ECSA GA4 Project Report

A.G. Jones

23767278

October 2022

Abstract

In this report, the focus is statistical process control and analysis of a given dataset. In section one of the report, an explanation is given on how the data was wrangled/sorted into a valid data set and an invalid data set. The data is also sorted in chronological order.

In section two, descriptive statistics and graphs are given for the data set. This includes explanations of the information conveyed by the graphs.

In section three, the data is sampled to use in following calculations. The control limits for the feature class and its categories were calculated. Statistical process control charts are also given and commented on.

In section four, some exploratory analysis is done on the data regarding the results of section three. This includes tabulating the xbar and sample standard deviations which are respectively out of the control limits and in the control limits. Some other statistical calculations are also done, such as calculating the probabilities of making type I and type II errors.

In section five, a Manova analysis is done and commented on.

In section six, some questions are done which are unrelated to the rest of the data in the project.

Contents

List of tables

List of figures

Introduction

The aim of this report was to showcase data analytics skills and understanding of statistical properties of data. The project was formulated by ECSA and is crucial to the curriculum of industrial engineering students.

The information given in the report is structured in a similar way to the project brief. Results are given per section, with appropriate comments and values being presented. Where necessary, simple graphs and tables are given to illustrate points or explain topics.

The instructions and questions in the project brief were followed and answered as understood by the student. All assumptions deemed necessary were stated.

1. Data Wrangling

For part one of the project, the data had to be formatted (wrangled) in such a way that any instances that included missing ("NA") values or impossible values (such as negative sales price) are removed. The data was first ordered by year, month, day. Next, it was split into valid data, as mentioned above, and invalid data containing maissing values and impossible values. The original index for these new datasets were kept intact as secondary indexes, while a new primary index was added.

	Primary_Key	X	ID	AGE	Class	Price	Year	Month	Day	Delivery.time	Why.Bought
463	1	463	47101	50	Clothing	1030.86	2021	1	1	9.0	Recommended
2627	2	2627	88087	21	Clothing	428.03	2021	1	1	10.0	Recommended
3374	3	3374	25418	68	Household	13184.41	2021	1	1	48.5	Website
5288	4	5288	13566	94	Household	7021.90	2021	1	1	42.0	Recommended
8182	5	8182	84692	35	Clothing	475.18	2021	1	1	9.0	Recommended
9272	6	9272	46305	72	Clothing	580.98	2021	1	1	8.5	Random
9712	7	9712	92105	45	Household	6877.00	2021	1	1	43.0	Recommended
12163	8	12163	21614	27	Clothing	513.13	2021	1	1	9.5	Recommended
12195	9	12195	12174	56	Household	14538.64	2021	1	1	41.5	EMail
20004	10	20004	84558	74	Food	255.41	2021	1	1	2.0	Recommended
20509	11	20509	15630	32	Clothing	164.56	2021	1	1	9.0	Recommended
21970	12	21970	81216	87	Clothing	173.76	2021	1	1	10.0	Recommended
27161	13	27161	56240	45	Household	17681.94	2021	1	1	45.5	Website
27638	14	27638	24396	30	Clothing	1018.21	2021	1	1	8.5	Recommended
30778	15	30778	12235	28	Technology	21096.86	2021	1	1	15.0	Website
34277	16	34277	30290	43	Household	10573.67	2021	1	1	51.0	Recommended
34950	17	34950	40035	77	Household	16548.61	2021	1	1	51.5	Recommended
35153	18	35153	36435	53	Technology	23304.75	2021	1	1	14.0	Browsing
37187	19	37187	49974	67	Sweets	332.46	2021	1	1	2.5	Recommended
42139	20	42139	36292	75	Food	205.96	2021	1	1	3.0	Recommended
43139	21	43139	39202	34	Clothing	353.28	2021	1	1	9.0	Recommended
44379	22	44379	92277	82	Food	219.76	2021	1	1	2.5	Recommended
45422	23	45422	12068	56	Gifts	1320.89	2021	1	1	6.5	Website

Table 1: Showcase of Valid data set

As can be seen above, the primary index runs from one to the number of instances in the dataset. The data itself does not containing any missing values or impossible values. The observations (sales) are also ordered by date, from oldest to newest, as is also clear from the table shown.

	Prim_key	X	ID	AGE	Class	Price	Year	Month	Day	Delivery.time	Why.Bought
98765	1	98765	64288	25	Clothing	NA	2021	1	24	8.5	Browsing
54321	2	54321	62209	34	Clothing	NA	2021	3	24	9.5	Recommended
34567	3	34567	18748	48	Clothing	NA	2021	4	9	8.0	Recommended
155555	4	155555	33583	56	Gifts	NA	2022	12	9	10.0	Recommended
144443	5	144443	37737	81	Food	-588.8	2022	12	10	2.5	Recommended
177777	6	177777	68698	30	Food	NA	2023	8	14	2.5	Recommended
16320	7	16320	44142	82	Household	-588.8	2023	10	2	48.0	EMail
56789	8	56789	63849	51	Gifts	NA	2024	5	3	10.5	Website
19998	9	19998	68743	45	Household	-588.8	2024	7	16	45.5	Recommended
87654	10	87654	40983	33	Food	NA	2024	8	27	2.0	Recommended
166666	11	166666	60188	37	Technology	NA	2024	10	9	21.5	Website
19541	12	19541	71169	42	Technology	NA	2025	1	19	20.5	Recommended
19999	13	19999	67228	89	Gifts	NA	2026	2	4	15.0	Recommended
155554	14	155554	36599	29	Luxury	-588.8	2026	4	14	3.5	Recommended
12345	15	12345	18973	93	Gifts	NA	2026	6	11	15.5	Website
23456	16	23456	88622	71	Food	NA	2027	4	18	2.5	Random
65432	17	65432	51904	31	Gifts	NA	2027	7	24	14.5	Recommended
144444	18	144444	70761	70	Food	NA	2027	9	28	2.5	Recommended
19540	19	19540	65689	96	Sweets	-588.8	2028	4	7	3.0	Random
76543	20	76543	79732	71	Food	NA	2028	9	24	2.5	Recommended
16321	21	16321	81959	43	Technology	NA	2029	9	6	22.0	Recommended
45678	22	45678	89095	65	Sweets	NA	2029	11	6	2.0	Recommended

Table 2: Table showcasing Invalid data set

The invalid dataset clearly contains only NA and impossible values. The data is also ordered in the same way as the valid dataset above.

The way in which the data was split was done in a repeatable manner, which means that it can quickly and efficiently be done again when new data is presented. The resulting valid dataset is used to generate the output presented in the following sections.

2. Section two

2.1. Descriptive statistics

In this section the data was analysed, and graphs were used to get a better understanding of the features of the data set. The main type of graphs used are a histogram for continuous features, while bar plots are used for categorical features. Below the graphs will follow, after which the key takeaways are given. First, a table is presented which gives some description of the features in the dataset.

