ECSA REPORT

Quality Assurance 344

Stephanie van Papendorp
22792937

21 October 2022

Table of Contents

List of Figures

List of Tables

Introduction

The results of an online business's client data analysis will be discussed in this report. The relationships between the various features of the data will be explored first, followed by a discussion of the control charts of the delivery times for each class. The Taguchi loss for the company's subsidiary Cool Food will be discussed, as will whether its other subsidiary, Magnaplex, is wasting money by having backup machines for its processes.

Part 1: Data Wrangling

In this part of the project, data wrangling takes place. This is the process of transforming raw data into more usable formats. The dataset 'salesTable2022.csv' was imported and separated into valid data and incomplete data. Instances in the dataset that contain missing values and negative values were regarded as incomplete data. In the case when instances had missing values and/or negative values, those instances were removed from the original dataset and stored in a separate dataset. The feature AGE had very large values, although it is possible for a person to be older than 100, it is highly unlikely to have 1094 instances in this dataset above 100 years old. These instances should be evaluated or removed from the dataset. For this report, all instances of the feature AGE were stored in the valid dataset.

After separating the raw dataset, two new datasets were created. The dataset 'validData' contained all the nonnegative and valid (i.e. not NA) instances. The instances that were removed from the original dataset because of their invalid nature were stored in a dataset named 'invalidData'. This invalid dataset consists of 22 instances of which 17 are missing values and 5 are negative values.

For the remainder of this report, only the 'validData' dataset will be used. The features, 'Year', 'Month', 'Day', and 'X' are chronologically ordered.

	X	ID	AGE	Class	Price	Year	Month	Day	Delivery.time	Why.Bought
1	1	19966	54	Sweets	246.21	2021	7	3	1.5	Recommended
2	2	34006	36	Household	1708.21	2026	4	1	58.5	Website
3	3	62566	41	Gifts	4050.53	2027	8	10	15.5	Recommended
4	4	70731	48	Technology	41843.21	2029	10	22	27.0	Recommended
5	5	92178	76	Household	19215.01	2027	11	26	61.5	Recommended
6	6	50586	78	Gifts	4929.82	2027	4	24	14.5	Random
7	7	73419	35	Luxury	108953.53	2029	11	13	4.0	Recommended
8	8	32624	58	Sweets	389.62	2025	7	2	2.0	Recommended
9	9	51401	82	Gifts	3312.11	2025	12	18	12.0	Recommended
10	10	96430	24	Sweets	176.52	2027	11	4	3.0	Recommended
11	11	87530	33	Technology	8515.63	2026	7	15	21.0	Browsing

Showing 1 to 11 of 180,000 entries, 10 total columns

Figure 1: Snippet of raw dataset

	X	ID	AGE	Class	Price	Year	Month	Day	Delivery.time	Why.Bought
1	463	47101	50	Clothing	1030.86	2021	1	1	9.0	Recommended
2	2627	88087	21	Clothing	428.03	2021	1	1	10.0	Recommended
3	3374	25418	68	Household	13184.41	2021	1	1	48.5	Website
4	5288	13566	94	Household	7021.90	2021	1	1	42.0	Recommended
5	8182	84692	35	Clothing	475.18	2021	1	1	9.0	Recommended
6	9272	46305	72	Clothing	580.98	2021	1	1	8.5	Random
7	9712	92105	45	Household	6877.00	2021	1	1	43.0	Recommended
8	12163	21614	27	Clothing	513.13	2021	1	1	9.5	Recommended
9	12195	12174	56	Household	14538.64	2021	1	1	41.5	EMail
10	20004	84558	74	Food	255.41	2021	1	1	2.0	Recommended
11	20509	15630	32	Clothing	164.56	2021	1	1	9.0	Recommended

Showing 1 to 11 of 179,978 entries, 10 total columns

Figure 2: Snippet of ordered valid dataset

Part 2: Descriptive Statistics

2.1 Data Visualisation

It is important to first understand the data within a dataset before applying any descriptive statistics to it.

```
> summary(ordData)
       X                ID             AGE             Class               Price          
 Min.   :     1   Min.   :11126   Min.   : 18.00   Length:179978      Min.   :    35.65  
 1st Qu.: 45004   1st Qu.:32700   1st Qu.: 38.00   Class :character   1st Qu.:   482.31  
 Median : 90005   Median :55081   Median : 53.00   Mode  :character   Median :  2259.63  
 Mean   : 90003   Mean   :55235   Mean   : 54.57                      Mean   : 12294.10  
 3rd Qu.:135000   3rd Qu.:77637   3rd Qu.: 70.00                      3rd Qu.: 15270.97  
 Max.   :180000   Max.   :99992   Max.   :108.00                      Max.   :116618.97  
      Year          Month             Day         Delivery.time   Why.Bought       
 Min.   :2021   Min.   : 1.000   Min.   : 1.00   Min.   : 0.5   Length:179978     
 1st Qu.:2022   1st Qu.: 4.000   1st Qu.: 8.00   1st Qu.: 3.0   Class :character  
 Median :2025   Median : 7.000   Median :16.00   Median :10.0   Mode  :character  
 Mean   :2025   Mean   : 6.521   Mean   :15.54   Mean   :14.5                     
 3rd Qu.:2027   3rd Qu.:10.000   3rd Qu.:23.00   3rd Qu.:18.5                     
 Max.   :2029   Max.   :12.000   Max.   :30.00   Max.   :75.0                     
> 
```

Figure 3: Summary of ordered dataset

Figure 3 above illustrates a summary of the dataset 'ordData' and gives a five number summary of all the features with numeric values and information about the length, class and mode of features that contain characters. The summary shows that the dataset contains 10 features. Two of these features, 'Class' and 'Why.Bought', are categorical features and the remaining eight features are numeric. After implementing the nrow() function on the dataset, the number of instances was found to be 179978. For this part of the project, descriptive statistics was performed with guidance from the exercises created by Fernando Hernandez (Hernandez, 2015).

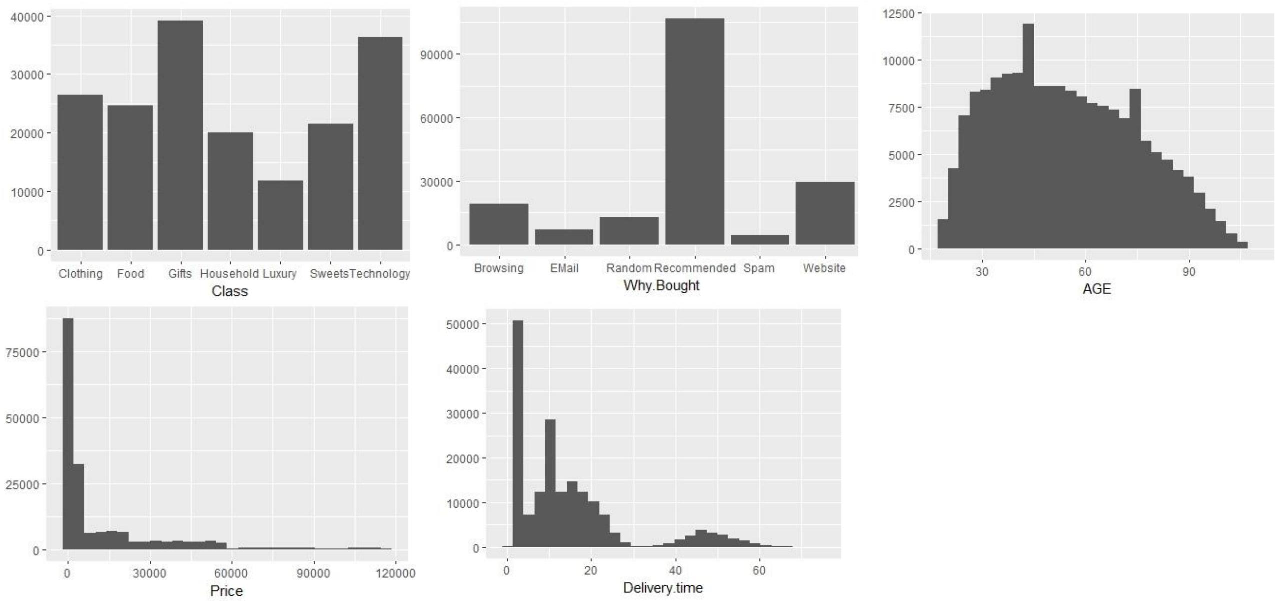


Figure 4: Distribution of features

Figure 4 illustrates the distribution of five of the features in the dataset 'ordData'. The distributions of the features X, ID, Year, Month and Day are not illustrated because they all have a normal distribution. From these illustrations, various conclusions can be made. The most products sold are gifts and the reason for purchasing items are mostly due to recommendations. The distribution of the 'AGE' feature is skewed to the right which indicates that most customers are of a younger age.

