ECSA Project

Duan Marshall

Quality Assurance 344

23930934

Contents

Introduction

This report offers an in-depth analysis of an online business' client data.

The report consists of six parts namely Data Wrangling, Descriptive Statistics, Statistical Process Control, Optimising the Delivery Process, DOE and MANOVA and Reliability of Service and Products.

In Data Wrangling all the data is sorted, which is then used for obtaining relevant information using Descriptive Statistics. X&s charts are then created and analysed in Statistical Process Control followed by Optimising the Delivery Process in which the delivery process is investigated. Hypotheses are evaluated using MANOVA and finally the Reliability of the Service and Products are determined.

The analysis of each part consists of writing R-code to achieve the required outputs which is then thoroughly inspected and evaluated.

Part 1: Data Wrangling

Part 1 consists of sorting the data and obtaining all the relevant entries. All the invalid records as well as all negative values are removed, therefore exclusively valid data remains. Both the valid and invalid data is stored in two separate files.

Part 2: Descriptive Statistics

There are 180000 entries of data. From these entries we will establish buying patterns, target markets and the capability of the process.

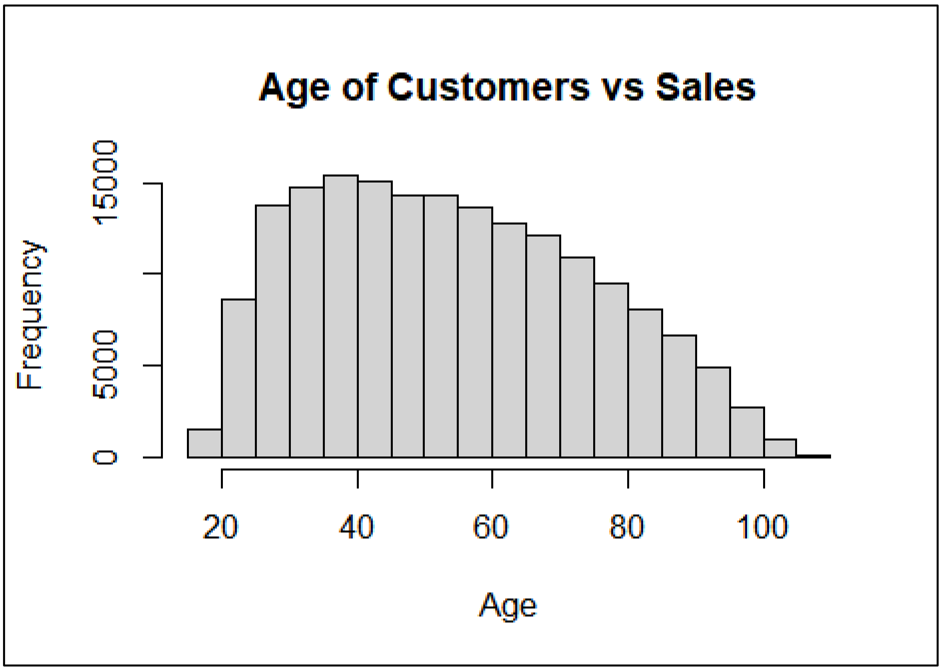

Figure 2.1: Most frequent age groups of consumers

As seen on the histogram above the majority of customers are between the ages of 35 and 40. This would be the target demographic. The histogram is right skewed or is a positively skewed distribution.

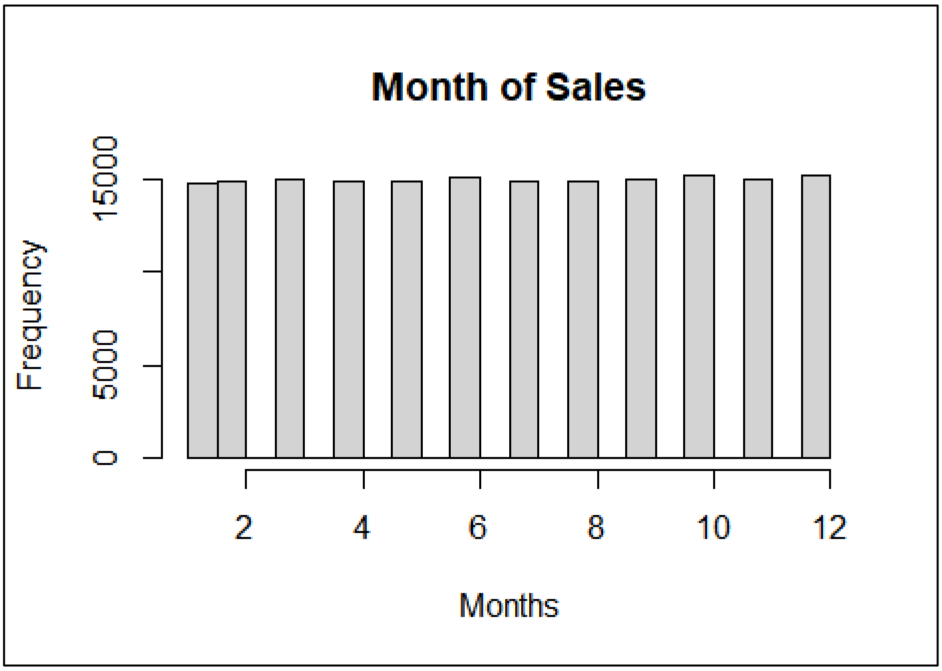

Figure 2.2: Number of sales in each month

Based on the histogram the month in which most sales occurred is December. We should thus ensure the maximum amount of stock is available in December. There is no significant difference in the number of sales per month throughout the year. Stock can therefore not be slacked off on.

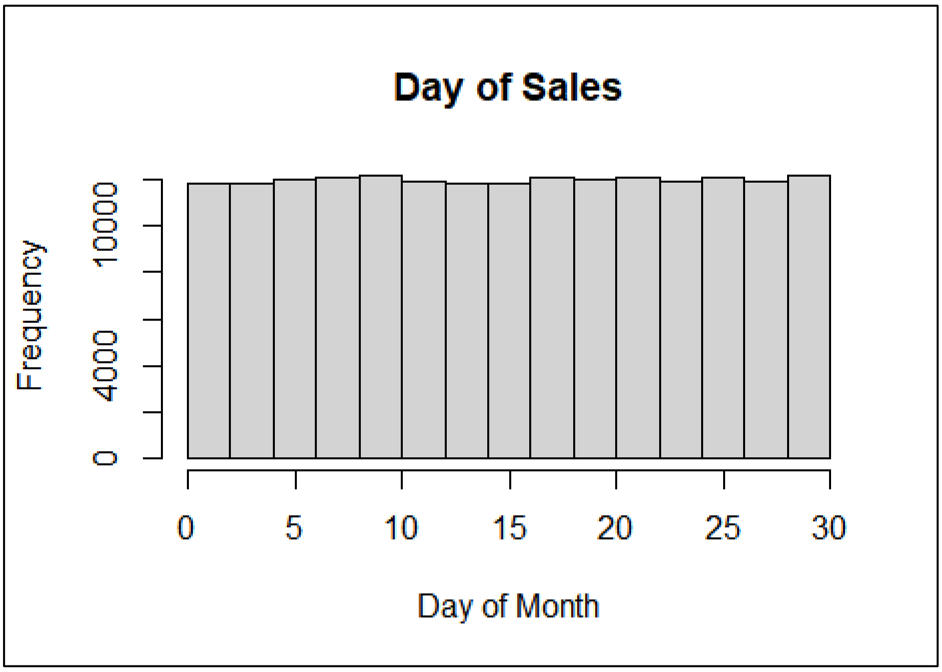

Figure 2.3: Number of sales throughout the month

Based on the histogram we see there is no significant reduction in the number of sales as the month progresses. This prompts us to ensure that we have sufficient safety stock to last throughout the entire month. Gifts are the most frequently bought class and should therefore be the items we stock the most of.

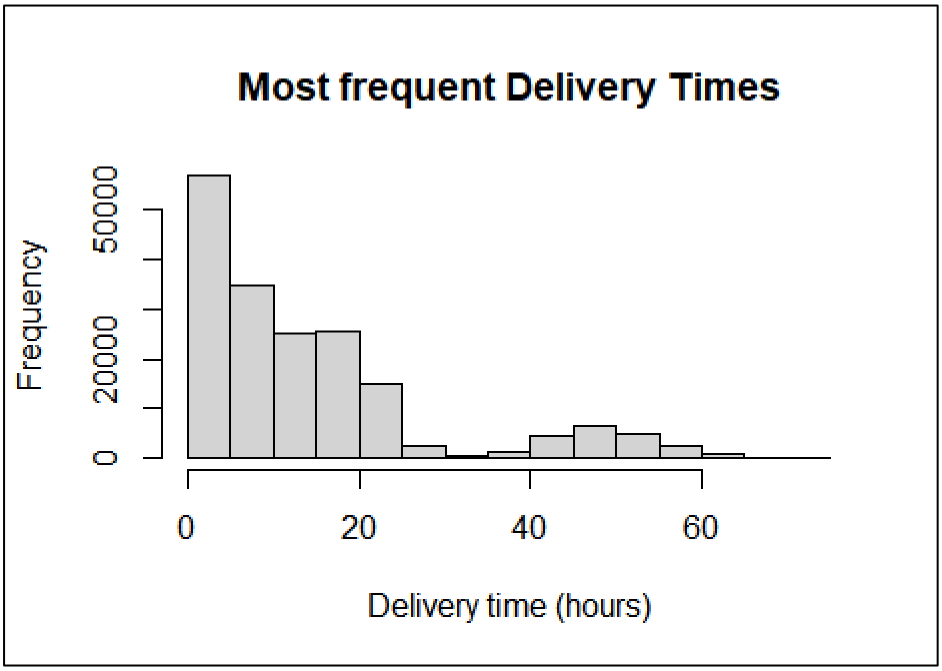

Figure 2.4: Most frequent delivery times in hours

Delivery times range from 0.5-75 hours with the most frequent delivery time of 3 hours and an average of 14.5 hours. Based on the histogram it is clear it is not a normal distribution, and the delivery times are right skewed to the minimum delivery time. The most frequent delivery time of 3 hours is close to the minimum. The average delivery time of 14.5 hours is still found in the lower quartile of the data however its decrease should be investigated.