Feature	Type	Description
Age	Numeric (discrete)	Age of customer who made purchase
Class	Character (categorical)	Type of item sold
Price	Numeric (continuous)	How much item cost
Year	Numeric (discrete)	Year in which sale was made
Month	Numeric (discrete)	Month of the year sale was made
Day	Numeric (discrete)	Day of the month on which sale was made
Delivery Time	Numeric (continuous)	Time from order to delivery
Why.Bought	Character (categorical)	Reason for making purchase

Table 3: Description of features in dataset

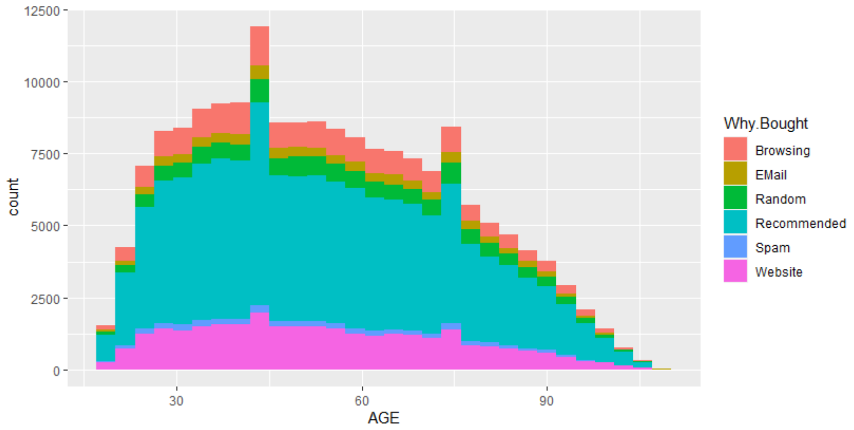

Figure 1: Graph showing distribution of AGE variable

For the age variable, the data follows a right skewed normal distribution, with some random spikes at approximately 45 and 75 years. The age of the buyer seems to have no correlation with their reason for purchasing, with the reasons for buying being proportionate to the count in their variance.

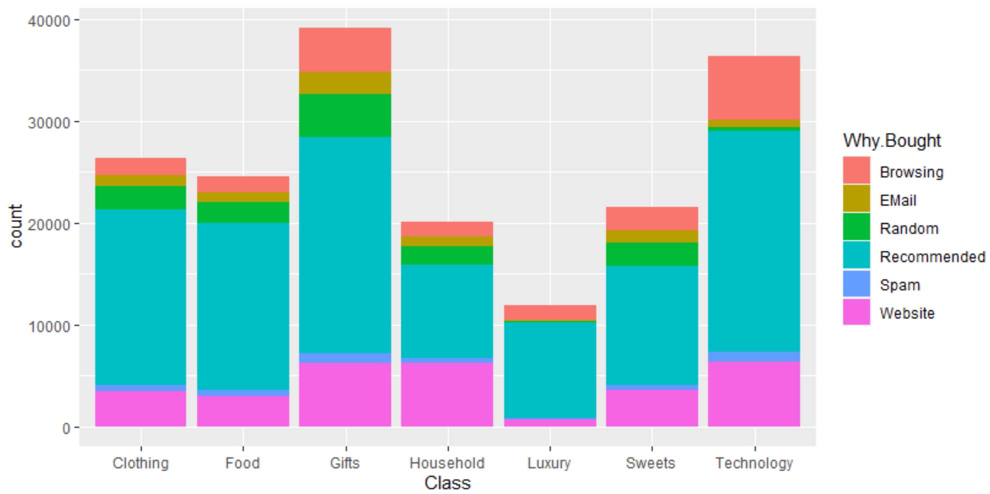

Figure 2: Graph showing distribution of Class variable.

The class variable is categorical and follows no distribution. The mode of the variable is Gifts, with the smallest count being the class Luxury. Technology is a close second to gifts. The reasons for buying do seem to be somewhat correlated with the class of purchase, with random buying being less prevalent among luxury and technology purchases. Luxury especially has a high proportion of recommended buys and low proportion of spam buys.

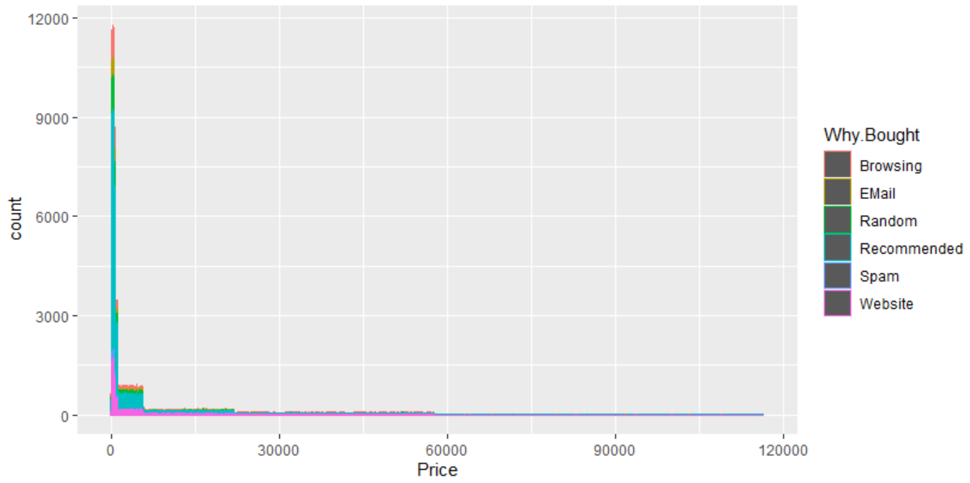

Figure 3: Graph showing distribution of Price variable

The variable price follows an exponential distribution, which means most sales are of low price. The data has a long tail, with sales of up to ~ 115000 still occurring. The reasons for buying are somewhat correlated with the price, with less random buys happening at higher prices (this can likely be linked to the class variable).

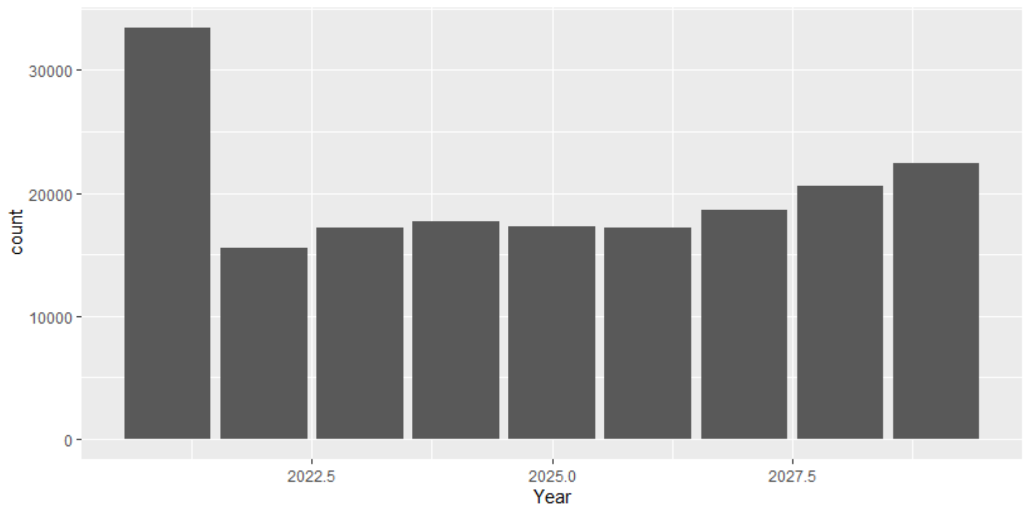

Figure 4: Graph showing number of sales per year

The distribution of sales by year shows that sales were highest in 2021, with sales increasing slightly year by year from 2022 to 2029. This leads to the conclusion that the company is growing at least somewhat over the period given.