'Price' is exponentially distributed, which means that a lot of products are purchased at a lower price and as the price increases, the number of products purchased decreases drastically. 'Delivery.time' has a multimodal distribution with three peaks.

In Figure 5 below, the delivery times for all seven classes are illustrated. Household items can be complex, personalised and handmade; therefore it is expected that these items will have a longer delivery time. Food and sweets are simple items with a limited lifetime, thus, shorter delivery times for these items are expected and necessary.

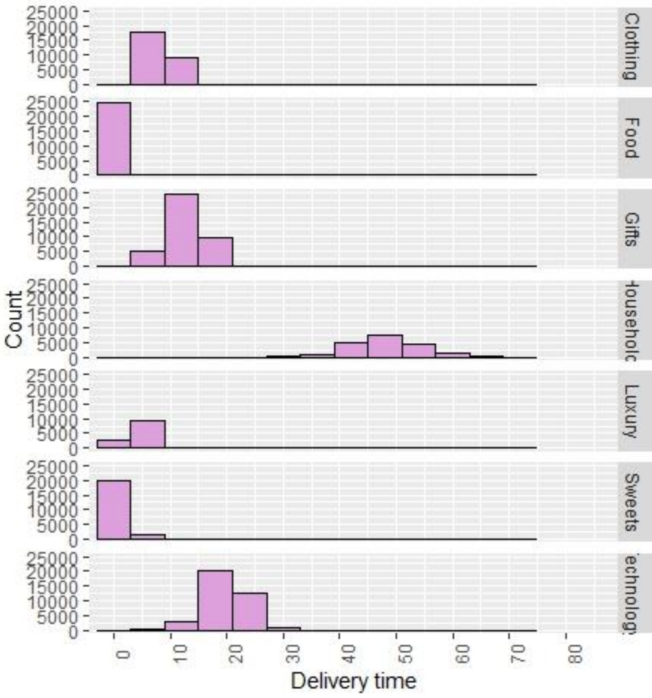


Figure 5: Delivery times of each class

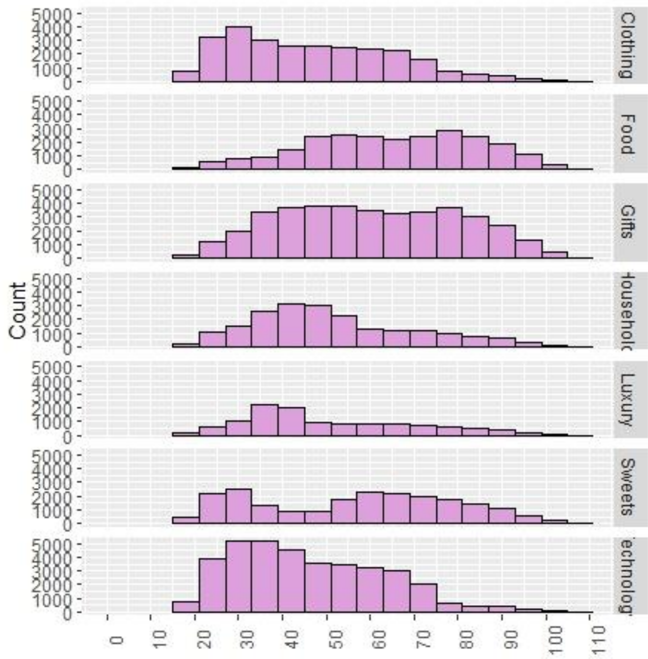


Figure 6: Age distribution in each class

The graphs in Figure 6 on the previous page, shows the distribution of the ages within each class. Clothing, household, and technology items are mostly purchased by customers younger than the age of 60 years. Luxury items are more popular amongst younger customers, especially between the ages of 35-45 years. Sweets are mostly purchased by customers in the age groups of 20-35 years and 55-75 years.

Figure 7 below shows the highest, lowest and the median price of items in each class. The overall highest price is that of a luxury item and the overall lowest price is that of a sweets item. The median price of luxury items are the highest and that of sweets is the lowest. This indicates that luxury items are the most expensive items and sweets are the least expensive items. Food and sweet items have the smallest difference in price range, which indicates that the prices of these items are fairly constant.

	Class	max_price	min_price	median_price
	<chr>	<dbl>	<dbl>	<dbl>
1	Clothing	1154.	128.	642.
2	Food	692.	128.	408.
3	Gifts	5774.	173.	2962.
4	Household	21935.	128.	10961.
5	Luxury	116619.	12825.	65342.
6	Sweets	576.	35.6	303.
7	Technology	57735.	935.	29654.

Figure 7: Min., max., and median price of each class

The same method was used to determine the shortest, longest, and median delivery times of items in each class. Figure 8 supports the conclusion made from Figure 5 that household items have the longest delivery time and sweets have the shortest delivery time. Food and sweets items have the shortest overall median delivery time.

	Class	max_delTime	min_delTime	median_delTime
	<chr>	<dbl>	<dbl>	<dbl>
1	Clothing	11.5	6.5	9
2	Food	3.5	1.5	2.5
3	Gifts	22	3.5	13
4	Household	75	27.5	48.5
5	Luxury	8	1	4
6	Sweets	4.5	0.5	2.5
7	Technology	33	4.5	20

Figure 8: Max., min., and median delivery time of each class

It was found that creating simple histograms of the price of each class did not provide useful results. To make it easier to understand and interpret the data, a logarithmic histogram, that compresses the range of data within each class, was used instead. The data is visualised in Figure 9 on the next page. As expected, the price range of food items is the smallest. As mentioned previously, household and luxury items can be complex and personalised, therefore it is expected that these two classes will show a large price range.

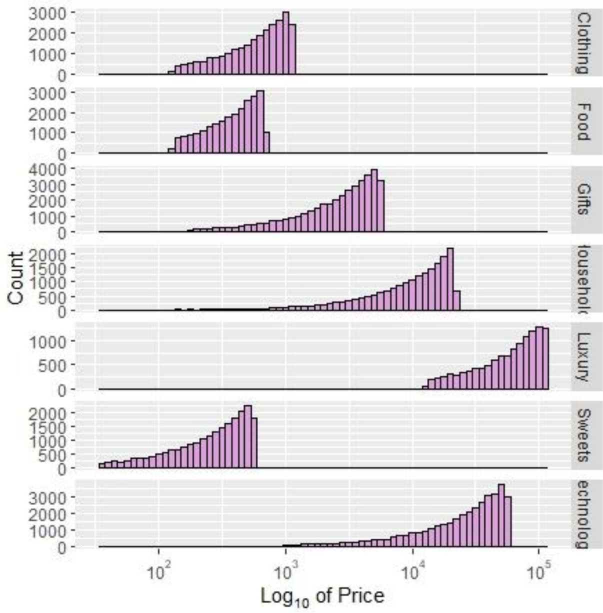


Figure 9: Logarithmic histograms of price in each class

To visualise the distribution of the delivery time and price of all the classes, boxplots were created. This is a good manner of illustrating all the data, it is clear and easy to understand and interpret. The boxplots in Figure 10 again supports the conclusions previously made; household items have the longest delivery time and, food and sweets have the shortest delivery time, while luxury items are the most expensive and food and sweets are the cheapest. The individual boxplots for each class can also be analysed to identify outliers. Food, household, technology, and luxury items have a lot of outliers.

