 hours and their correlating total cost for being late
  - = the sum of [R329*(delivery time – 26)   per delivery item]
- The cost of reducing the average delivery time hours to the current iteration point
  - R2.5 * (total number of delivery instances) * average's hours decrease

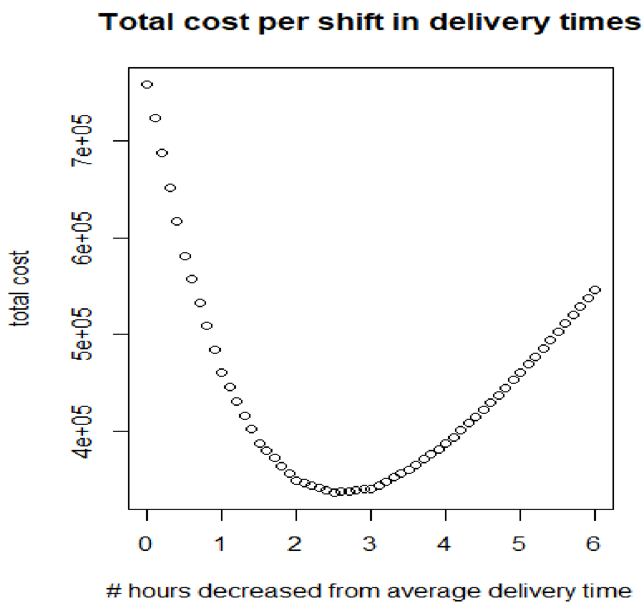
Table 8: Section of total cost, cost of being overtime, cost of shifting the average per iteration of average hours shifted

	hours_shift	C_tot	C_over	C_shift
1	0.0	758674.0	758674.0	0.0
2	0.1	723149.1	714061.6	9087.5
3	0.2	687624.2	669449.2	18175.0
4	0.3	652099.3	624836.8	27262.5
5	0.4	616574.4	580224.4	36350.0
6	0.5	581049.5	535612.0	45437.5
7	0.6	557039.6	502514.6	54525.0
8	0.7	533029.7	469417.2	63612.5
9	0.8	509019.8	436319.8	72700.0
10	0.9	485009.9	403222.4	81787.5
11	[illegible]	[illegible]	[illegible]	[illegible]

Table 8 illustrates a section of the table created and calculating the cost of the entries being overtime (more than 26 hours), the cost of shifting the average time with the fraction of an hour and the resulting total cost being the sum of the cost of being overtime and cost of shifting the average per iteration of fraction of an hour shifted.

Figure 14 shows the total costs (cost to shift the data + cost of data instances over 26 hours long) compared to the number of hours the average delivery times was shifted to be less hours:

Figure 14: Total cost per shift in delivery times



The solution to this problem is therefore to reduce or shift the average delivery times to the left with 2.5 hours to minimize total costs and therefore maximize the total profit. The delivery process was initially centred around 20.01 hours and **should therefore rather be centred at** (20.01-2.5) = **17.51 hours** of delivery time to minimize the total cost due to shifting the average hours and delivery times being longer than 26 hours**. This will move the average and data spread to be more centralized between limits of LSL=0 and USL=26 in this case but also ensure that the shift is not at the cost of profit or does not increase total costs.**

Figure 14 indicates how the total cost (per increase in number of hours the average delivery time was shifted), decreased until a point (2.5) but then increased again as the number of iterations still increased, creating a parabolic curve of loss. This is similar to the Taguchi Loss that also follows a parabolic curve of loss and both figure 14's curve and the Taguchi loss function indicate how loss is a parabolic factor and increases gradually the more it deviates from the ideal and that there is not suddenly loss the moment the instance falls outside the limits, such as with the goal post theorem. Figure 14 shows that this total cost (similar to loss) also gradually decreases to a minimum and then increases, therefore having a minimum loss at the point where customers (or in this case the production) will be most satisfied and thereafter, it will increase again in total cost.

4.4 Type II error for A

Since the upper control limit for Technology's sample averages is 22.975, the type II error (area of the new distribution that falls between the UCL = 22.975 and LCL = 17.77) where the sample values are actually faulty but we don't realize it, will therefore be the probability or area of the blue, new curve that falls below the UCL of 22.975.