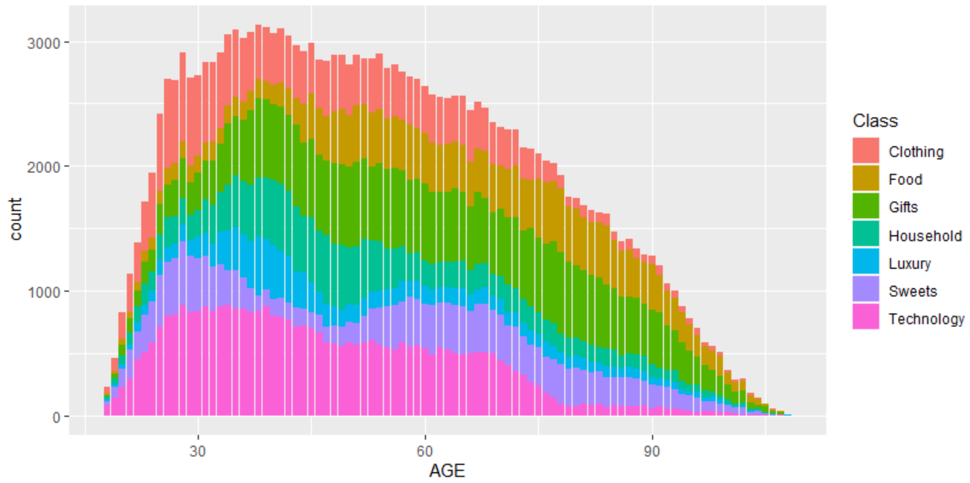

Figure 5: Graph showing distribution of class sales among ages

When analysing the distribution of class of sale by age group, it is interesting to note that younger age groups have a higher tendency to buy clothes and technology (percentage wise), while older age groups have a higher tendency to buy food and gifts. The other classes seem to be spread out proportionately equally over the age groups.

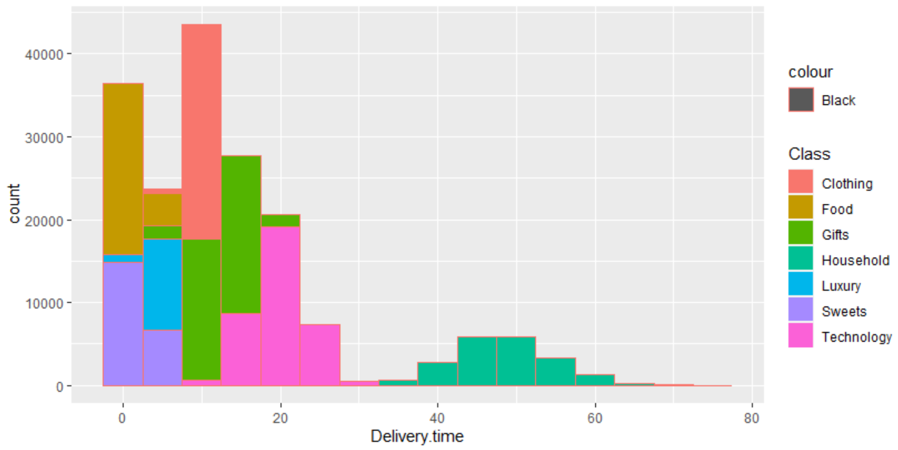

Figure 6: Graph showing distribution of delivery times

The distribution of the delivery time variable follows a multimodal distribution with a small peak at Delivery time ~ 50 hours. Items such as food and sweets have the shortest delivery time, with household products having the longest delivery time. This is as one would expect, as for food and sweet items to be relevant, they must be delivered while still fresh and in time to satisfy the customer's possible immediate need.

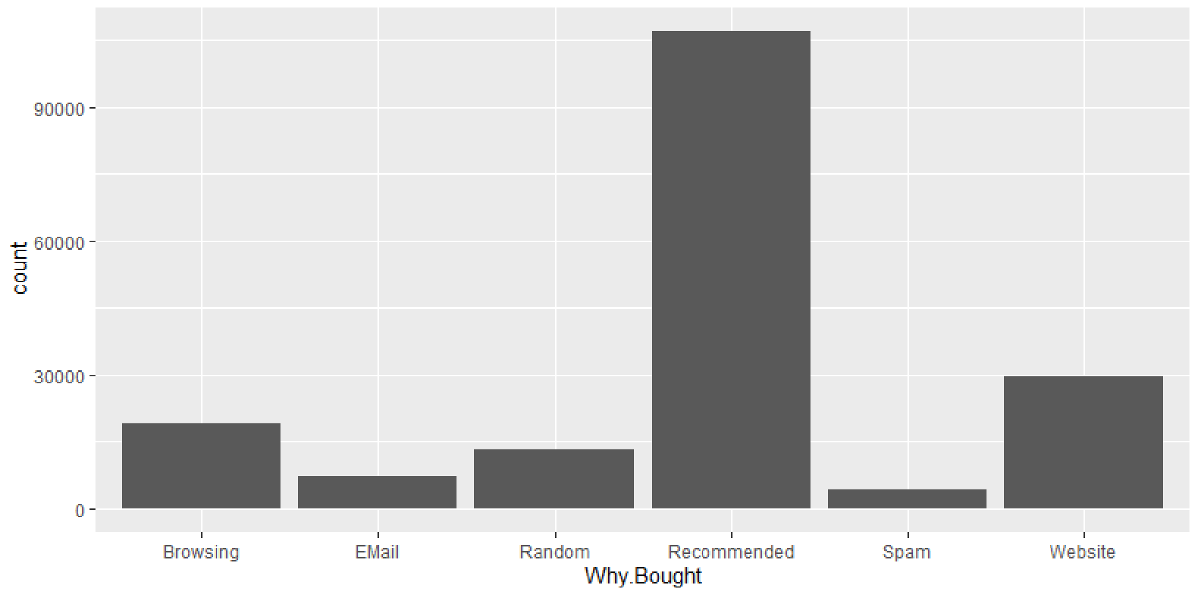

Figure 7: Graph showing distribution of Why.Bought variable

The Why.Bought variable is categorical, as such, it does not follow any distribution. The mode is "recommended", the least occuring category is "spam". The takeaway from this graph is that most customers are buying because of word-of-mouth. Next in line is the website and browsing, which are closely connected. Any marketing campaigns would thus be less effective than simply having good quality products which advertise themselves.