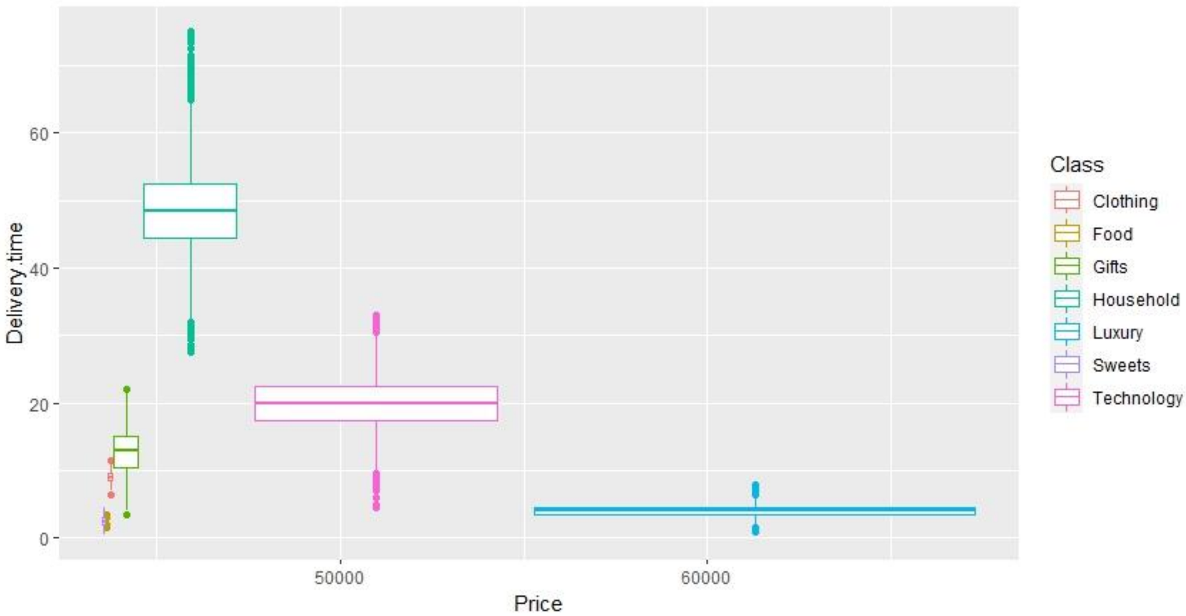


Figure 10: Boxplots of delivery time and price for each class

Figure 11 below, illustrates the age distribution for every reason why a purchase was made. Most items were purchased because it was recommended by others. Therefore, the company should focus on customer satisfaction, because the more satisfied current customers are, the more likely they are to recommend the products to others. The younger age groups are more hands-on with technology, thus customers who purchase items by making use of the website fall within the younger age groups.

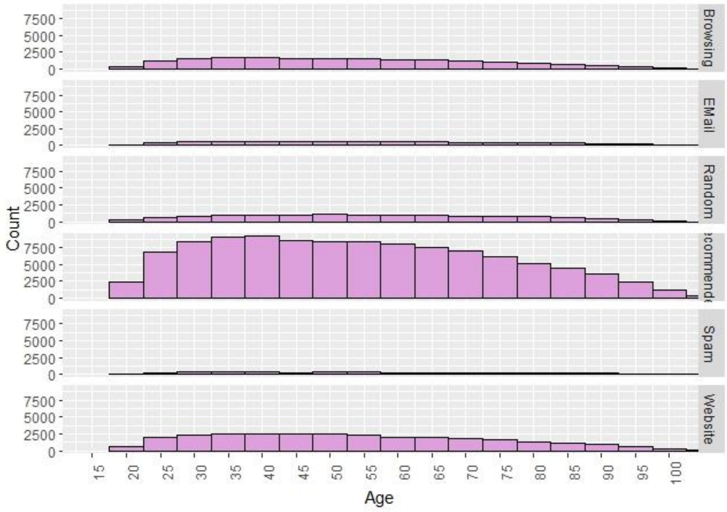


Figure 11: Age distribution for every reason an item was purchased

The price distribution from 2021 through to 2029 is illustrated in Figure 12. The most sales were made during the year 2021, which also has a fairly uniform price distribution. The other years have a right skewed distribution, which is caused by the outliers in these years.

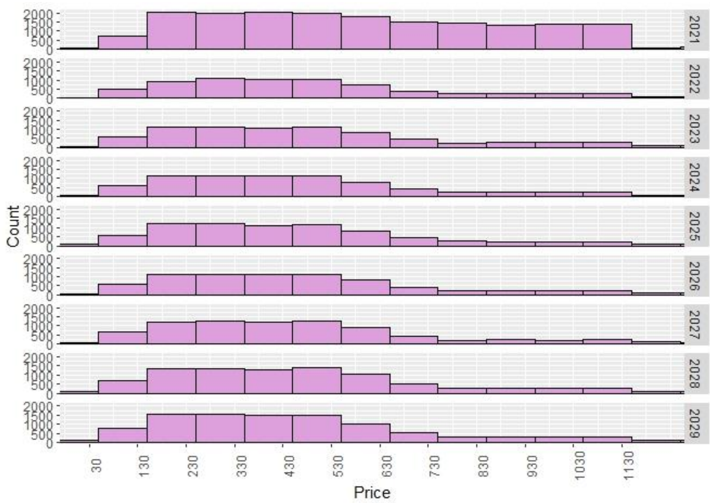


Figure 12: Price distribution for 2021 to 2029

Another useful data visualisation, that of the fluctuation in the delivery time over the years is illustrated in Figure 13. Clothing, food, luxury and sweets items remained constant over the nine-year period, while gift, household and technology items fluctuated.

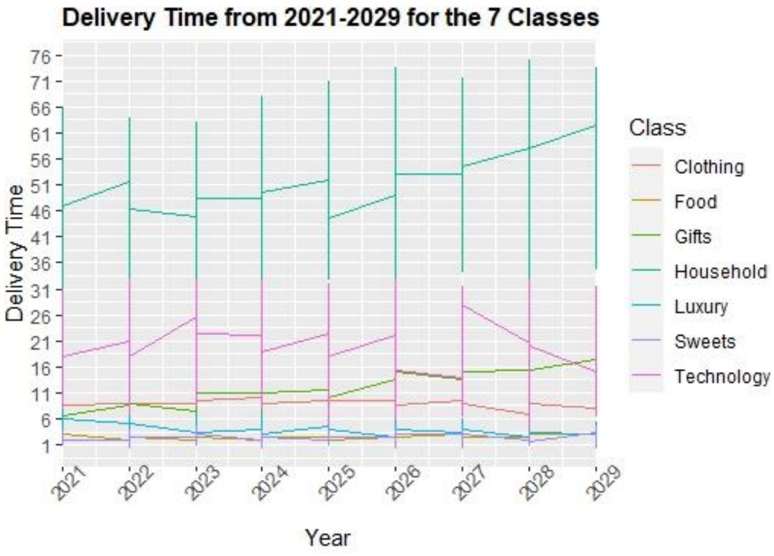


Figure 13: Delivery time from 2021-2029 for each class

2.2 Process Capability

The process capability indices gauge how much variance a process encounters in comparison to its specification parameters. One might compare various procedures in terms of the ideal circumstance or whether they live up to certain expectations.

With the assumption that the USL = 24 hours and the LSL = 0, the process capability indices for the 'ordData' dataset's delivery time of technology items are listed in Error! Reference source not found.. To calculate these indices, the mean and standard deviation of the delivery time of the class 'Technology' was first determined to be 20.01095 and 3.501993 respectively.

Process Capability Index	Value
C_p	1.142207
C_{pu}	0.3796933
C_{pl}	1.90472
C_{pk}	0.3796933

Table 1: Process capability indices

A LSL value of zero is logical, because it is a natural boundary. When calculating the C_{pk}, with a lower bound limit of 0 hours, the C_{pk} value will also approach 0. The C_p value obtained is larger than 1, which means that the process meets specifications, and it is capable. The C_{pk} value is well below 1, indicating that the delivery time that the business achieves for technology items is very poor and they should improve their delivery system.

Part 3: Statistical Process Control (SPC) for the X&s-charts

For this part of the project statistical process control (SPC) will be done for the X&s-charts. The control charts are created by using the first 30 samples of 15 sales each from all seven classes in the dataset 'ordData'.

3.1 Control charts for the first 30 samples

A control chart makes it easier to capture data and shows you when an unusual occurrence, such as an observation that is significantly higher or lower than "normal" process performance, takes place. Both the X-bar and S-charts were initialised the same way. The X-bar chart was initialised by determining the mean of the first 30 samples' means, and then this value was used to determine all of the other parameters included in the chart.

The X-bar chart table, illustrated in Table 2, shows the mean of the delivery times for the first 30 samples of each class.

	UCL	U2Sigma	U1sigma	CL	L1Sigma	L2Sigma	LCL
Technology	22.974616	22.107892	21.241168	20.374444	19.507721	18.640997	17.774273
Clothing	9.404934	9.259956	9.114978	8.970000	8.825022	8.680044	8.535066
Household	50.248328	49.019626	47.790924	46.562222	45.333520	44.104818	42.876117
Luxury	5.493965	5.241162	4.988359	4.735556	4.482752	4.229949	3.977146
Food	2.709458	2.636305	2.563153	2.490000	2.416847	2.343695	2.270542
Gifts	9.488565	9.112747	8.736929	8.361111	7.985293	7.609475	7.233658
Sweets	2.897042	2.757287	2.617532	2.477778	2.338023	2.198269	2.058514

Table 2: X-bar chart table for the first 30 samples of each class

The S-chart table, illustrated in Table 3, shows the standard deviation of the delivery times for the first 30 samples of each class.