The main reason customers bought from the business was because it was recommended to them, we should thus focus on keeping our customer satisfaction and engagement high in order to maintain our good reputation and further spread the positive word-of-mouth about the business.

Process capability indices

Cp	1.142
Cpu	0.380
Cpl	1.905
Cpk	0.380

Table 2.1: Process capabilities

Cp is more than one but less than 1.3, therefore the process is capable with tight control. Since Cp does not equal Cpk the process is not centered. Cpl>Cpu therefore the process is more likely to violate the UCL than the LCL. To centre the process the mean must thus be shifted to the left. Cpk is quite low, and the process thus requires improvement. This improvement entails either adjustments in order to centre the process on target or to reduce variation. The LSL cannot be less than zero because time cannot be negative.

Part 3: Statistical Process Control

X-chart

Class	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
Technology	22.975	22.107	21.24	20.374	19.507	18.641	17.774
Clothing	9.405	9.26	9.115	8.97	8.825	8.68	8.535
Household	50.25	49.02	47.79	46.56	45.33	44.10	42.88
Luxury	5.49	5.24	4.99	4.74	4.48	4.23	3.98
Food	2.71	2.64	2.56	2.49	2.42	2.34	2.27
Gifts	9.49	9.11	8.74	8.36	7.99	7.61	7.23
Sweets	2.9	2.76	2.62	2.48	2.34	2.19	2.06

Table 3.1: X-chart values of Delivery Times

s-chart

Class	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
Technology	5.1805	4.552	3.923	3.296	2.667	2.038	1.410
Clothing	0.867	0.761	0.656	0.551	0.446	0.341	0.236
Household	7.34	6.45	5.56	4.67	3.78	2.89	2
Luxury	1.51	1.33	1.14	0.96	0.78	0.59	0.41
Food	0.44	0.38	0.33	0.28	0.23	0.17	0.12
Gifts	2.25	1.97	1.7	1.43	1.16	0.88	0.61
Sweets	0.84	0.73	0.63	0.53	0.43	0.33	0.23

Table 3.2: s-chart values of Delivery Times

After the data has been ordered chronologically both the X&s- bar chart values of the delivery times have been calculated from thirty samples of fifteen sales each. These values can be found in Table 3.1 and Table 3.2 above.

Example of Initialised X-chart

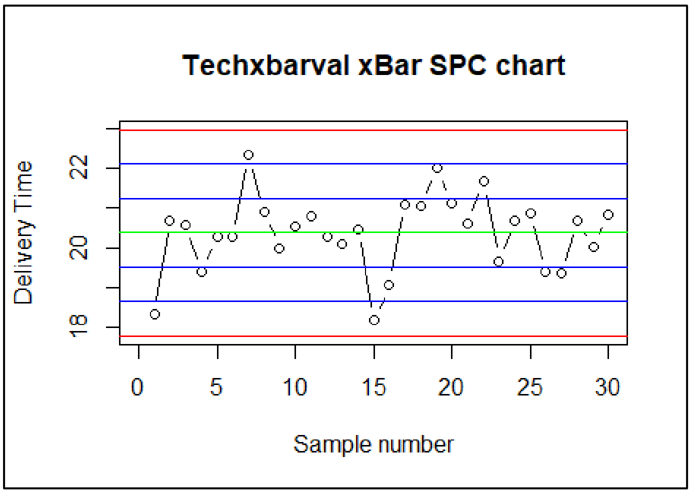

Figure 3: Example of initialised x chart with UCL, U2Sigma, U1Sigma, CL, L1Sigma, L2Sigma, LCL

Using the data in the above tables graphs for each class was constructed as seen below

3.1 Sample of thirty

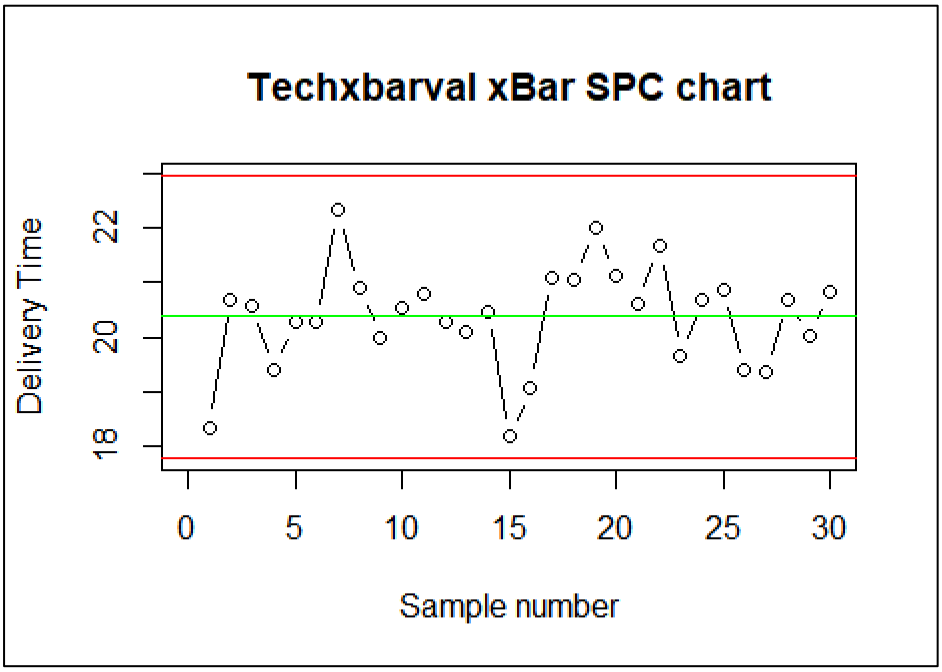

Figure 3.1: Technology X-bar chart

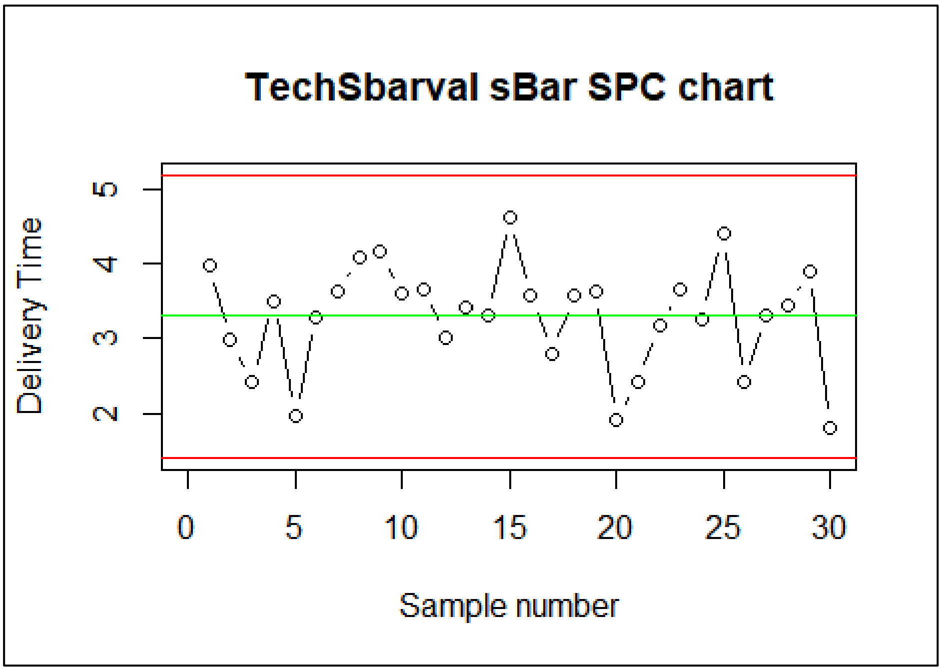

Figure 3.2: Technology s-bar chart

The s-bar chart for technology is in control which allows us to correctly interpret the X-bar chart. The X-bar chart indicates that the technology delivery process times are in control.