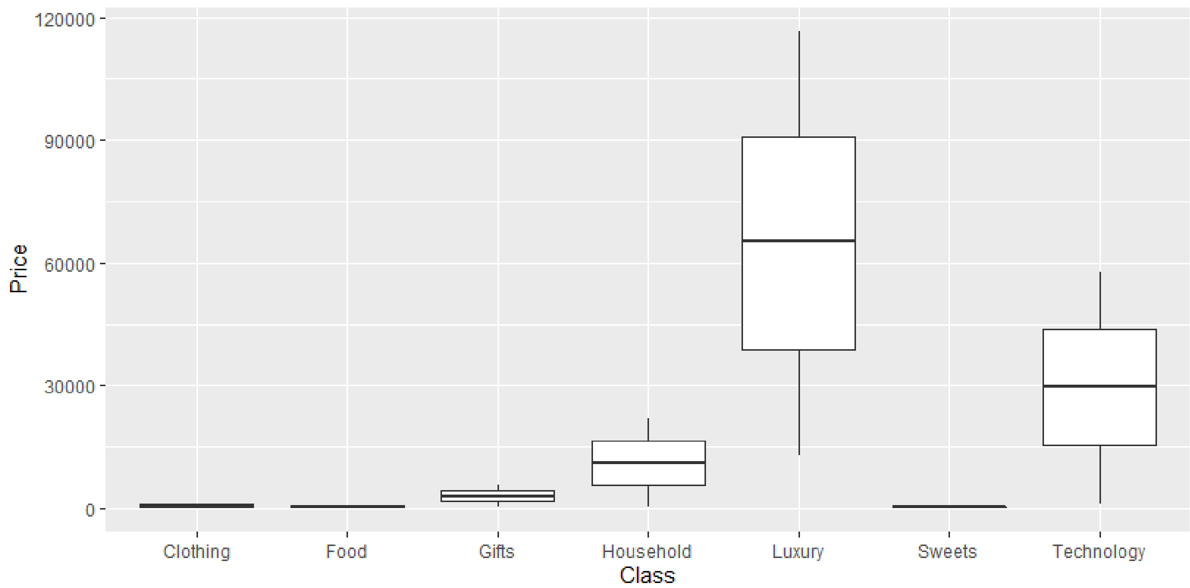

Figure 8: Graph showing boxplots of price distribution per class of item bought

This graph is the Price as a function of the class of purchase. The class with the highest mean price is luxury, while the class with the lowest mean price is food. The results of this chart, in combination with the sales per class bar chart, indicates that technology is producing the highest revenue out of all the classes.

2.2. Process Capability Indices

The process capability indices are values which show whether a process is under control or not. The calculation of these values and result of these calculations will be given below.

```
USL <- 24;
LSL <- 0;
SD <- sd(Valid[which(Valid["Class"] == "Technology"), "Delivery.time"])
mean <- mean(Valid[which(Valid["Class"] == "Technology"), "Delivery.time"])

Cp <- round((USL - LSL)/(6*SD), 3)
Cpu <- round((USL-mean)/(3*SD), 3)
Cpl <- round((mean - LSL)/(3*SD), 3)
Cpk <- round(min(Cpu, Cpl), 3)
```

Figure 9: Calculation of CP values

An LSL of 0 is logical as the delivery time cannot be less than zero (delivered before the order is placed).

```
Cp     1.142
Cpk    0.38
Cpl    1.905
Cpu    0.38
```

Figure 10: Process Capability values

From the values above, it can be said that the process is not under statistical control. Most companies would require a Cp value of above 1.33 and a Cpk value of between 1 and 3 (Process Capability | Quality-One, 2022).

3. Section three

In this section, the control limits for the delivery times of each item class are calculated. First, samples are drawn for each class, with oldest data being first and newer data being later. They are then used for the calculations mentioned above. The values are tabulated for both the sample means (xbar) and for the sample standard deviations. For the sample means, the first thirty are plotted to check whether they are under statistical control. For the sample standard deviations, all the samples are plotted.

3.1. Tables containing control limits for SPC charts.

To calculate the values in the tables below, a standard i-j for loop was utilized to loop through the samples in each class and use them to calculate the required control chart values.

	Class	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
1	Clothing	9.404934	9.259956	9.114978	8.970000	8.825022	8.680044	8.535066
2	Household	50.248328	49.019626	47.790924	46.562222	45.333520	44.104818	42.876117
3	Food	2.709458	2.636305	2.563153	2.490000	2.416847	2.343695	2.270542
4	Technology	22.974616	22.107892	21.241168	20.374444	19.507721	18.640997	17.774273
5	Sweets	2.897042	2.757287	2.617532	2.477778	2.338023	2.198269	2.058514
6	Gifts	9.488565	9.112747	8.736929	8.361111	7.985293	7.609475	7.233658
7	Luxury	5.493965	5.241162	4.988359	4.735556	4.482752	4.229949	3.977146

Table 4: X-chart containing control limits for sample means of delivery times of each class

	Class	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
1	Clothing	0.8665596	0.7614552	0.6563509	0.5512465	0.4461422	0.3410379	0.2359335
2	Household	7.3441801	6.4534101	5.5626402	4.6718703	3.7811003	2.8903304	1.9995605
3	Food	0.4372466	0.3842133	0.3311800	0.2781467	0.2251134	0.1720801	0.1190468
4	Technology	5.1805697	4.5522224	3.9238751	3.2955278	2.6671805	2.0388332	1.4104859
5	Sweets	0.8353391	0.7340215	0.6327039	0.5313862	0.4300686	0.3287509	0.2274333
6	Gifts	2.2463333	1.9738773	1.7014213	1.4289652	1.1565092	0.8840532	0.6115971
7	Luxury	1.5110518	1.3277775	1.1445032	0.9612289	0.7779546	0.5946803	0.4114060

Table 5: S-chart table containing control limits for sample standard deviations of delivery times for each class

3.2. SPC charts (statistical process control)

Two types of control charts were generated to monitor whether the processes are under statistical control or not. The first type is the xbar charts, which plot the sample means of the first thirty samples in each class. The second is the s charts, which plot the sample standard deviations for all the samples in each class.

For each chart, the green line indicates the CL (target) value, the red lines are the upper and lower control limits, and the blue lines are the upper and lower sigma and two sigma lines. To monitor if a process is under statistical control, you first look at the s SPC chart. If the s chart is within the control limits, you can then look at the xbar chart to see whether the process is under control or not. If the s chart is out of the control limits, the process is not under control, and you can disregard the xbar chart. It is thus important to use these charts in tandem to monitor process performance.