	UCL	U2Sigma	U1sigma	CL	L1Sigma	L2Sigma	LCL
Technology	5.1805697	4.5522224	3.9238751	3.2955278	2.6671805	2.0388332	1.4104859
Clothing	0.8665596	0.7614552	0.6563509	0.5512465	0.4461422	0.3410379	0.2359335
Household	7.3441801	6.4534101	5.5626402	4.6718703	3.7811003	2.8903304	1.9995605
Luxury	1.5110518	1.3277775	1.1445032	0.9612289	0.7779546	0.5946803	0.4114060
Food	0.4372466	0.3842133	0.3311800	0.2781467	0.2251134	0.1720801	0.1190468
Gifts	2.2463333	1.9738773	1.7014213	1.4289652	1.1565092	0.8840532	0.6115971
Sweets	0.8353391	0.7340215	0.6327039	0.5313862	0.4300686	0.3287509	0.2274333

Table 3: S-chart table for the first 30 samples of each class

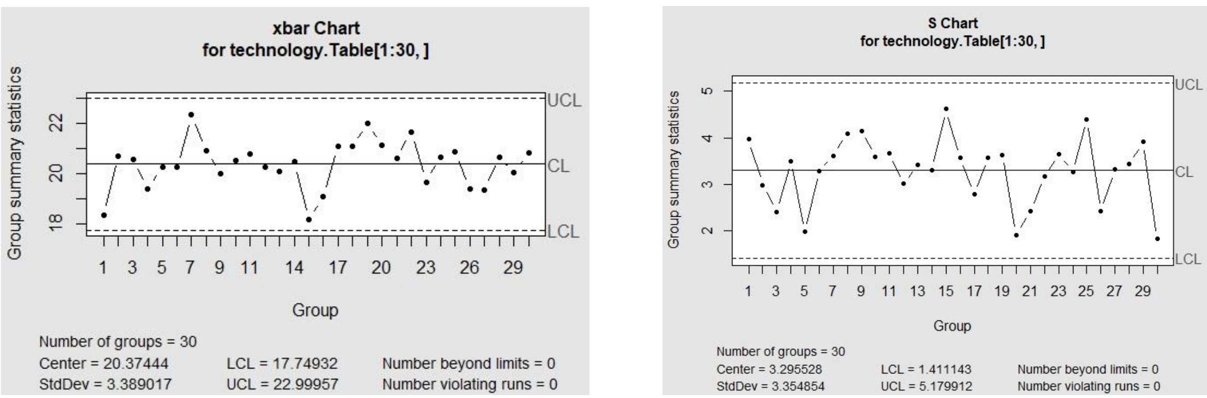


Figure 14: X-bar and S-chart for Technology

Figure 14 on the previous page shows the X-bar and S-chart for delivery times of the ‘Technology’ class. The S-chart on the left shows that no points are out of control, and some variation is present in the standard deviation of the samples. Because the S-chart is in control, the control limits on the X-bar chart are accurate. The X-bar chart on the right shows that no points are out of control, and some variation is present in the averages of the samples.

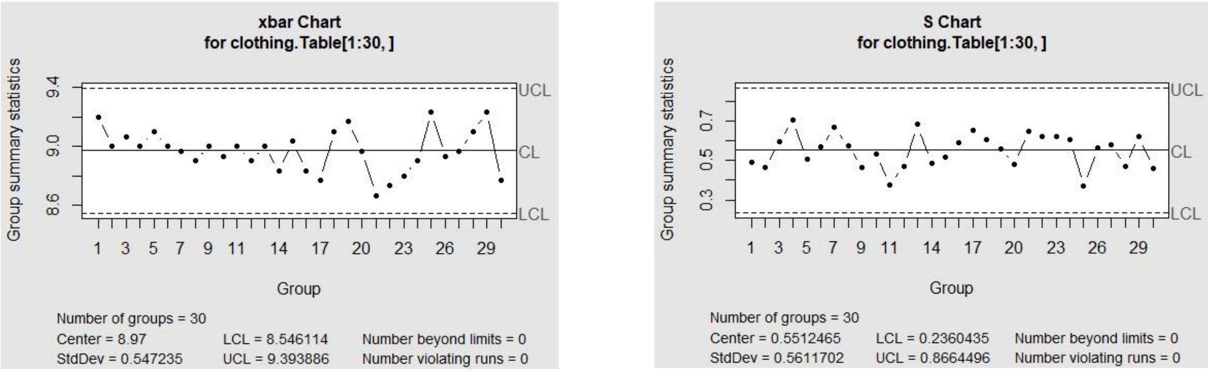


Figure 15: X-bar and S-chart for Clothing

Figure 15 above shows the X-bar and S-chart for delivery times of the ‘Clothing’ class. The S-chart on the left shows that no points are out of control, and the variation in the standard deviations of the samples is very little. The control limits on the X-bar chart are accurate because the S-chart is in control. The X-bar chart on the right shows that no points are out of control, and little variation is present in the averages of the samples.

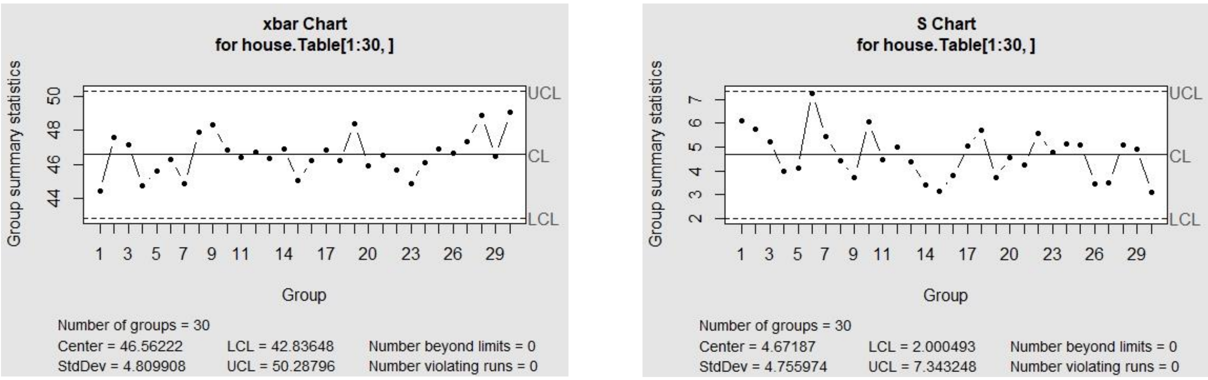


Figure 16: X-bar and S-chart for Household

Figure 16 above shows the X-bar and S-chart for delivery times of the ‘Household’ class. The S-chart on the left shows that no points are out of control, however, point 6 is on the verge of being beyond the upper control limit. There is some variation in the standard deviations of the samples. Because the S-chart is in control, the control limits on the X-bar chart are accurate. The X-chart on the right shows that no points are out of control, and there are some variations in the averages of the samples.

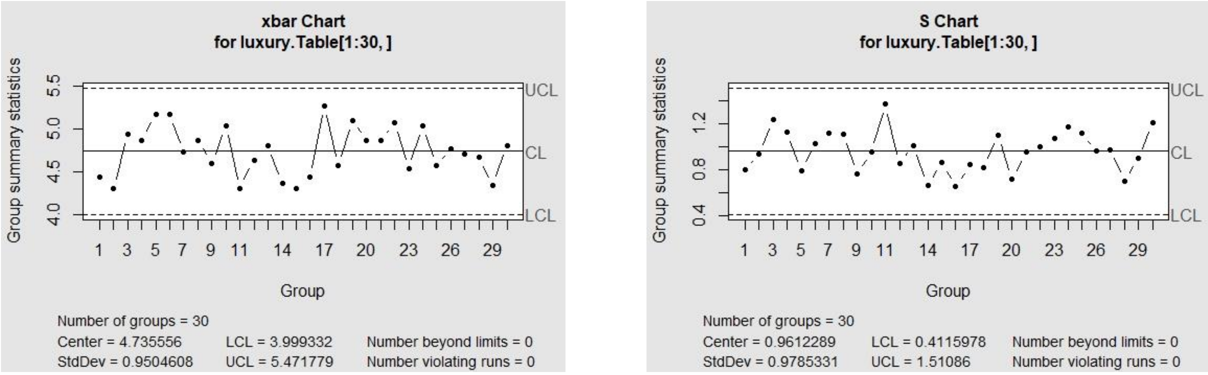


Figure 17: X-bar ans S-chart for Luxury

Figure 17 above shows the X-bar and S-chart for delivery times of the 'Luxury' class. The S-chart on the left shows that no points are out of control. There is some variation in the standard deviations of the samples. Because the S-chart is in control, the control limits on the X-bar chart are accurate. The X-chart on the right shows that no points are out of control, and there are some variations in the averages of the samples.