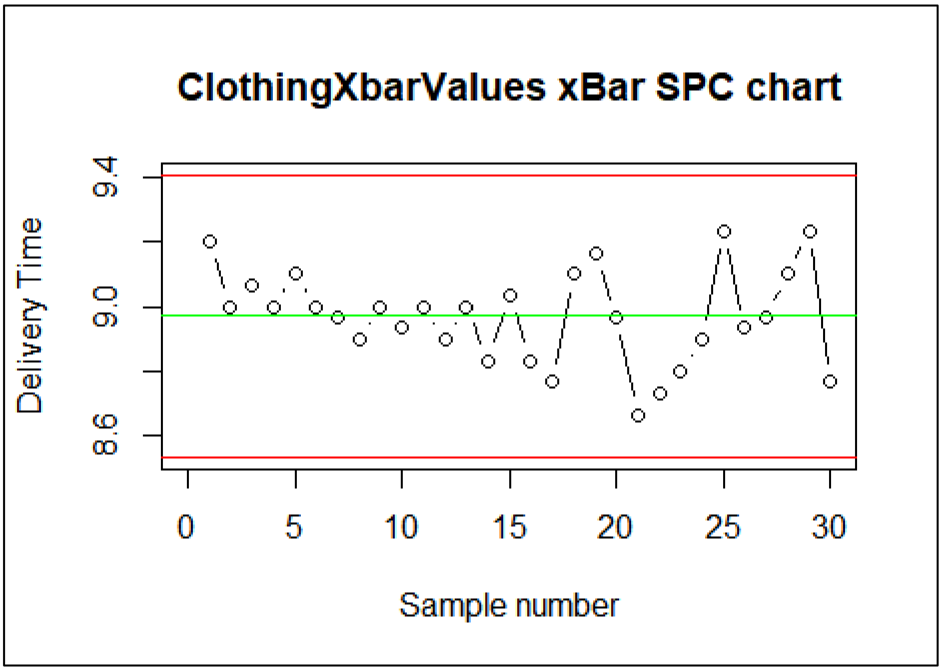

Figure 3.3: Clothing X-bar chart

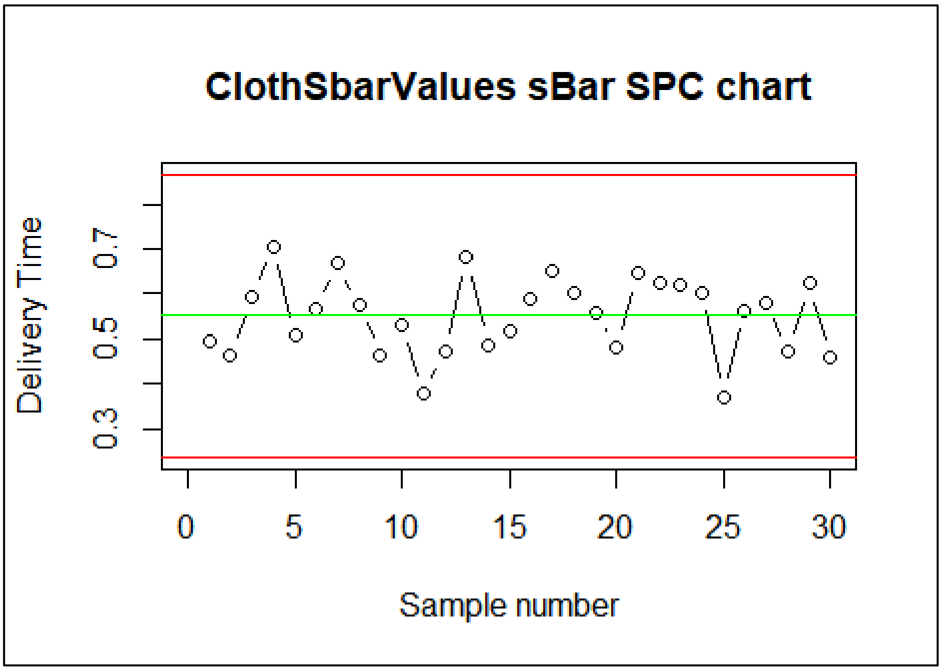

Figure 3.4: Clothing s-bar chart

The s-bar chart for clothing is in control which allows us to correctly interpret the X-bar chart. The X-bar chart indicates that the clothing delivery process times are in control.

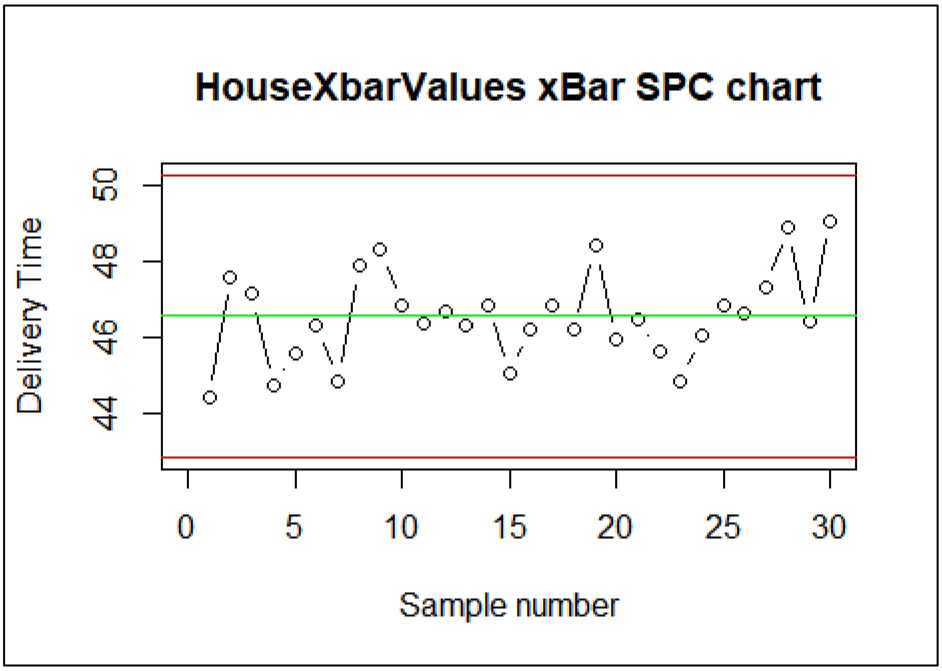

Figure 3.5: Household X-bar chart

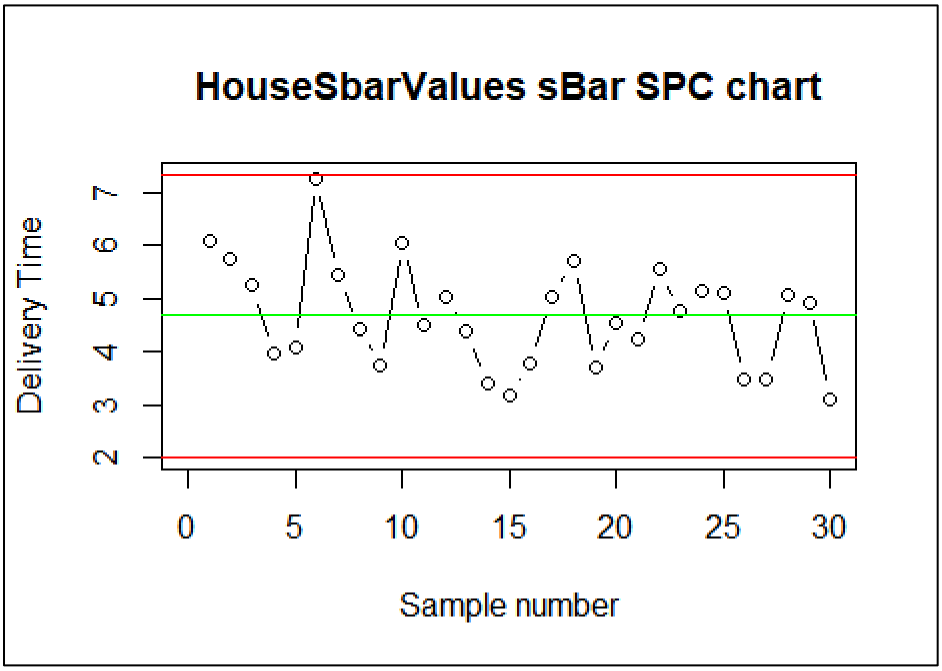

Figure 3.6: Household s-bar chart

The s-bar chart for household is in control which allows us to correctly interpret the X-bar chart. The X-bar chart indicates that the household delivery process times are in control.