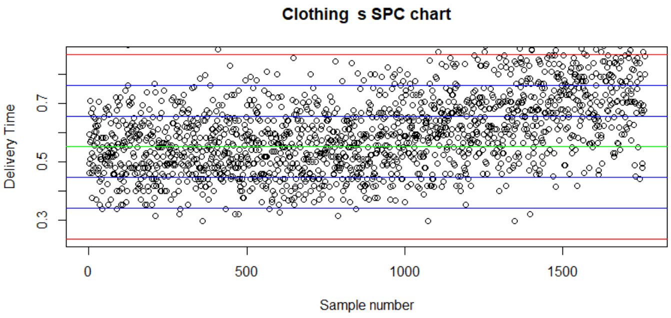

Figure 11: Clothing s SPC chart

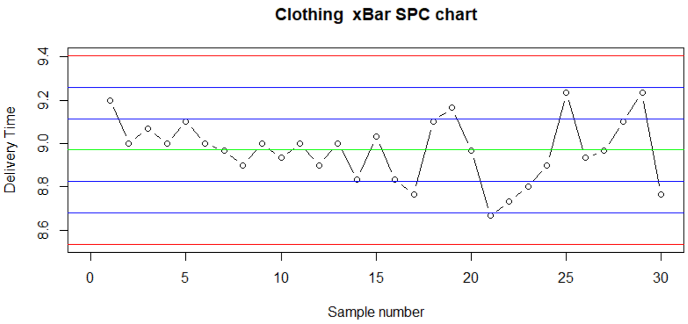

Figure 12: Clothing Xbar SPC chart

The s SPC chart is under control for the first 1000 or so samples, with the odd sample being out of the control limits. After this however the process tends toward being out of control. The xbar chart for the first thirty samples show that the process is under control, it is however tending to spiralling out of control.

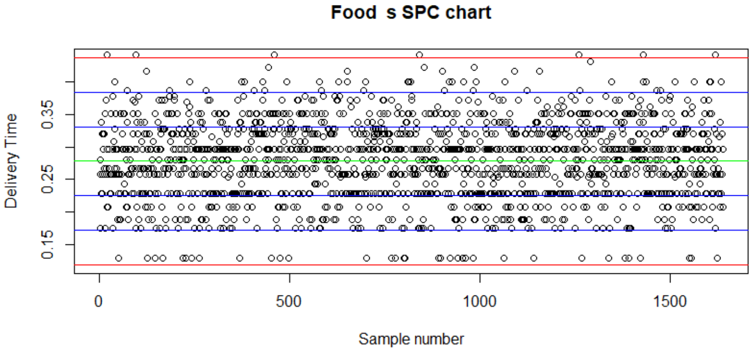

Figure 13: Food s SPC chart

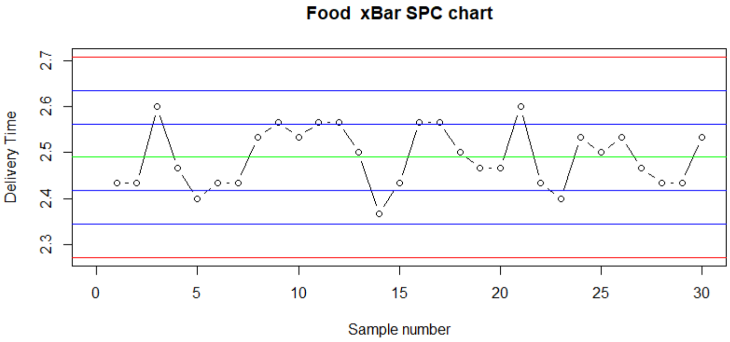

Figure 14: Food Xbar SPC chart

The s SPC chart indicates that the process remains relatively stable over time, with the odd outlier that is out of the control limits. The xbar chart for the first thirty samples seems to oscillate around the CL line, with no indication that the process as an underlying trend to become unstable.

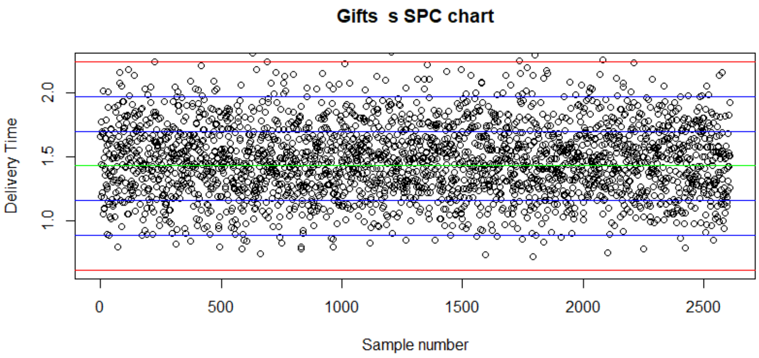

Figure 15: Gifts s SPC chart

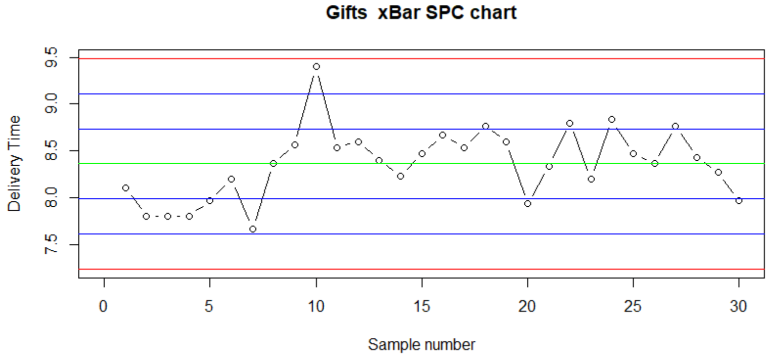

Figure 16: Gifts Xbar SPC chart

The s SPC chart indicates no upwards or downwards trend in the standard deviations of the samples. There is the odd outlier, which is outside the control limits, but there is no clear indication that the process is not under statistical control. The xbar SPC chart seems to indicate that over the first thirty samples, the process has become more stable, with smaller oscillations occurring between later samples than at the start.