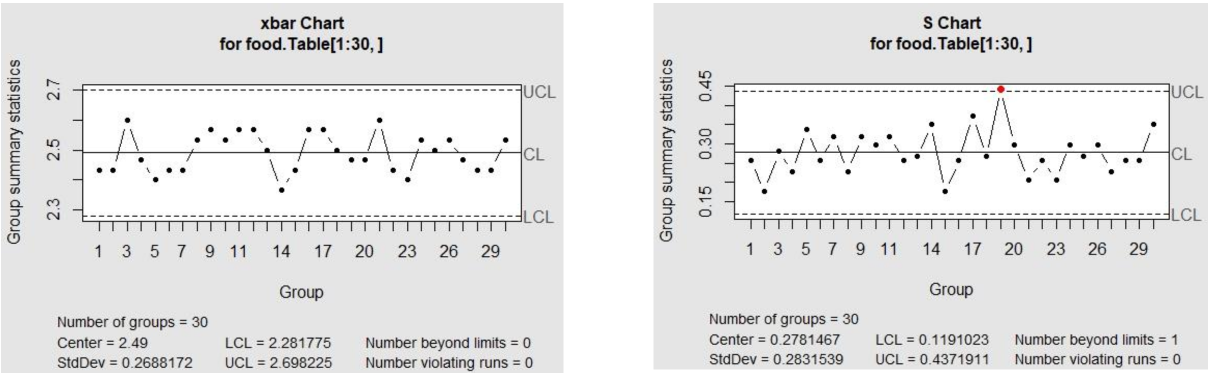


Figure 18: X-bar and S-chart for Food

Figure 18 above shows the X-bar and S-chart for delivery times of the 'Food' class. The S-chart on the left shows that point 19, marked in red, is out of control. Sample 19 is more than three standard deviations above the centre line. There is some variation in the standard deviations of the samples. Because the S-chart is out of control, the control limits on the X-bar chart are inaccurate. The X-chart on the right shows that no points are out of control, and there are variations in the averages of the samples.

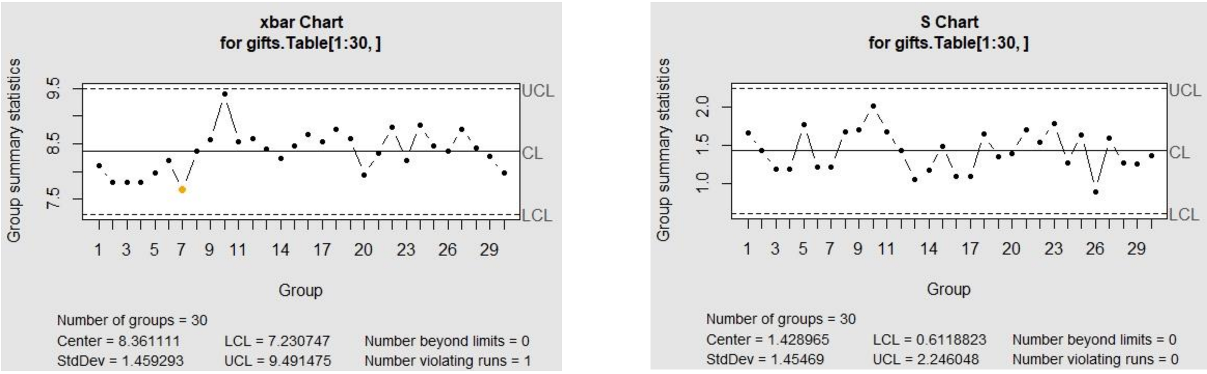


Figure 19: X-bar and S-chart for Gifts

Figure 19 above shows the X-bar and S-chart for delivery times of the 'Gifts' class. The S-chart on the left shows that no point is out of control. There is some variation in the standard deviations of the samples. Because the S-chart is in control, the control limits on the X-bar chart are accurate. The X-chart on the right shows that there are seven consecutive points (1-7) below the centre line which indicate out of control signals.

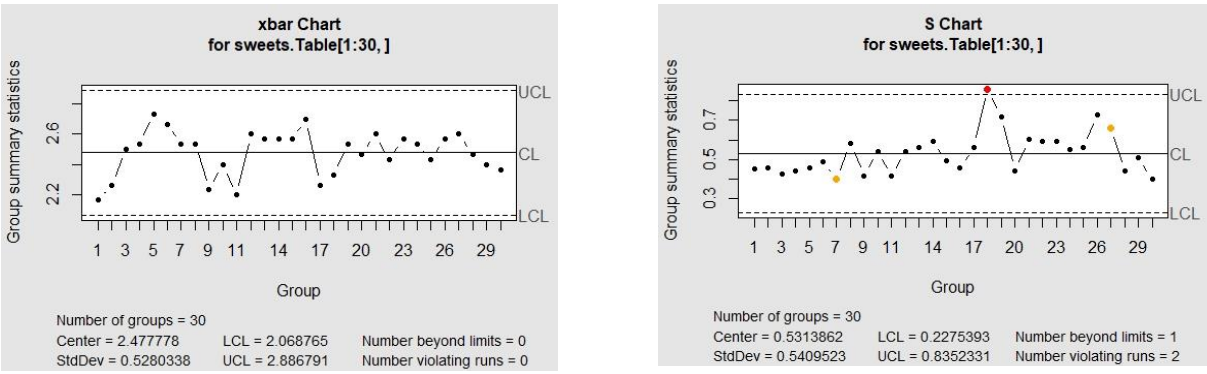


Figure 20: X-bar and S-chart for Sweets

Figure 20 above shows the X-bar and S-chart for delivery times of the 'Sweets' class. The S-chart on the left shows that one point is out of control. Sample 18 is more than three standard deviations above the centre line. Furthermore, there are two cases where there are seven consecutive points (1-7 and 21-27) below and above the centre line respectively. This is indicative of out-of-control signals. There is some variation in the standard deviations of the samples. Because the S-chart is out of control, the control limits on the X-bar chart are inaccurate. The X-chart on the right shows that no points are out of control and there are some variations in the averages of the samples.

3.2 “Control” samples from 31 until the end

The X-bar and S-chart for ‘Technology’ seen in Figure 21, shows the variation in the means and standard deviation of the samples from 31 to the end. The S-chart on the right shows that there are six samples with standard deviations beyond the control limits. From the X-bar chart on the left it can be seen that there are 16 outliers, marked in red, and 95 out of control signals, marked in orange. This is a relatively stable process.

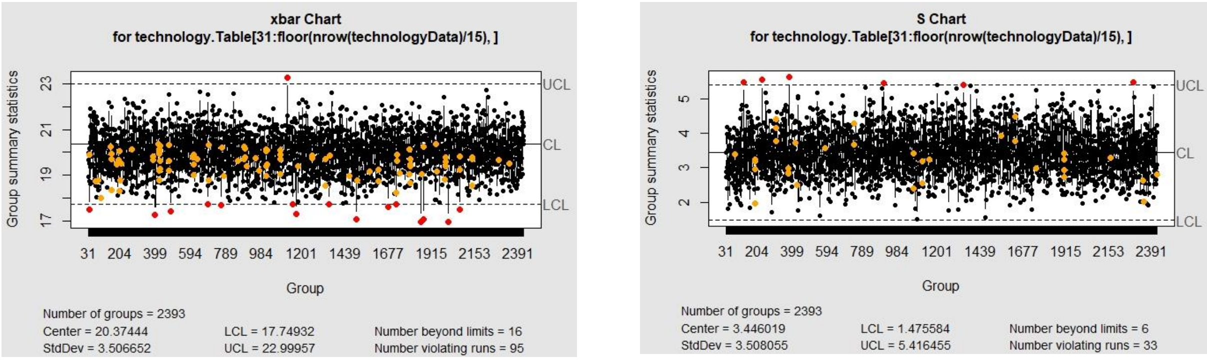


Figure 21: X-bar and S-chart for Technology [31:end]

The X-bar and S-chart for ‘Clothing’ seen in Figure 22, shows the variation in the means and standard deviation of the samples from 31 to the end. The S-chart on the right shows that there are 41 samples with standard deviations beyond the control limits. From the X-bar chart it can be seen that there are 22 outliers, marked in red, and 23 out of control signals, marked in orange. The few samples below the lower control limit are not necessarily a negative aspect, as this means that the business managed to deliver these clothing samples in far less time than the average delivery time. This is a relatively stable process; however the delivery time of this class must be analysed because the amount of samples outside of the control limit increases as time passes.