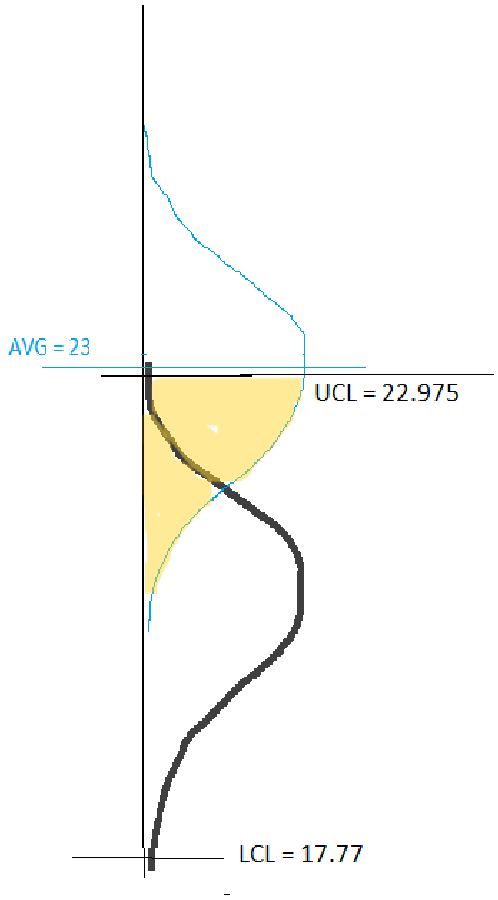
This probability was calculated using the normal distribution function (pnorm()) along with the the UCL value of 22.975, the mean of 23 and a standard deviation of (22.975-17.77)/6

P(type II error for A) = 48.85 %

There is in this example therefore 48.85% chance of making a type II error, meaning that the item can be faulty but we won't correctly identify it as faulty. By re-centring the process means, the process loss can therefore be reduced as shown in 4.3 but it might increase the type II error and reliability of the process.

The following illustration demonstrates this concept.

Figure 15: Visualization of Type II error for Technology rule A



Part 5: DOE and MANOVA

5.2 Hypotheses and MANOVA results

Based on Part 2,3,4's results and choosing reliable features, the hypothesis to test is the following:

H1 = The class groups will have a great and significant impact on the outcomes of Delivery Times and Price

H0 = The class groups will NOT have a great and significant impact on the outcomes of Delivery Times and Price

These hypotheses and tests are done to make sure of the reliability in making assumptions regarding the effect one factor (such as class types) has on another (such as price and delivery times) to use in further future predictions and planning. If the assumptions aren't proved and are perhaps unreliable, then our entire production planning will also be unreliable and lead to poor service.

For the class impact on delivery time and class impact on price, the p-value was both 2.2*e^-16 or = 2.475*10^-7, which is very close to a value of 0, **making it safe to reject the H0 hypothesis and to conclude that Class certainly has a great impact on both Delivery time and Price**. Our predictions and conclusions will therefore indeed be reliable when we associate the effect class groups has on delivery time and price.

Class has an effect factor on Delivery time and Price of 0.84, which is very big and indicates a great effect size.

The following are the Manova results that was created to get the p value and prove that Hypothesis H1 is true and that the Class groups indeed has a great effect on Delivery times as well as Price.

The graphs thereafter also further prove and visualizes the accuracy of our findings.

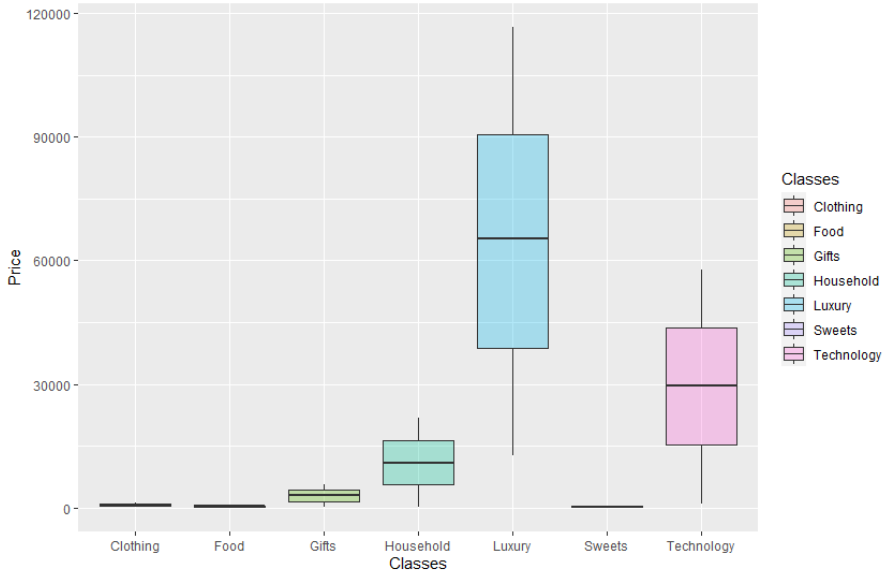
Table 9: results of summary(manova_data)