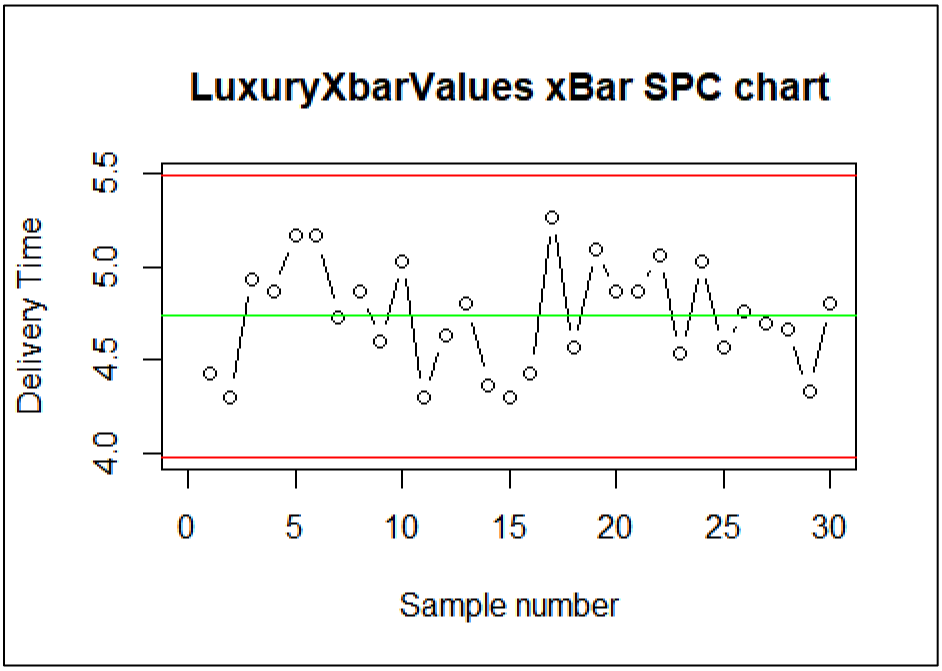

Figure 3.7: Luxury X-bar chart

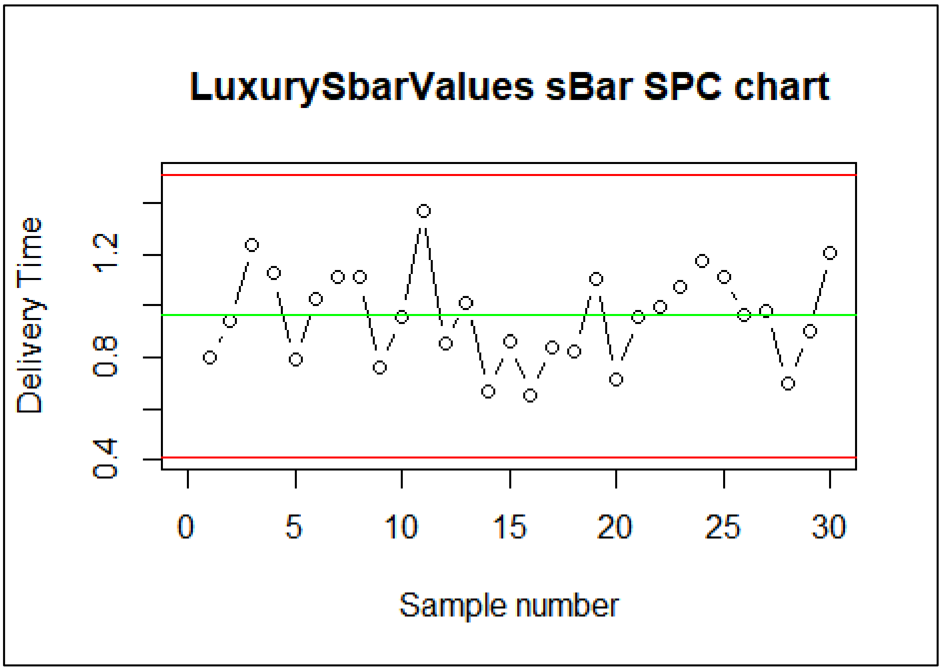

Figure 3.8: Luxury s-bar chart

The s-bar chart for Luxury is in control which allows us to correctly interpret the X-bar chart. The X-bar chart indicates that the luxury delivery process times are in control.

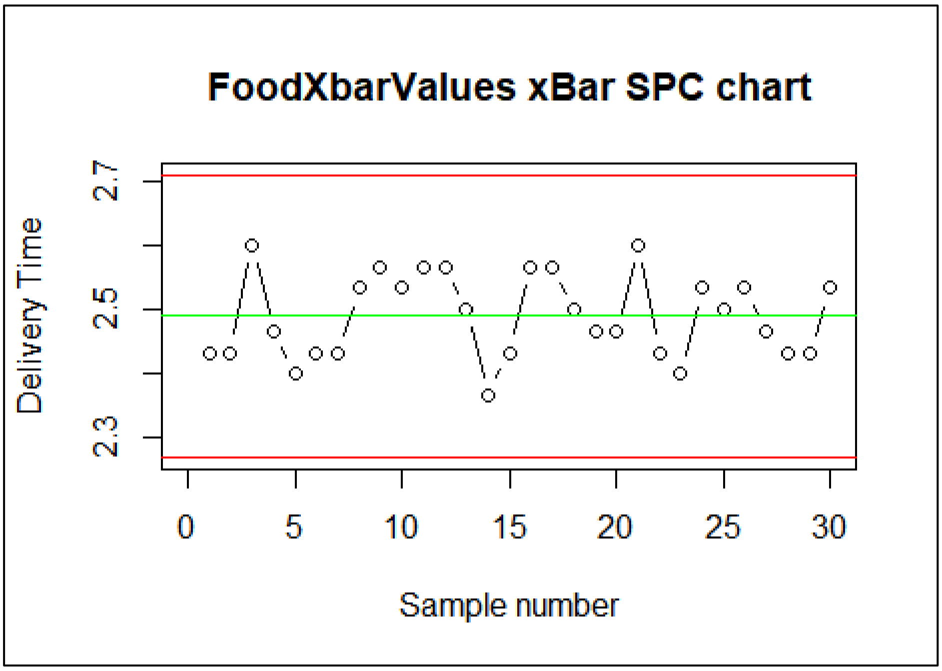

Figure 3.9: Food X-bar chart

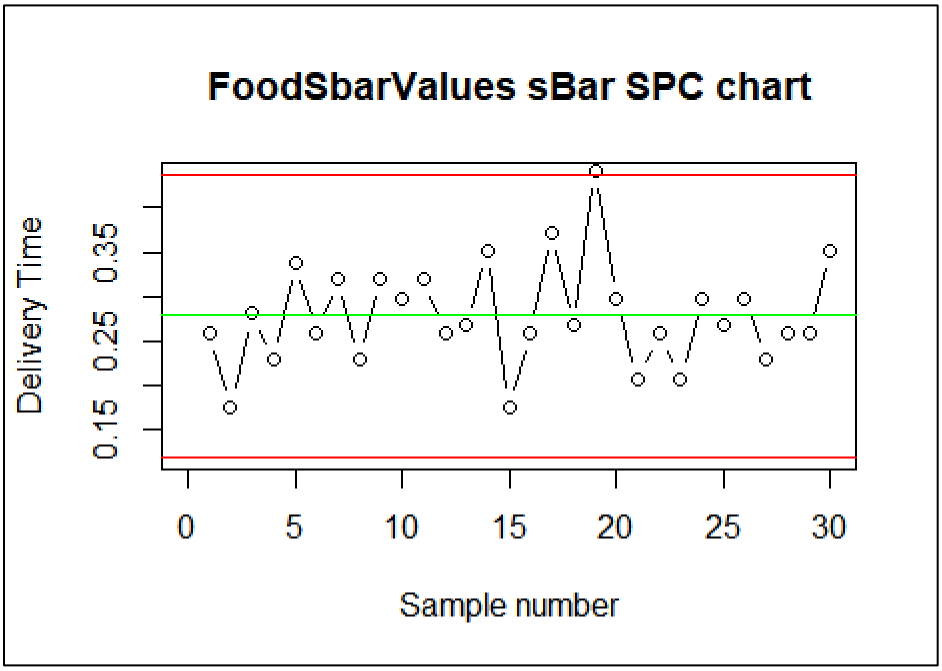

Figure 3.10: Food s-bar chart

The s-bar chart for food seems in control, however sample nineteen's standard deviation is not within the limits. Sample nineteen should thus be removed. After the removal the X-chart would still display an in-control process.

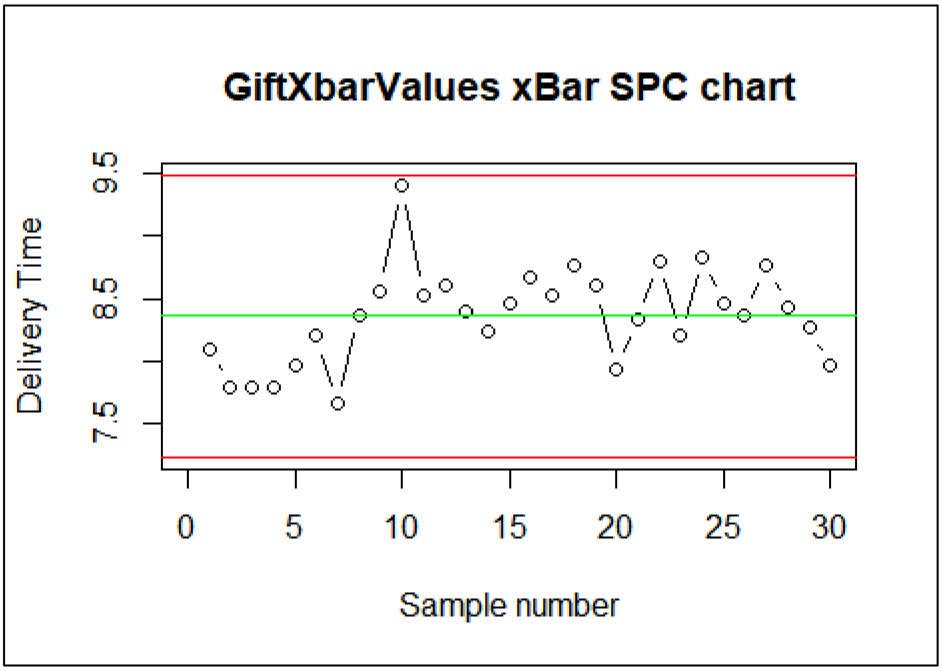

Figure 3.11: Gift X-bar chart

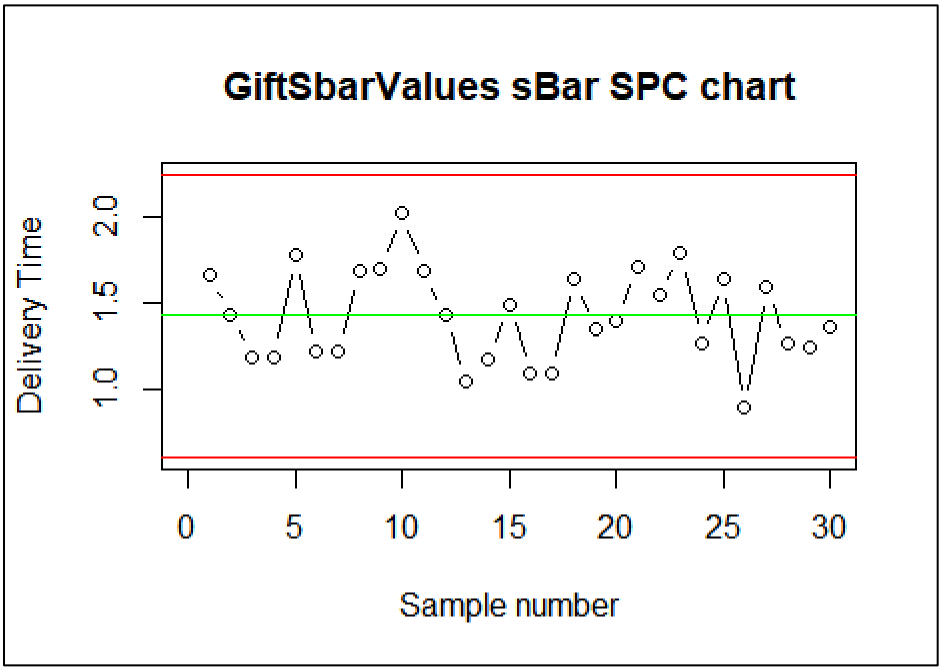

Figure 3.12: Gift s-bar chart

The s-bar chart for gifts is in control which allows us to correctly interpret the X-bar chart. The X-bar chart indicates that the gifts delivery process times are in control.