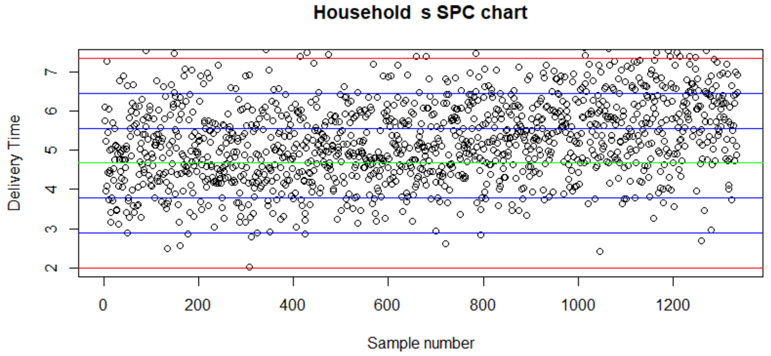

Figure 17: Household s SPC chart

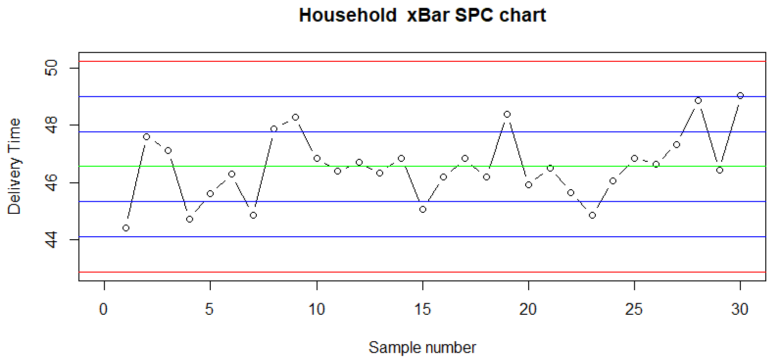

Figure 18: Household Xbar SPC chart

For the delivery times of the household items, the s SPC chart has an upward trend over time. There are quite a few samples with standard deviation out of the control limits. This means that even though the xbar chart shows no indication of the process being out of control, many instances within the samples might fall out of the control limits due to the high standard deviation.

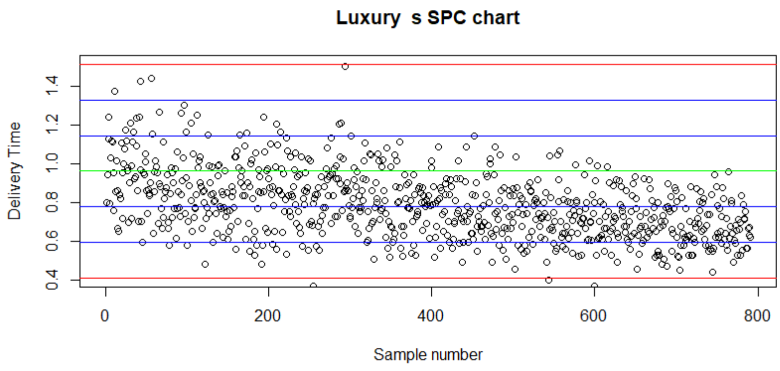

Figure 19: Luxury s SPC chart

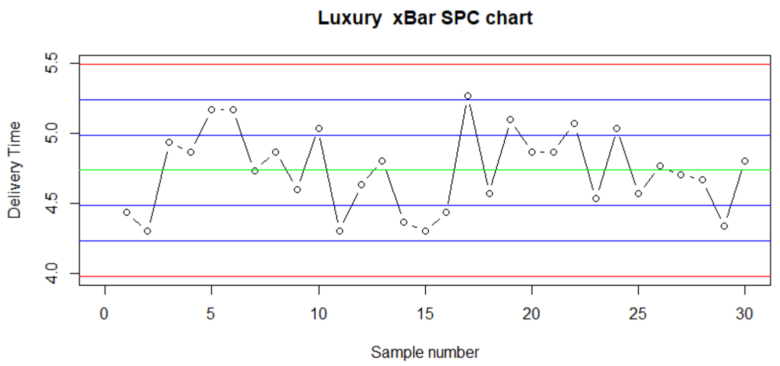

Figure 20: Luxury Xbar SPC chart

For the delivery times of the luxury items, the s SPC chart indicates a clear downward trend over time. There are, however, not many samples standard deviations out of the control limit yet. The xbar SPC chart shows oscillations around the CL line with high amplitudes, but there is no indication of a trend towards being out of the control limits. The concern for this variable should be that the standard deviations are clearly trending towards being under the lower control limit.

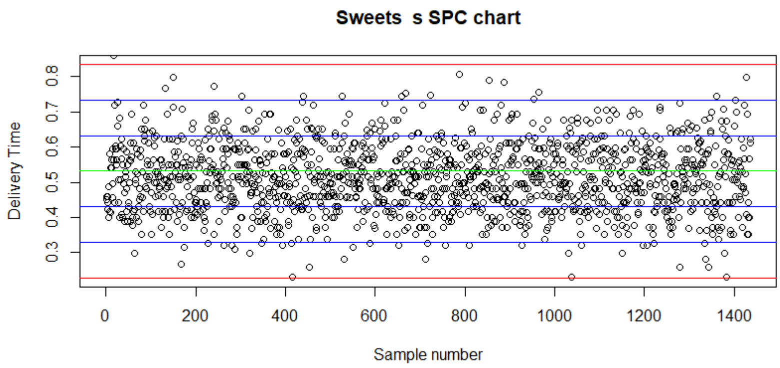

Figure 21: Sweets s SPC chart

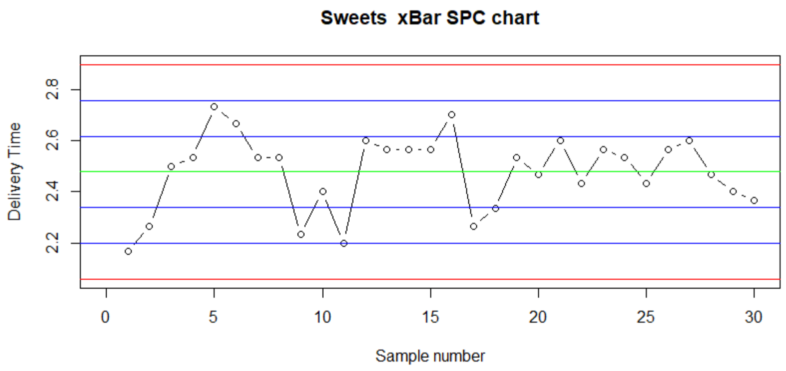

Figure 22: Sweets Xbar SPC chart

For the delivery times of the sweets item class, the s SPC chart shows no clear trend in any direction and seems to remain spread out rather evenly around the CL line. There are a couple sample standard deviations out of the control limits, but not enough to indicate an out-of-control process. The xbar SPC chart for the first thirty samples indicates that the process has become more stable towards the later samples regarding sample means. This is deduced from the smaller oscillations around the CL line versus at the start.