	Df Pillai	approx F	num Df	den Df	Pr(>F)
H1Class	6	1.6797	157291	12 359942	< 2.2e-16 ***
Residuals 179971					

Table 10: results of eta_squared(manova_data)

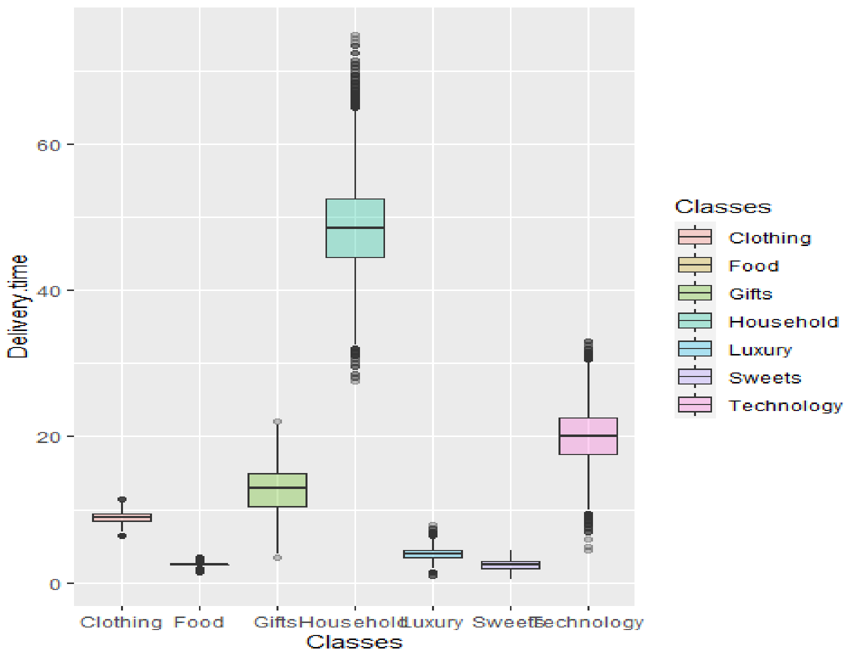
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1		
Parameter	Eta2 (partial)	95% CI
----------------	--------------	-----------
H1Class	0.84	[0.85, 1.00]

Figure 16: Boxplots of Price vs Class types



This figure therefore proves that the class types have an influence of Price. The luxury class gives much greater average prices than that of the rest and for example if the class type is clothing, the item can from the graph be sure to have a low Price.

Figure 17: Boxplots of Delivery Times vs Class types



This figure also proves that class types indeed have a great effect on delivery times. The class Household items will therefore have the greatest delivery times and if the class type is sweets, the expected delivery times can be very low.

These outcomes prove that the service regarding price and delivery time will definitely be influenced by the class group and that by doing so, the predicted data will be reliable, since we have proven that the class groups have a very strong effect on the delivery time and price and the outcomes are expected to be reliable.

Part 6: Reliability of the service and products

6.1.1 Problem 6 p 363

Question: 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.

L = k * (y - m)^2

45 = k * (0.04) ^2

k = 45/(0.04)^2

k = 28125

Thus **L = 28125*(y – 0.06)^2**

This Taguchi loss function will therefore be the parabolic function of loss and will in the end be minimized to find the minimum loss for the project and ensure high service performance and reliability.