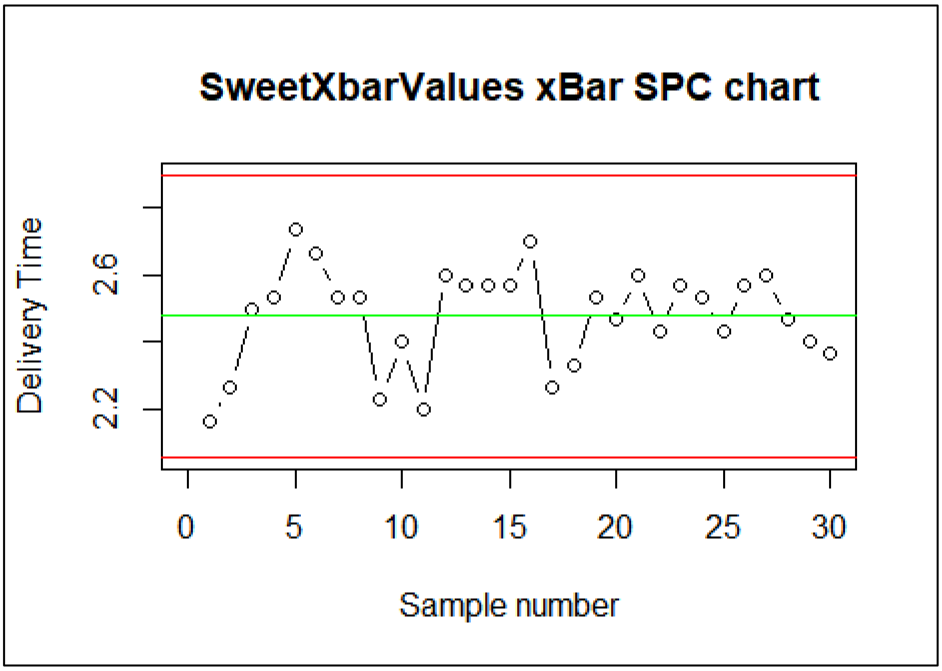

Figure 3.13: Sweets X-bar chart

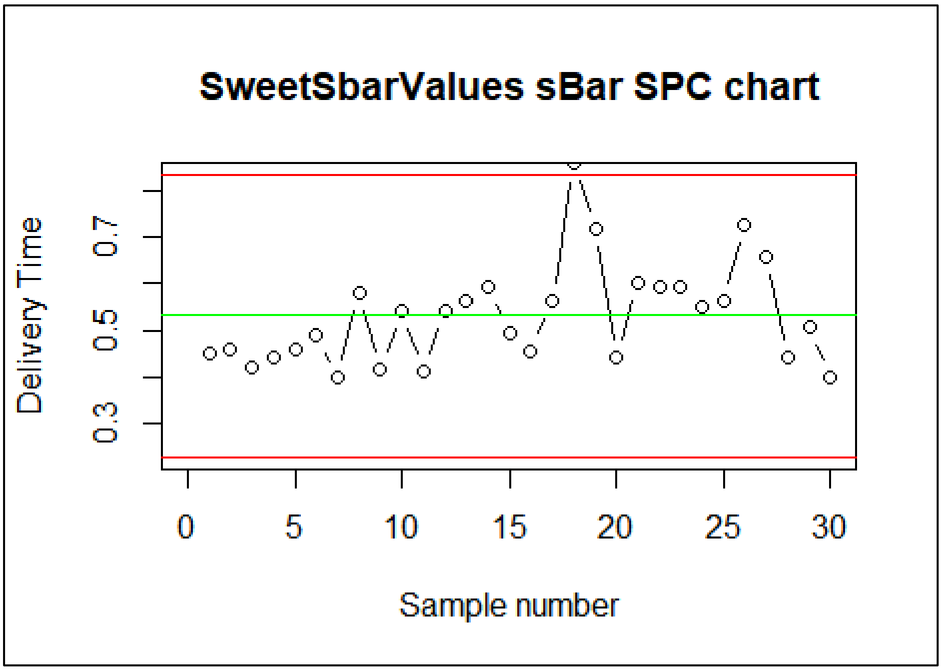

Figure 3.14: Sweets s-bar chart

The s-bar chart for sweets seems in control, however sample eighteen's standard deviation is not within the limits. Sample eighteen should thus be removed. After its removal, the X-chart would still display an in-control process.

3.2 Entire samples

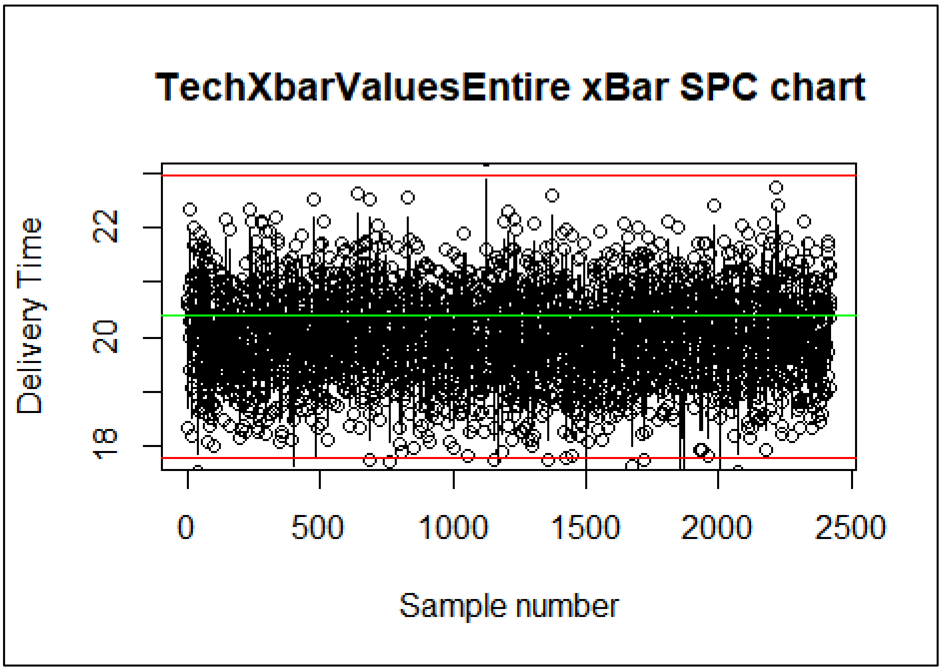

Figure 3.15: Technology X-bar chart

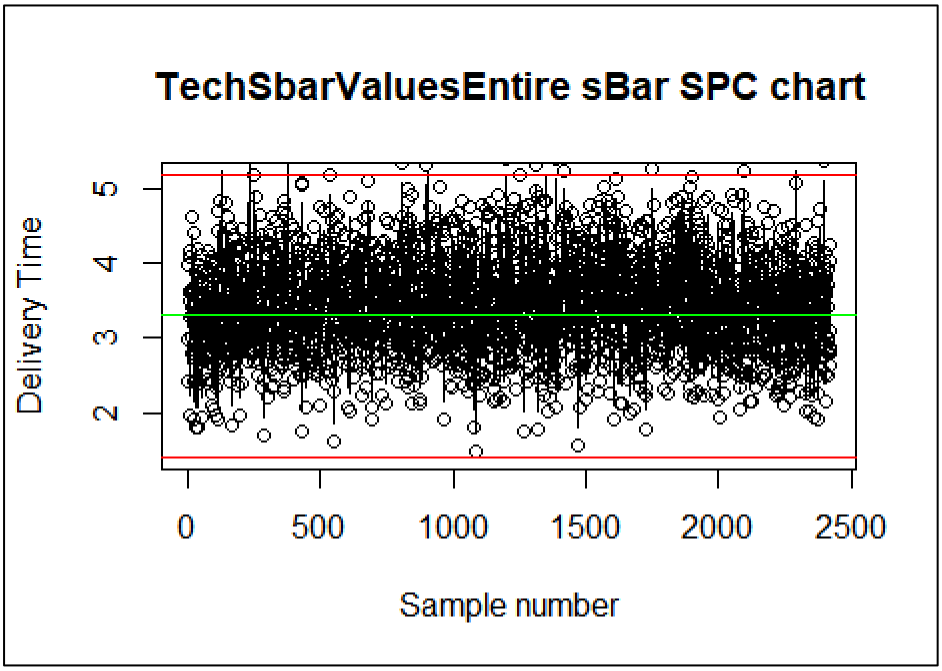

Figure 3.16: Technology S-bar chart

The s-bar chart only has sixteen standard deviations outside the limits. The interpretation of the X-chart would thus be satisfying, even though these sixteen samples should be removed. The majority of the X-chart is within the limits, except for seventeen samples, and technology process delivery times seems to be in control.