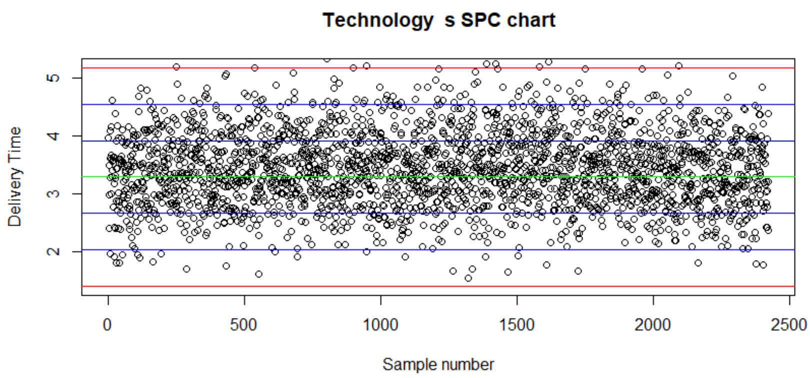

Figure 23: Technology s SPC chart

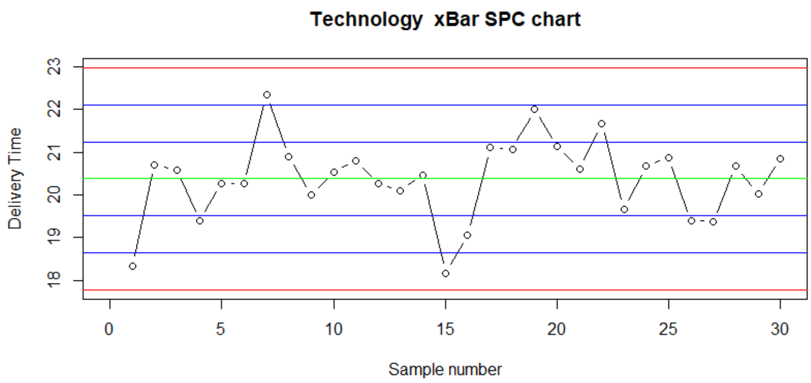

Figure 24: Technology Xbar SPC chart

For the delivery times of the technology class items, the s SPC chart has a slight negative parabolic curve. There are multiple sample standard deviations which are out of the control limits, but overall, the trend does not seem to indicate that the process is moving towards being out-of-control. The xbar

SPC chart also shows no trend towards fluctuating out of control, even though it has large oscillations at certain times.

4. Section four

In this section, some statistical analysis is done on the results produced in section three. This includes the indexes and total count of all xbar values of samples that are out of the control limits, as well as the count of the maximum consecutive sample standard deviations which are out of the control limits.

4.1. Control Limits

4.1.1. A

Table containing first and last three indexes of xbar values that were out of bounds. It also has the total count of out of bounds values.

	Class	Indexes	Total_Count
1	Clothing	282,1222,1337,1713,1723,1756	15
2	Household	252,387,629,1335,1336,1337	396
3	Food	75,432,1149,1408	4
4	Technology	37,398,483,2009,2071,2218	20
5	Sweets	942,1243,1294,1358	4
6	Gifts	213,216,218,2608,2609,2610	2291
7	Luxury	142,171,184,789,790,791	434

Table 6: Xbar table of out of bounds values

The classes food and sweets have few values outside the control limits, while gifts has a startling amount that are outside. This indicates that the delivery times for gifts is not under statistical control, and this must be addressed immediately.

4.1.2. B

Table containing information regarding s values which are out of the control limits.

	Class	Maximum consecutive amount	Last consecutive index	Total within count
1	Clothing	14	1758	1027
2	Household	20	1335	767
3	Food	9	1638	932
4	Technology	18	2421	1666
5	Sweets	17	1437	1025
6	Gifts	17	2610	1792
7	Luxury	9	780	341

Table 7: S table containing information for question 4.1B

The classes with the least amount of consecutive standard deviations within the control limits are food and luxury. This could be taken to mean that they are not statistically as under control as the other classes. If the assumption is made that the last sample standard deviation value that was within the control limits is close to the end (i.e., total amount of samples), then it is also clear that these classes have a low proportion of standard deviations within the control limits. The class with the highest amount of consecutive samples within the control limits is household, followed by technology.

4.2. Type I Error Calculation

To calculate the probability of making a type I error, we convert the data into a normal distribution and mean of zero. The standard deviation is taken to be one. Therefore, we want to know what the probability of our indicators telling us that the process is out of control (outside three standard deviations) when it is actually in control (within three standard deviations).

The value obtained was that the probability of making a type I error is 0.2700 %

4.3. Minimum cost calculation

In this question, the objective was to find the value of x (change in the mean delivery time of technology class items) that minimizes the total cost of being late. The information given was that it costs R329/item-late-hour (if delivery is late), and R2.5/item/hour to change the delivery times of each sale. The cost of increasing delivery time is -R2.5, so money is then saved.

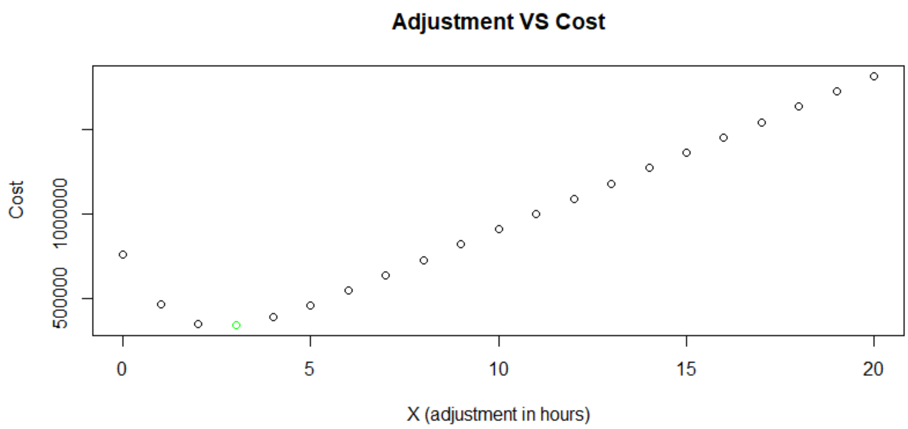

Figure 25: Plot of adjustment in hours versus cost

The minimum cost occurs at the green point, with the adjustment in hours being shifting the mean down with 3 hours. The minimum cost achieved was R340,892.5

4.4. Type II Error calculation

In this question, the probability of making a type II error is calculated. This is when it is assumed that a process is under control, when in reality, it is not. As with question 4.2, it is assumed that the process follows a normal distribution. The mean is taken as zero, with standard deviation being one. The process is considered out of control when it is more than three standard deviations from the mean.