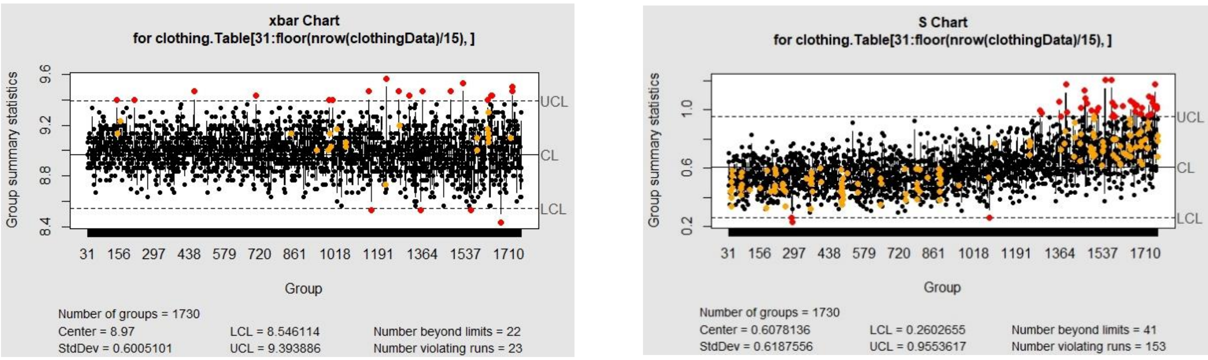


Figure 22: X-bar and S-chart for Clothing [31:end]

The X-bar and S-chart for ‘Household’ seen in Figure 23 on the next page, shows the variation in the means and standard deviation of the samples from 31 to the end. The S-chart on the right shows that there is a total of seven samples beyond the control limits. From the X-bar chart on the left it can be seen that there are 396 outliers, marked in red, and 557 out of control signals, marked in orange. Most of the outliers lie above the upper control limit, indicating that the mean of the delivery times has increased as time passed, making this an unstable process.

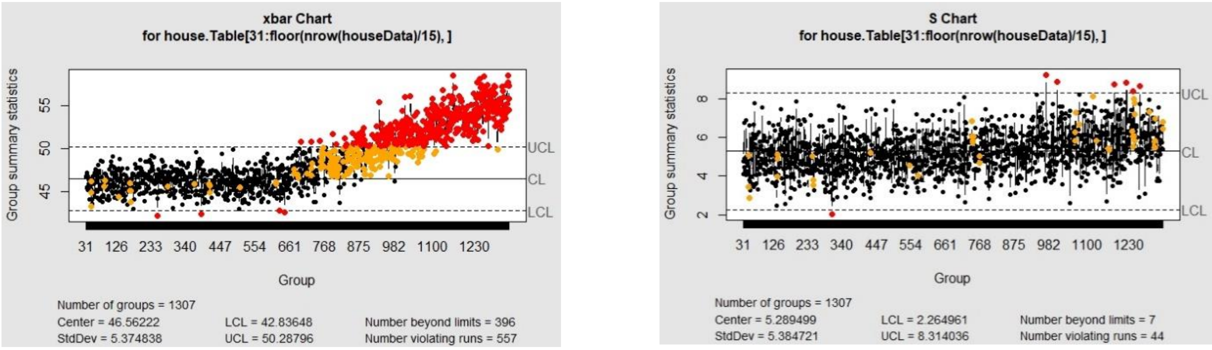


Figure 23: X-bar and S-chart for Household [31:end]

Figure 24 below illustrates the X-bar chart of the variation in means and the S-chart of the variation in standard deviations of the samples from 31 to the end for the class ‘Luxury’. The S-chart on the right shows that there is a total of 11 samples beyond the control limits. The X-bar chart on the left shows that the process has shifted out of control, as the points shifted to the lower half and mostly below the lower control limit. This phenomenon indicates that some sort of strategy has been successfully implemented to reduce the delivery time of ‘Luxury’ items. This strategy should be investigated and implemented in other classes as well with the aim to reduce their delivery time.

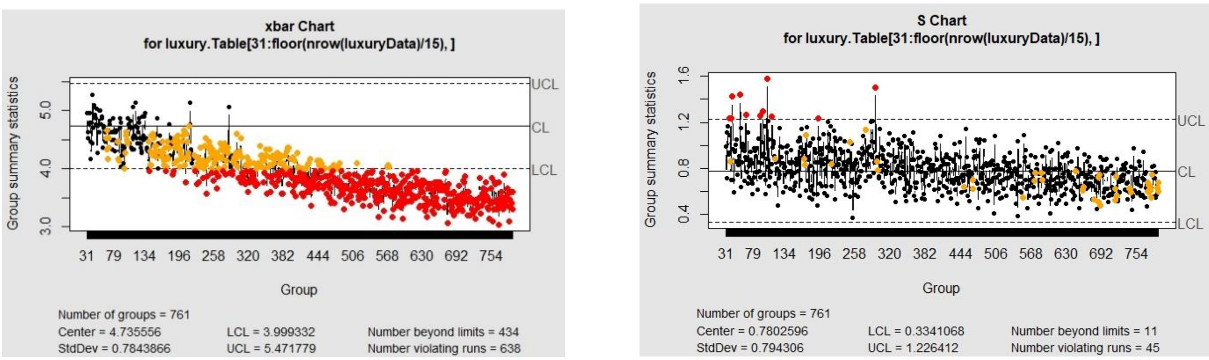


Figure 24: X-bar and S-chart for Luxury [31:end]

The X-bar and S-chart for ‘Food’ seen in Figure 25 on the next page, shows the variation in the means and standard deviation of the samples from 31 to the end. The S-chart on the right shows that there is a total of 15 samples beyond the control limits. The X-bar chart on the left shows that there are 18 outliers, marked in red, and 58 out of control signals, marked in orange. The outliers are not too far above the upper- and lower control limits, and therefore the process can be described as stable.

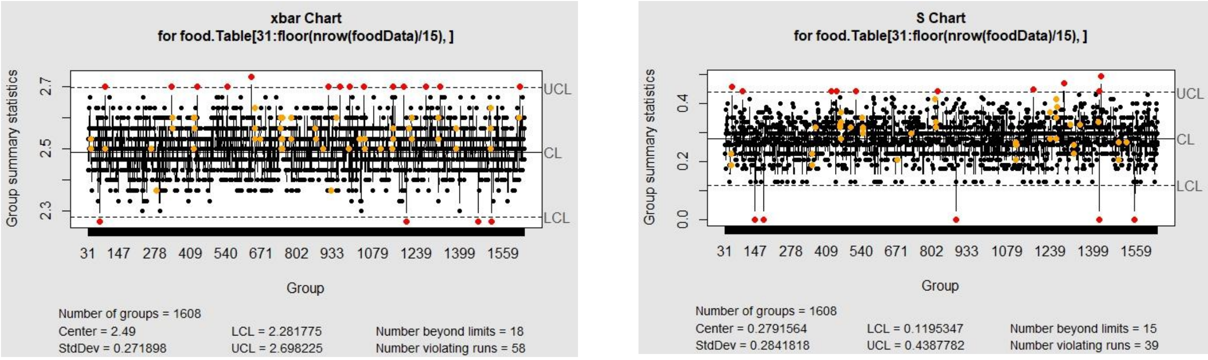


Figure 25: X-bar and S-chart for Food [31:end]

Figure 26 below illustrates the X-bar chart of the variation in means and the S-chart of the variation in standard deviations of the samples from 31 to the end for the class 'Gifts'. The S-chart on the right shows that there are merely five samples that are beyond the control limits. However, the X-bar chart on the left shows that there are 2290 samples, marked in red, beyond the control limits and 2397 out of control signals, marked in orange. The out-of-control samples are all located very close to or above the upper control limit. This sort of tendency towards and beyond the upper control limit means that as time passed, the mean delivery time of 'Gift' items increased. This is indicative of a severely unstable process and should be investigated and improved.