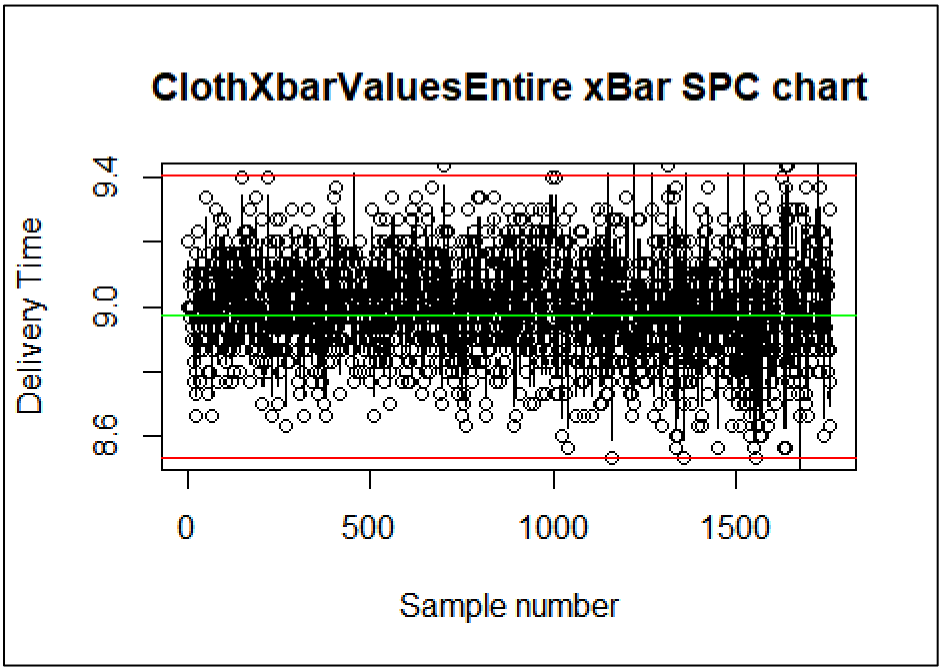

Figure 3.17: Clothing X-bar chart

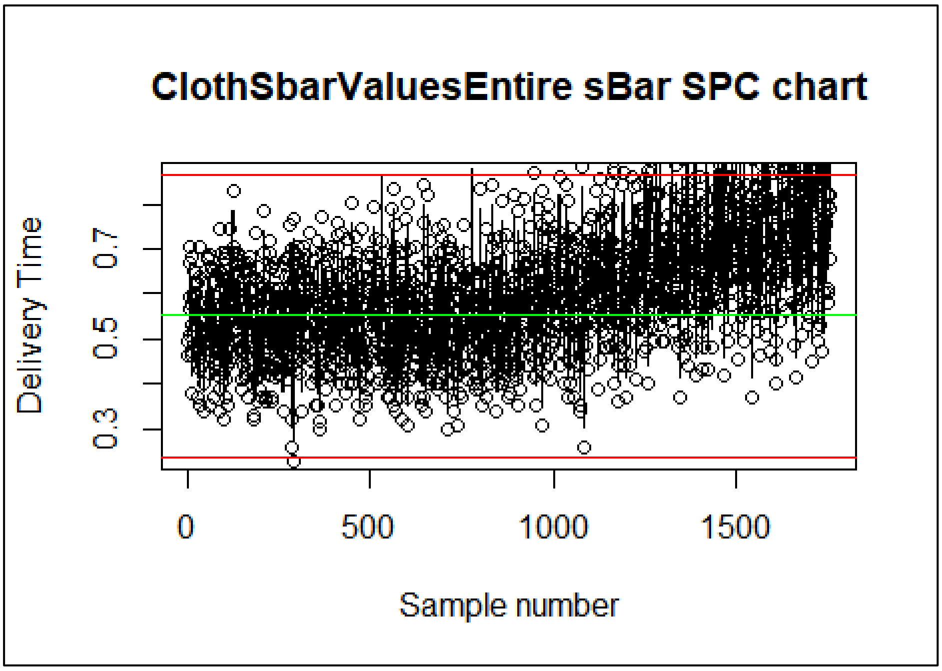

Figure 3.18: Clothing S-bar chart

According to the s-chart 98 samples are outside the limits. These have seemed to occur more recently as a slight upward trend towards the UCL has been observed and should be investigated and the samples removed. Even though the majority is still in control, the interpretation of the X-bar chart would be adequate at best. These samples should be removed for a more accurate interpretation of the X-bar chart. Keeping this in mind the X-chart still displays the majority samples within the limits with the exception of seventeen cases, and therefore ought to be in control.

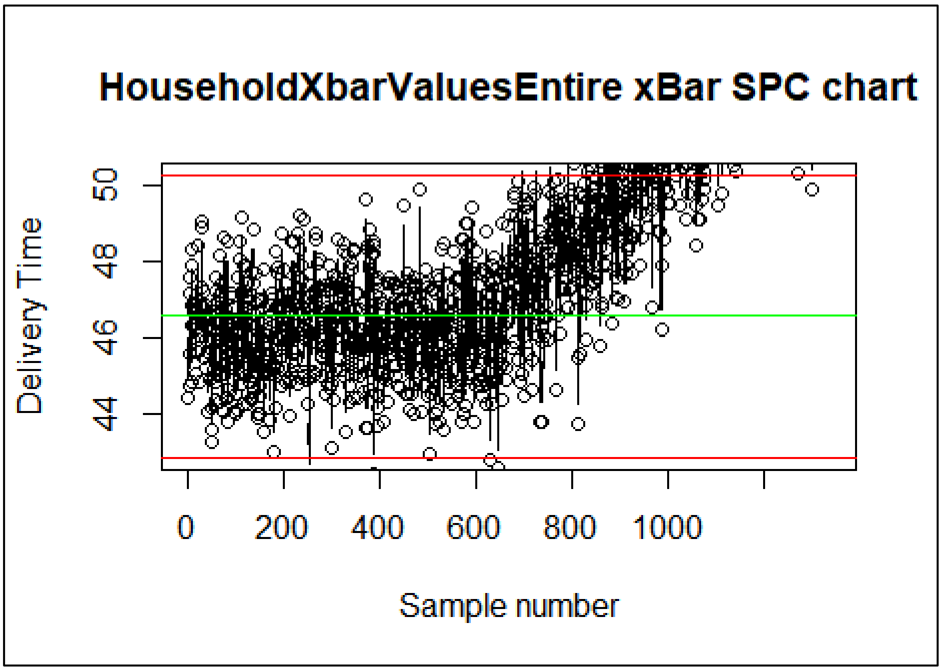

Figure 3.19: Household X-bar chart

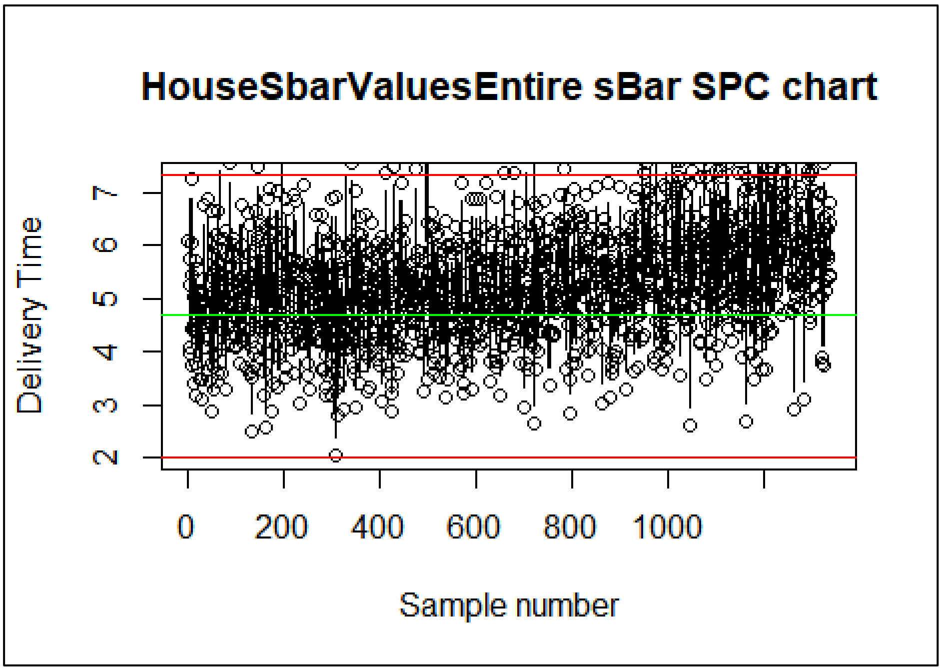

Figure 3.20: Household S-bar chart

The s-chart has fifty-four standard deviations outside the limits. The interpretation of the X-chart would therefore be somewhat satisfying. The X-chart displays a significant increase in delivery times and has four hundred samples outside the limits. This has occurred more recently and must be investigated immediately as the process is not in control at all.