The calculated probability for making a type II error is 0.1025

5. Manova

In this section, a Manova analysis is done. The target features are delivery time, with the predictor variables used being class, age and why.bought. The summary of the manova is given below.

```
> summary(res.man_1)
                Df  Pillai approx F num Df den Df Pr(>F)    
Class            6 1.67972   157303     12 359930 <2e-16 ***
Why.Bought       5 0.00009        2     10 359930 0.0805 .  
AGE              1 0.00052       47      2 179964 <2e-16 ***
Residuals   179965                                          
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 26: Summary of Manova

```
> summary.aov(res.man_1)
 Response Price :
                Df     Sum Sq    Mean Sq    F value Pr(>F)    
Class            6 5.7168e+13 9.5281e+12 80257.7518 <2e-16 ***
Why.Bought       5 5.5624e+08 1.1125e+08     0.9371 0.4555    
AGE              1 1.2094e+07 1.2094e+07     0.1019 0.7496    
Residuals   179965 2.1365e+13 1.1872e+08                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Response Delivery.time :
                Df   Sum Sq Mean Sq    F value   Pr(>F)    
Class            6 33458565 5576427 6.2978e+05 < 2e-16 ***
Why.Bought       5      107      21 2.4099e+00 0.03412 *  
AGE              1      835     835 9.4325e+01 < 2e-16 ***
Residuals   179965  1593510       9                       
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 27: Summary per response variable

The above summary shows that the variable class does affect Price and delivery time, while age only somewhat influences price but not delivery time. Why.Bought does not seem to have any correlation with either price or delivery time.

6. Reliability of service and products

In this section, some statistical questions are answered which are unrelated to the data.

6.1. Lafrideradora

For question 6, the values used in the plot were calculated as follows:

```
# q6
C <- 45
N <- 0.06
d <- 0.04
k <- C/(d^2)
x <- seq(-1, 01, by = 0.01)
L <- k*(x-N)^2
```

Figure 28: Values used for Q6.1 A

The resulting Taguchi loss plot is as given below.

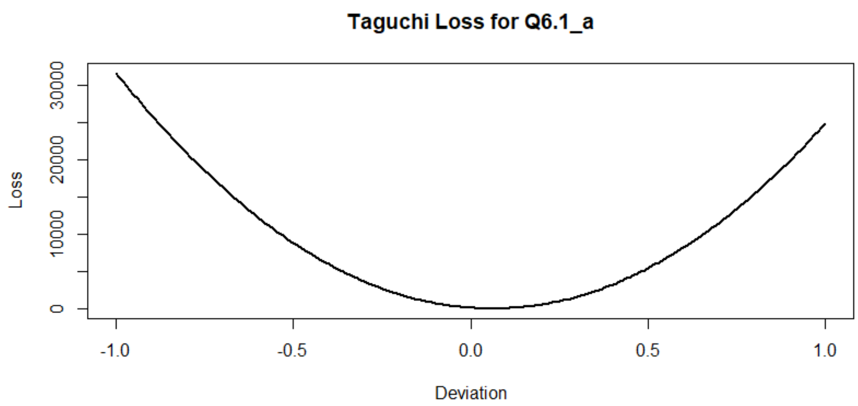

Figure 29: Taguchi loss for Q6.1 , question 6

For question 7, the values used in the plot were calculated as follows:

```
# Values are initialized to
C <- 35
k <- C/(d^2)
x <- seq(-1, 1, by = 0.01)
L <- k*(x-N)^2
```

Figure 30: Values used for question 6.1, question 7

The resulting Taguchi loss plot is given below.

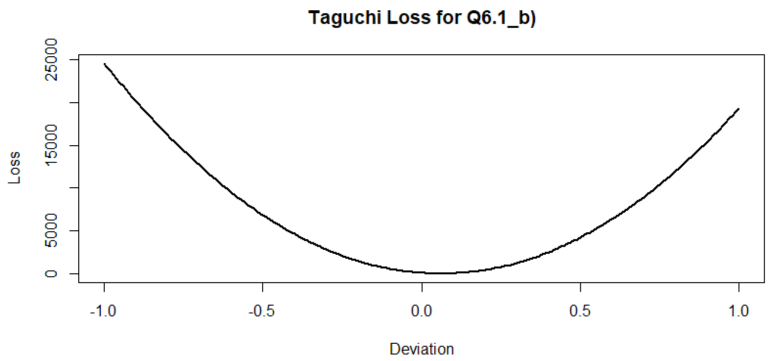

Figure 31: Taguchi loss function for textbook question 7

For part b of question 7, the calculated Taguchi loss was 23.822.

6.2. Magnaplex

In this question, two scenarios are given relating to the reliability of a production line which is dependant on the working of machines.

Scenario one is that there is only one of each machine, and they are operating in series.

For scenario two, there are two of each machine operating in parallel.

The calculations for this question are given below.

```
#6.2
#a)
A <- 0.85
B <- 0.92
C <- 0.9 # Assign dependability of machines
ReliabilityA<- A*B*C # All machines have to work when in series
#b)
ReliabilityB <- (1-(1-A)^2)*(1-(1-B)^2)*(1-(1-C)^2) # Both machines have to fail when in parallel
```

Figure 32: Calculations for Question 6.2

The calculated reliability for scenario one was 0.7038.

The calculated reliability for scenario two was 0.9615

6.3. Delivery Reliability

In this question, the proportion of days for which a certain number of drivers and vehicles were available was given. It is also stated that for reliable delivery, at least 19 vehicles (by association, also 19 drivers) are needed. The question for scenario one is then to estimate how many days in a year reliable delivery can be expected.

The logic used in answering the question is that the probability for success on each day will be the total proportion of days for which there were sufficient vehicles available, times the proportion of days for which sufficient drivers were available. This “success” probability was then used in a binomial distribution, with the number of trials occurring being 365 days.

Using this logic, it was found that for scenario one, the expected number of days in a year for which reliable delivery can be expected is 3 days.

For scenario two, it was found that the expected number of days on which reliable delivery can be expected stays the same at three days.

Conclusion

In section one, the data was split in such a way as to make the process repeatable. The different data sets created are showcased and the methods used were explained.

In section two, the descriptive statistics needed to provide a basic understanding of how the data is spread out is presented. Some of the important findings are that the class of purchase strongly influences both price and delivery time of the product. The age of the customer also somewhat influences what they choose to purchase. The Why.Bought variable is also correlated with the class of item purchased, with more expensive items being less prone to be bought randomly and stem more from recommendation.

In section three, the control limits for each class item's delivery time are calculated and presented in a table. The control charts for each process are also given and any trends or problems observed are stated. The processes for which gave indication of being out of control were clothing, household, and luxury items.

In section four, some numerical descriptors are calculated and provided for the work done in section three. Regarding sample means, gifts was the class which by far had the most samples which were out of control. With the sample standard deviations, luxury had the least samples within specification percentage wise. The type I and II errors were calculated and given.

In section five, the Manova analysis indicated that the class variable strongly influences both the price and delivery time of the sale. The age variable somewhat influenced price, but this is mostly because of autocorrelation between the age and class variable. The reason for buying does not influence either price or delivery time.

In section six, all the questions were answered, and results displayed.

References

1. : Quality-One | Quality and Reliability Consulting - Training - Facilitation. 2022. *Process Capability | Quality-One*. [online] Available at: <https://quality-one.com/process-capability/#:~:text=Process%20capability%20indices%20Cp%20and%20Cpk%20evaluate%20the,your%20process%20is%20centered%20between%20the%20specification%20limits.> [Accessed 19 October 2022].