6.1.2 Problem 7 p 363

Question: 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.*

L = k * (y - m)^2

35 = k * (0.04)^2

k = 21875

Thus **L = 21875* (y – 0.06)^2**

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

L = 21875* (y – 0.06)^2

L = 21875*(0.027)^2

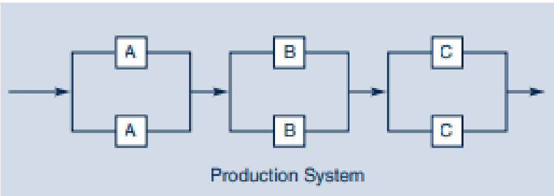
L = 15.95

The Taguchi loss function was determined to predict how much loss will be made if the process deviation target is adjusted such as in question b. This ensures that the company can plan ahead to ensure changes in production (such as deviation changes) will only be implemented if the change will bring acceptable losses to the company and to therefore ensure desired service levels are always attained and that the production remains reliable.

6.2 Problem 27 Chapter 7 p 363

Question: Magnaplex, Inc. has a complex manufacturing process, with three operations that are performed in series. Because of the nature of the process, machines frequently fall out of adjustment and must be repaired. To keep the system going, two identical machines are used at each stage; thus, if one fails, the other can be used while the first is repaired (see accompanying figure). B

Figure 18: Pr 27 p 363 given figure



The reliabilities of the machines are as follows:

Machine	Reliability	Machine	Reliability	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).

System reliability = 0.85*0.92*0.9

= **0.7038**

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

New system reliability= (1-(1-0.85)^2)*(1-(1-0.92)^2)*(1-(1-0.9)^2)

= 0.9775*0.9936*0.99

= 0.9615

Improvement = 100* (0.9615-0.7038)/0.7038

= **36.62%**

This proves that machine reliability can be improved by ensuring that a machine type has a backup machine parallel to it, to ensure that if there are problems with the one machine, the backup in parallel can easily continue with the work and avoid stops in production. This will ensure the required reliability is maintained since there were no setbacks with no stops and will also ensure that a high service level is maintained since the faulty machine have the opportunity to be fixed while in the meantime the backup machine continues in producing products in good quality.

6.3 Given Question

6.3.1 Initial question

Question: 21 vehicles, 19 operatable to be reliable

Tot 1560 days *- only 20 available = 190 days*

- only 19 available = 22 days

- only 18 available = 3 days

- only 17 available = 1 day

21 drivers, 8 hours shift per day

Tot 1560 days *- 20 drivers available = 95 days*

- 19 drivers available = 6 days

- 18 drivers available = 1 day

<u>How many days per year expect reliable delivery times</u>?

To determine accurate probabilities, the resulting days of no vehicles or drivers being unavailable was added to the dataset (tot days – sum of days vehicles or drivers are unavailable).

For the vehicles available:

The probability of 0, 1, 2, 3, 4 vehicles being unavailable was calculated using

P = (correlating number of days)/(total days)

Equation 1: binomial distribution

$$P_x = \binom{n}{x} p^x q^{n-x}$$

P = binomial probability

x = number of times for a specific outcome within n trials

$\binom{n}{x}$ = number of combinations

p = probability of success on a single trial

q = probability of failure on a single trial

n = number of trials

For each number of vehicles unavailable, the probability (pi or small p in the binomial equation) of success on a single trial was determined by using a trial and error method to determine the appropriate p-value to use in the dbinom() function or equation to result in the correct correlating P probability previously calculated.

Table 11: The trial and error resulting p-values per number of vehicles unavailable

	num.cars.down	sample.Prob	pi.val	pi_given_Prob
1	0	0.8615384615	0.0070720	0.8615349738
2	1	0.1217948718	0.1501447	0.1217949453
3	2	0.0141025641	0.3200946	0.0141025390
4	3	0.0019230769	0.4609620	0.0019230736
5	4	0.0006410256	0.5527245	0.0006410254

To find a final p-value, the weighted average (dependant on the total number of out of the total sample days) was used to determine a **final p-value for probability of a vehicle being unavailable of = 0.006026916.**

To determine the actual (more accurate) probabilities (P) of the number of vehicles being unavailable, the binomial distribution (dbinom()) was used as well as the new weighted average p-value.