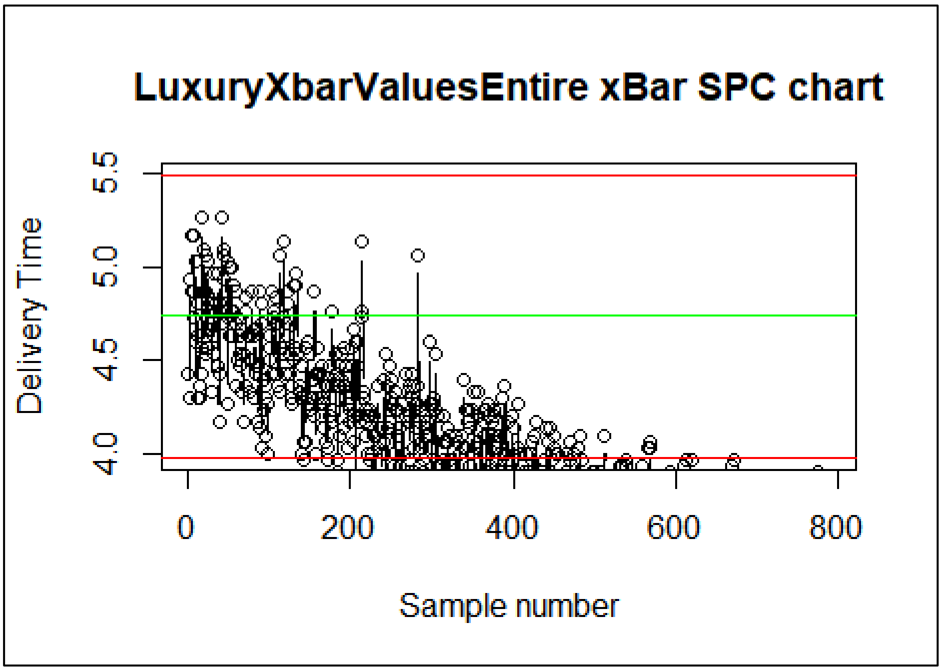

Figure 3.21: Luxury X-bar chart

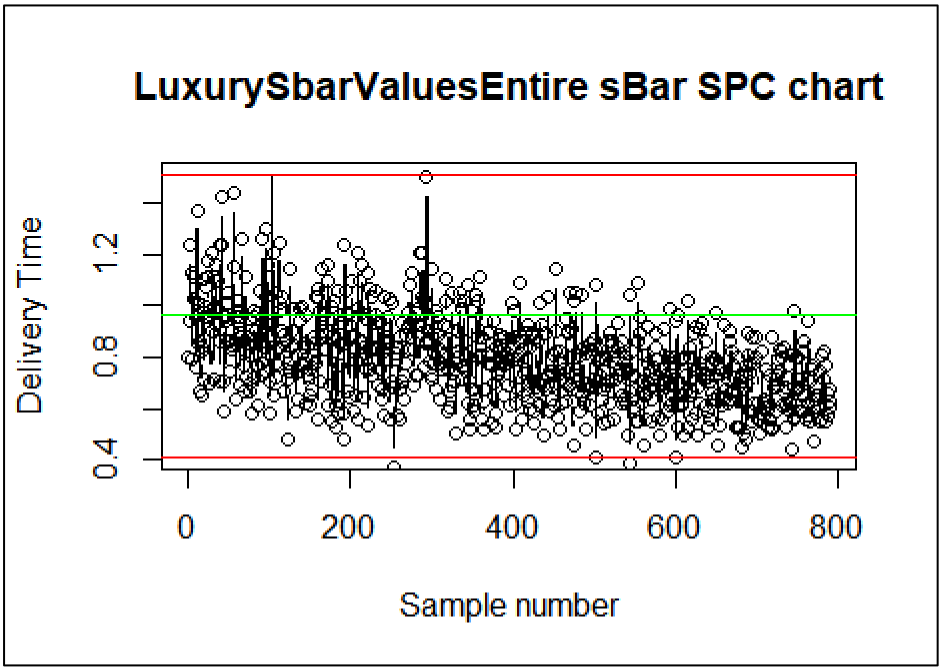

Figure 3.22: Luxury S-bar chart

The s-chart shows only four samples outside the limits. This shows our interpretation of the X-chart would be satisfying. The X-chart shows a significant decrease in delivery times with four hundred and thirty-four samples outside the limits. This decrease must be investigated as the process is not in control at all.

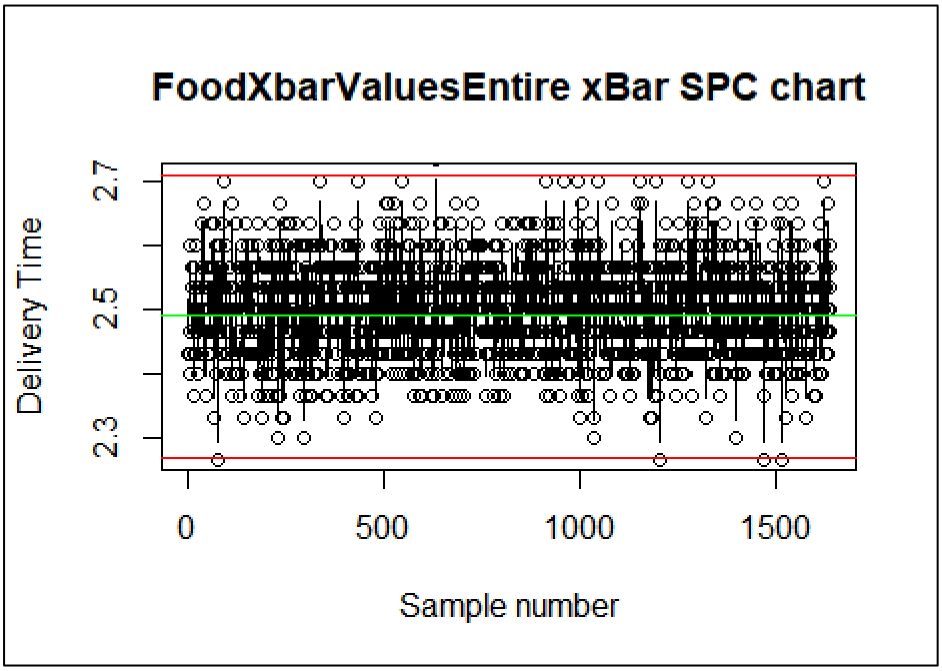

Figure 3.23: Food X-bar chart

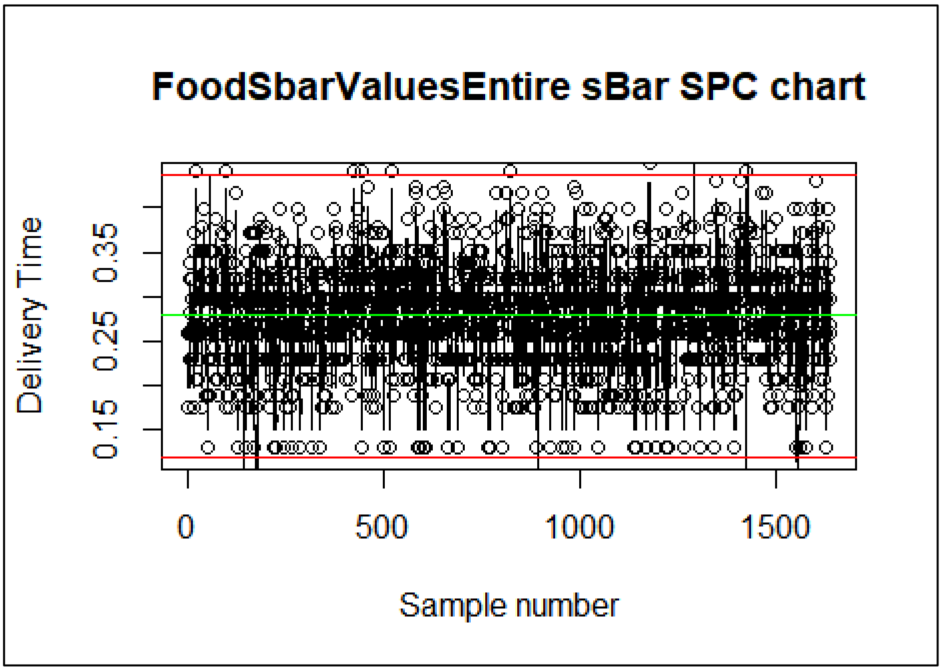

Figure 3.24: Food S-bar chart

The s-bar chart shows only sixteen standard deviations outside the limits. The interpretation of the X-chart would still be appropriate. The X-chart shows that the process is in control with only five samples outside the limits.