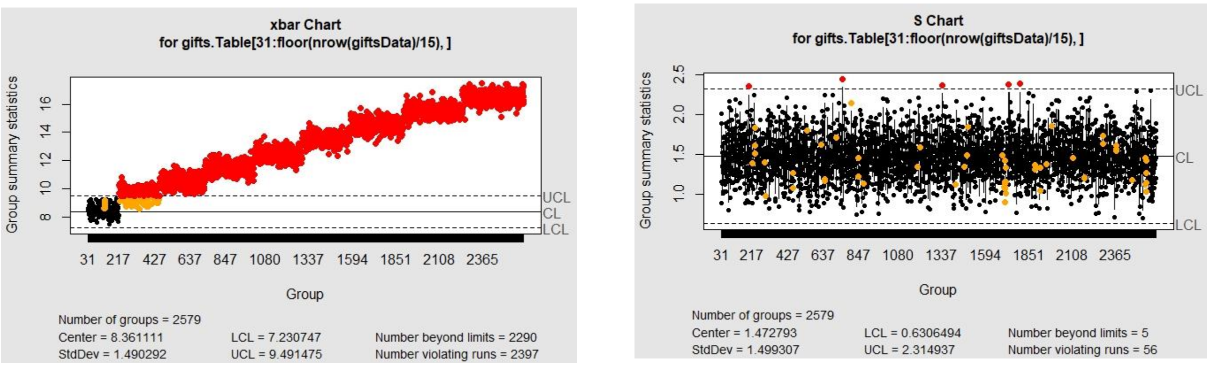


Figure 26: X-bar and S-chart for Gifts [31:end]

The X-bar chart illustrating the variation of the means and the S-chart illustrating the variation of the standard deviations of the samples from 31 to the end of the class 'Sweets' can be seen in Figure 27. The S-chart on the right shows that there is only one sample beyond the control limits. The X-bar chart on the left shows that there are five samples beyond the control limits, marked in red, and 31 out of control signals, marked in orange. The samples are evenly spread between the control limits, with very little drastic outliers which indicates that the process is stable.

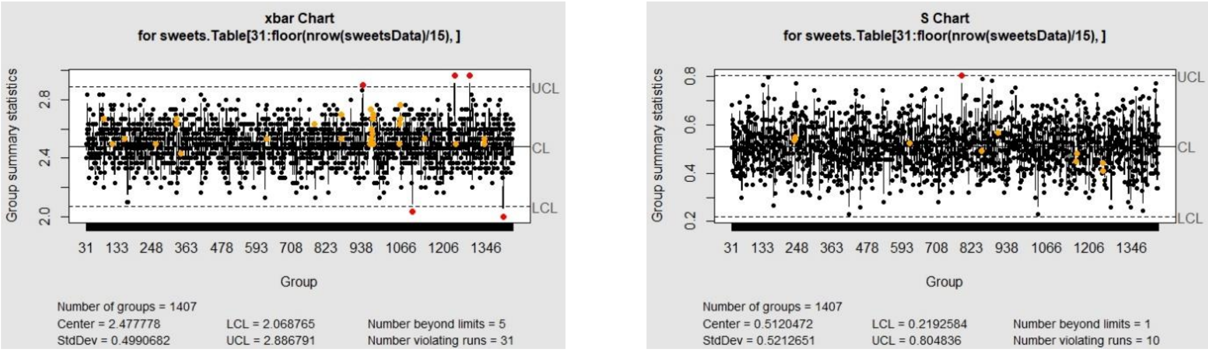


Figure 27: X-bar and S-chart of Sweets [31:end]

Part 4: Optimising the delivery processes

4.1 Out of control samples

4.1.1 Sample means outside the control limits (A)

Table 4 below summarises the first and last thee samples located outside the outer control limits for each of the seven classes, also referred to as outliers. The total amount of outliers in each class are in the first column of the table. 'Gifts' has the largest number of outliers, indicating that there is a large degree of variation in sample means for this class. 'Food' and 'Sweets' have the least number of outliers, which indicates that there is little variation in the sample means for these two classes.

	Number of outliers	**First**	**Second**	**Third**	**Third last**	**Second last**	**Last**
Technology	17	37	398	483	1872	2009	2071
Clothing	17	455	702	1152	1677	1723	1724
Household	400	252	387	629	1335	1336	1337
Luxury	434	142	171	184	789	790	791
Food	5	75	633	1203	1467	1515	-
Gifts	2290	213	216	218	2607	2608	2609
Sweets	5	942	1104	1243	1294	1403	-

Table 4: Summary of sample means outside control limits

4.1.2 Most consecutive samples between -0.3 and +0.4 sigma-control limits (B)

Table 5 on the next page shows the most consecutive sample standard deviations between the -0.3 and +0.4 sigma-control limits together with the last sample value. Technology and food has the most consecutive samples between the specified sigma-control limits, and household has the least. Table 5: The most consecutive sample standard deviations between -0.3 and +0.4 sigma-control limits

	Max. consecutive samples	Ending sample
Technology	7	2419
Clothing	4	1013
Household	3	908
Luxury	4	63
Food	7	952
Gifts	5	1651
Sweets	4	1292

Table 5: The most consecutive sample standard deviations between -0.3 and +0.4 sigma-control limits

4.2 Type I Error

A type 1 error occurs when you incorrectly reject the null hypothesis and claim that something significant happened when it did not (Anon., n.d.).

4.2.1 Type I Error for A

To calculate the probability of making a type I error for one standard deviation, this equation is used $P(type\ I) = pnorm(-1) * 2$. Fot his data, the probability of making a type I error for one standard deviation is 31.73%, which is relatively high.

To calculate the probability of making a type I error for three standard deviations, this equation is used $P(type\ 1) = pnorm(-3) * 2$. The type I error for three standard deviations for this data is 0.2699796%, which is very small. Therefore, the probability that the manufacturer will stop the process when he suspects the process is out of control is very small.

4.2.2 Type I Error for B

To calculate the probability of making a type I error for B , this equation is used $Prob(type\ 1) = pnorm(0.4) - pnorm(-0.3)$. The type I error for B is 27.33332%, which is significantly larger than for A. Therefore, the probability that the manufacturer will stop the process when he suspects the process is out of control is not very small.

4.3 Minimum cost with reduced delivery time

The total cost (R) vs every possible reduced delivery time (hours), is illustrated in Figure 28 below. The optimal delivery time by which the delivery of technology items should be reduced is three hours. On the graph, the optimal reduced delivery time at the least total cost is highlighted in blue. The total cost associated with a reduced delivery time of three hours is R340 870.

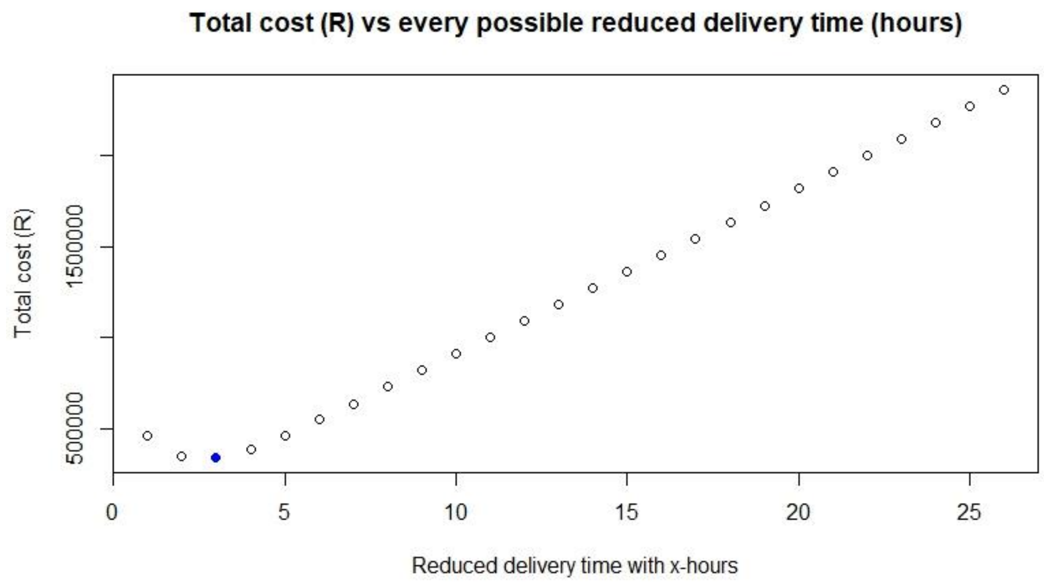


Figure 28: Total cost (R) vs every possible delivery time (hours)

4.4 Type II Error

A type II error is a term used to describe the error that occurs when a false null hypothesis is not rejected. The likelihood of making a type II error in the class 'Technology', given that the delivery process average moves to 23 hours, is illustrated in Figure 29 below.