The probabilities of only 0,1,2 vehicles unavailable was determined and added together to find the final probability that sufficient vehicles will be available (at least 19 of 21) and therefore the delivery being reliable being $\boldsymbol{P_{reliable\ vehicle} = 0.0.9997316}$.

The same was done for the number of employees unavailable:

Table 12: The trial and error resulting p-values per number of employees unavailable

	num.emp.down	sample.Prob	pi.val	pi_given_Prob
1	0	0.9346153846	0.00321485	0.9346149961
2	1	0.0608974359	0.18834500	0.0608979618
3	2	0.0038461538	0.37563900	0.0038462741
4	3	0.0006410256	0.49964250	0.0006410283

The final weighted average p-value of number of employees unavailable proved to be

$\boldsymbol{p_{employee}}$ **= 0.004059856**

The probability of enough employees being available and therefore having reliable deliveries used only the sum of the binomial probabilities of 0,1,2 employees unavailable as well (at least 19 of 21 available).

The resulting probability was $\boldsymbol{P_{reliable\ employee} = 0.999157.}$

The final Probability of days having reliable delivery times were calculated by multiplying the $\boldsymbol{P_{reliable\ vehicle}}$ **and** $\boldsymbol{P_{reliable\ employee}}$ values, since both play a role in having reliable deliveries on the day.

The resulting Probability of being reliable on a day was $\boldsymbol{P_{reliable} = 0.9996474}$

The final resulting number of days in a year expected to have reliable deliveries are therefore

#Days = 365*$P_{reliable}$

#Days = 364.8713

#Days = 364 (rounded down)

6.3.2 Increase the number of vehicles to 22

<u>If we increased our numbers to 22 vehicles in total. How many days per year should we expect reliable delivery times?</u>

Since the total number of vehicles available does not change the determined probability p-value, the new probability of a having reliable deliveries due to the effect of available vehicles was determined by using the sum of binomial probabilities of 0,1,2 and 3 vehicles being unavailable (at least 19 but out of 22 available).

This resulted in a new $\boldsymbol{P_{reliable\ vehicle\ 22} = 0.9999912}$

The resulting probability of having reliable deliveries due to the effect of available employees however did not change since the number of employees stayed the same. $\boldsymbol{P_{reliable\ employee} = 0.9999157}$

The resulting Probability of being reliable on a day was $\boldsymbol{P_{reliable\ 22} = 0.9999069}$

The final resulting number of days in a year expected to have reliable deliveries for 22 vehicles are therefore

#Days = 365*$P_{reliable}$

#Days = 364.966

#Days = 364 (rounded down)

The increase in total number of cars therefore increased the probability of giving reliable service, even though the difference did not change the total number of full days of being reliable, since the change was not big enough.

Conclusion:

In this report, the dataset proved to indeed have problem factors such as outliers and consecutive instances in all class groups, that re-centring the average delivery times can indeed reduce the total loss or cost and are quite similar to that of the Taguchi loss function, but when the process mean is re-centred, it can increase the chance of having a type II error and reduce the system reliability. Hypotheses can be identified and tested with manova to be proven and to make decisions. The Taguchi loss function can be used to identify and predict losses according to planned changes made in the system and maintain the required service level and reliability. The system reliability can also be improved by parallel machine positioning. Finally, the increase in total vehicles or employees proved to increase the probability of providing a reliable service.

In conclusion, the report content proved that statistics and data analytics can be used to effectively analyse historical trends, identify quality problems to be adjusted and can accurately predict what changes need to be made and how that could potentially influence the future performance of the business.

References:

Evans, J.R, Lindsay W.M. 2019. *Managing for Quality and Performance Excellence.* Boston USA: Cengage Learning Inc.

Potential (within) capability for Normal Capability Analysis. [S.a.]. [Online]. Available: https://support.minitab.com/en-us/minitab/21/help-and-how-to/quality-and-process-improvement/capability-analysis/how-to/capability-analysis/normal-capability-analysis/interpret-the-results/all-statistics-and-graphs/potential-within-capability/ [2022, September 21]

Xbar-s Control Charts: Part 1 | BPI Consulting - SPC for Excel. [S.a.]. [Online]. Available: https://www.spcforexcel.com/knowledge/variable-control-charts/xbar-s-control-charts-part-1

[2022, September 30]