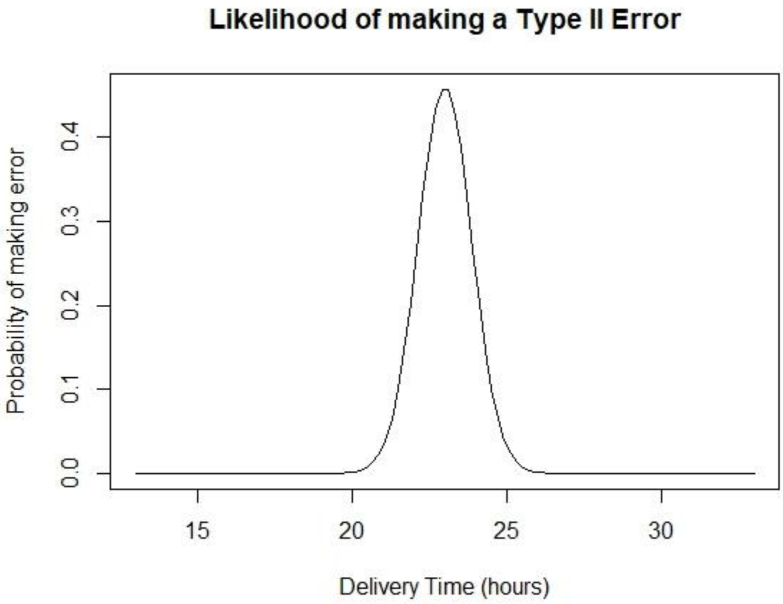


Figure 29: Normal distribution of a Type II Error

The z-value is calculated as follows:

$$P(Type\ II\ Error) = P_{norm} \times \left(\frac{A}{B}\right) \times 3$$

$$Z - value = \frac{UCL - new\ mean}{UCL - old\ mean} \times 3$$

$$= \frac{22.97462 - 23}{22.97462 - 20.37444} \times 3$$

$$= -0.02928744$$

From the standard z-table for normal distributions, the probability is 0.4880. Thus, the probability of making a type II error is 48.8%. The conclusion that can be made from this result is that the chance that a customer waits for a long period of time for items from the 'Technology' class to be delivered is high.

Part 5: DOE and MANOVA

When there are multiple response variables, a multivariate analysis of variance (MANOVA) can be used to test them all at once (Anon., n.d.). The results obtained from the MANOVA test performed on the delivery time and price of each of the seven classes is displayed in Figure 30.

```
> summary(MANOVA.test)
              Df Pillai approx F num Df den Df    Pr(>F)
Class          6 1.6797   157291     12 359942 < 2.2e-16 ***
Residuals 179971
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> # Look to see which differ
> summary.aov(MANOVA.test) # Summarize an Analysis of Variance Model
 Response Delivery.time :
                Df   Sum Sq Mean Sq F value    Pr(>F)
Class            6 33458565 5576427  629429 < 2.2e-16 ***
Residuals   179971  1594452       9
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response Price :
                Df     Sum Sq    Mean Sq F value    Pr(>F)
Class            6 5.7168e+13 9.5281e+12   80258 < 2.2e-16 ***
Residuals   179971 2.1366e+13 1.1872e+08
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 30: Results of MANOVA test on delivery time and price of each class

A test is deemed to be significant if the p-value is smaller than 0.05. Therefore, are both the features ‘Delivery.time’ and ‘Price’ highly significant each with a very small p-value of 0.00000000000000022.

Part 6: Reliability of the service and products

6.1 Problem 6 and 7 of chapter 7

6.1.1 Problem 6

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

$$k = \frac{45}{0.04^2} = 28125$$

$$Taguchi\ loss\ function = L(x) = k \times (x - T)^2$$
$$= 28125 \times (x - T)^2$$

6.1.2 Problem 7

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

a) *Determine the Taguchi loss function for this situation.*

$$k = \frac{35}{0.04^2} = 21875$$

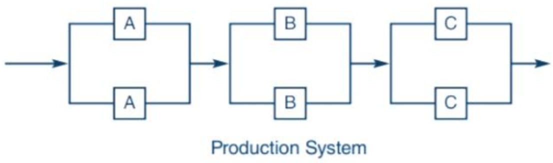
$$Taguchi\ loss\ function = L(x) = k \times (x - T)^2$$
$$= 21875 \times (x - T)^2$$

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

$$Taguchi\ loss\ function = L(x) = k \times (x - T)^2$$
$$= 21875 \times (0.027)^2$$
$$= 15.94687$$

6.2 Problem 27 of chapter 7

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.



The reliabilities of the machines are as follows:

Machine	Reliability
A	0.85
B	0.92
C	0.90

Figure 31: Screenshot of problem 27

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

$$Rs = RA \times RB \times RC$$
$$= 0.85 \times 0.92 \times 0.90$$
$$= 0.7038$$
$$= 70.38\%\ Reliable$$

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

$$Rp = 1 - (1 - RA)^2 \times (1 - RB)^2 \times (1 - RC)^2$$
$$= 1 - (1 - 0.85)^2 \times (1 - 0.92)^2 \times (1 - 0.90)^2$$
$$= 0.9999986$$
$$= 99.99986\%$$
$$\approx 100\%$$

The system reliability has increased with 29.62%, and thus it would be a good idea to implement a parallel system to increase the system reliability.

6.3 Delivery reliability problem

There are 21 delivery vehicles, 21 drivers and a total of 1560 days. The number of vehicles and drivers available over the past 1560 days is summarised in

Vehicles available	**Number of days vehicles available**	**Drivers available**	**Number of days driver available**
20	190	20	95
19	22	19	6
18	3	18	1
17	1	17	

Table 6: Vehicle and driver availability

The probabilities are calculated using the binomial distribution formula below.

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

In R, the $dbinom()$ function was used to calculate the probabilities listed below.

$P(no\ vehicles\ available) = \ 0.007064264$

$P(no\ drivers\ available) = \ 0.00306887$

The expected reliable days for delivery vehicles are 361.5013 days and the expected reliable days for drivers are 364.9844 days.

$$P(reliable) = P(vehicles\ relaible) \times P(drivers\ available)$$

$$= \ 0.9904146 \times 0.9999574$$

$$= 0.99037241$$

$$Expected\ reliable\ days = 0.99037241 \times 365$$

$$= 361.485929$$

Part 7: Conclusion

Following the analysis of the data, a few conclusions were reached. There is a correlation between age groups and the type of products purchased, and recommendation is the most common reason for customers to purchase products The control process reveals that the household and gifts classes have a lot of variation and that the process is unstable. The delivery times of these classes have increased over time. Management must determine what is causing the increase in average delivery times. Furthermore, the process average for technology must be reduced by 3 hours to reduce the cost of late items. The subsidiary Magnaplex's use of backup machines were found to be beneficial rather than wasteful. Finally, 361 days were discovered to be the number of days that reliable delivery can be expected.

References

Anon., 2022. *Process Capability Analysis Cp, Cpk, Pp, Ppk - A Guide.* [Online]
Available at: https://www.1factory.com/quality-academy/guide-to-process-capability-analysis-cp-cpk-pp-ppk.html#:~:text=In%20general%2C%20the%20higher%20the,2.00%20or%20higher%20where%20possible.
[Accessed 1 October 2022].

Anon., n.d. *MANOVA Test in R: Multivariate Analysis of Variance.* [Online]
Available at: http://www.sthda.com/english/wiki/manova-test-in-r-multivariate-analysis-of-variance
[Accessed 12 October 2022].

Anon., n.d. *Type I Error.* [Online]
Available at: https://www.isixsigma.com/dictionary/type-i-error/#:~:text=A%20type%201%20error%20is%20when%20you%20incorrectly%20reject%20the,significant%20happened%20when%20it%20did.
[Accessed 15 October 2022].

Hernandez, F., 2015. *Data Ananlysis with R - Exercises.* [Online]
Available at: http://fch808.github.io/Data-Analysis-with-R-Exercises.html
[Accessed 18 September 2022].