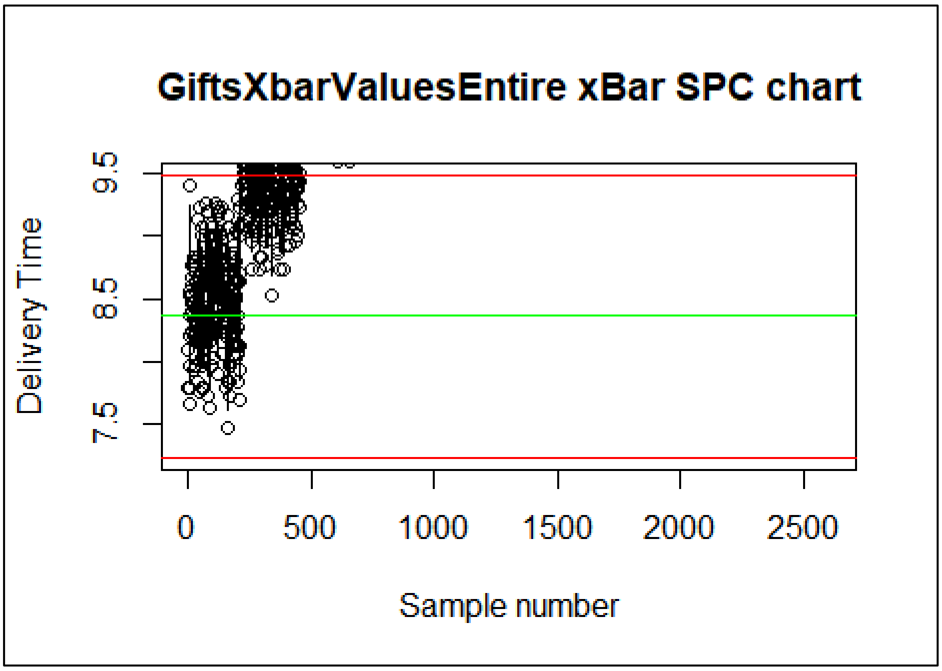

Figure 3.25: Gifts X-bar chart

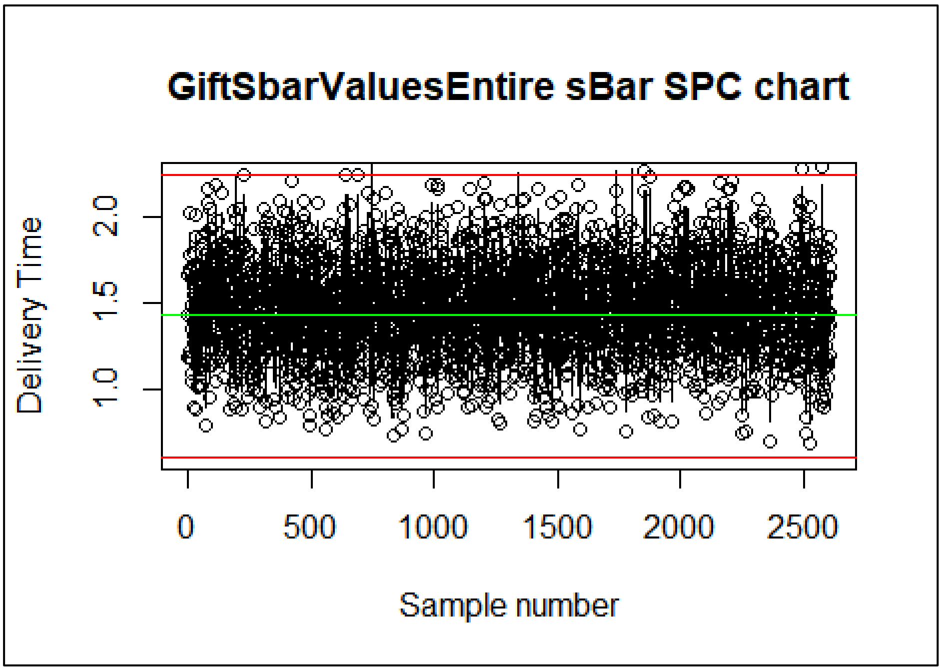

Figure 3.26: Gifts S-bar chart

The s-chart shows only eight samples outside the limits. The interpretation of the X-chart would thus be satisfying. The X-chart shows a significant increase in delivery times with 2290 samples outside the limits. The process is not in control at all, and an immediate investigation should be launched to find the root of the problem.

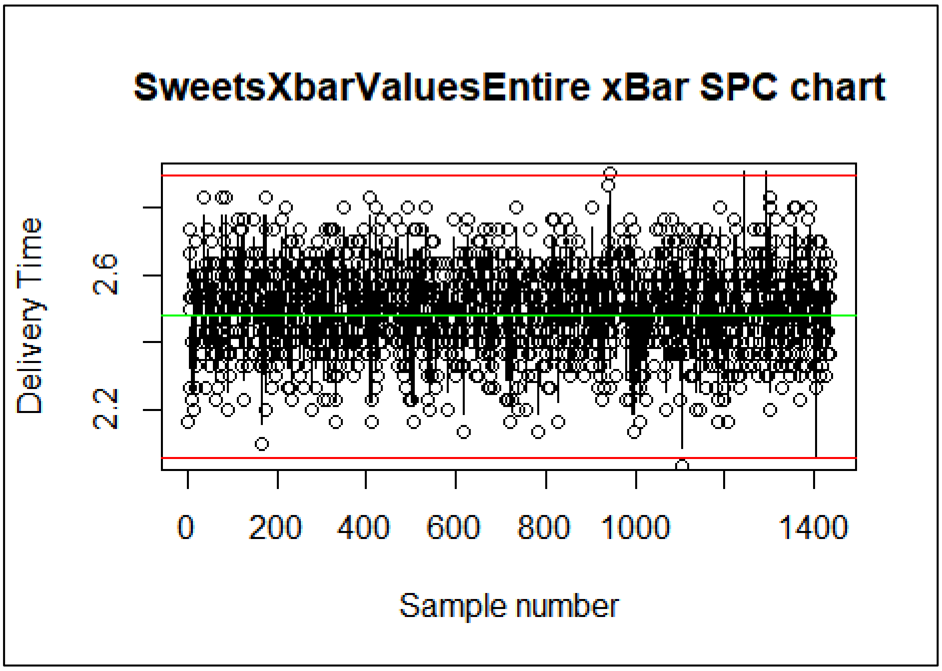

Figure 3.27: Sweets X-bar chart

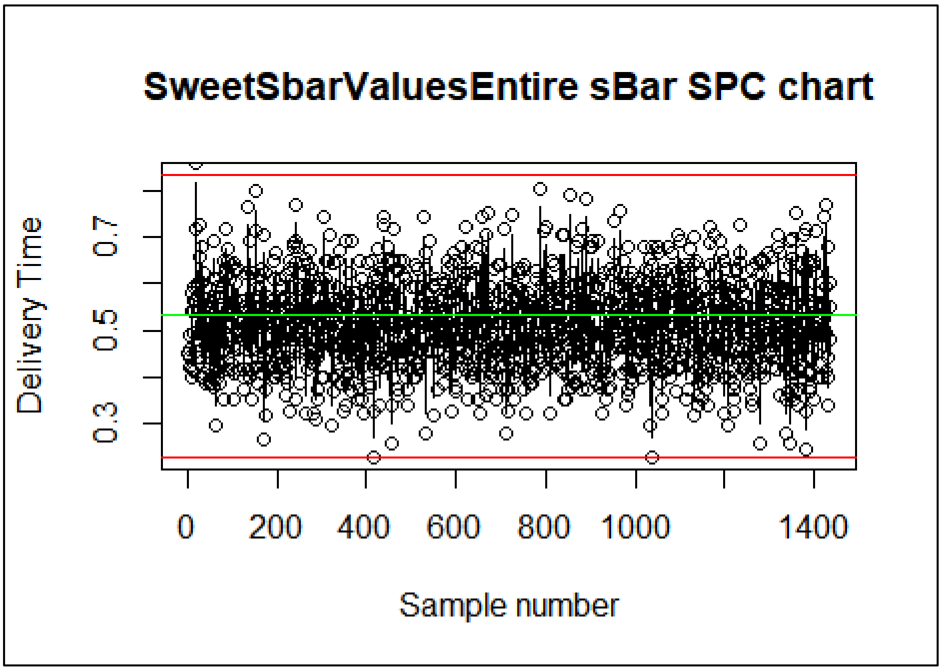

Figure 3.28: Sweets S-bar chart

The s-chart shows only one sample outside the limits. The interpretation of the X-chart is thus satisfying. The X-chart shows only five samples outside the limits. The process is thus in control.

Part 4: Optimising delivery process

4.1 A

Technology

	1st	2nd	3rd	Third from last	Second form last	Last	Total outside limit
Under LCL	37	398	483	1872	2009	2071	16
Above UCL	1122	-	-	-	-	-	1

Table 4.1: Samples outside the control limits

Clothing

	1st	2nd	3rd	Third from last	Second from last	Last	Total outside limit
Under LCL	1161	1359	1557	-	-	1677	4
Above UCL	455	702	1152	1644	1723	1724	13

Table 4.2: Samples outside the control limits

Household

	1st	2nd	3rd	Third from last	Second from Last	Last	Total outside limit
Under LCL	252	387	629	-	-	643	4
Above UCL	693	725	752	1335	1336	1337	396

Table 4.3: Samples outside the control limits

Luxury

	1st	2nd	3rd	Third from last	Second from Last	Last	Total outside limit
Under LCL	142	171	184	789	790	791	434
Above UCL	-	-	-	-	-	-	0

Table 4.4: Samples outside the control limits

Food

	1st	2nd	3rd	Third from last	Second from last	Last	Total outside limit
Under LCL	75	1203	1467	-	-	1515	4
Above UCL	633	-	-	-	-	-	1

Table 4.5: Samples outside the control limits

Gifts

	1st	2nd	3rd	Third from last	Second from Last	Last	Total outside limit
Under LCL	-	-	-	-	-	-	0
Above UCL	213	216	218	2607	2608	2609	2290

Table 4.6: Samples outside the control limits

Sweets

	1st	2nd	3rd	Third from last	Second from last	last	Total outside limit
Under LCL	1104	-	-	-	-	1403	2
Above UCL	942	1243	-	-	-	1294	3

Table 4.7: Samples outside the control limits

Based on these tables Technology, Clothing, Food and Sweets have very few samples outside the control limits and therefore are very much in control. Household, Luxury and Gifts have too many samples outside the limits to qualify for in control. Further investigation is required to establish the root of the problem and implement a solution.

4.1 B

Class	**Most Consecutive**
Technology	6
Clothing	4
Household	3
Luxury	4
Food	4
Gifts	7
Sweets	5

Table 4.8: Most consecutive sample standard deviations between -0.3 and 0.4 sigma control limits

From Table 4.8 it is clear that the maximum consecutive samples are seven which stems from the class Gifts. This is a low number and therefore many samples occur outside the limits of -0.3 and 0.4 sigma control limits.

4.2 Type l Error

Ho: Process is in control and centred on centreline.
Ha: Process is not in control and has moved from centreline or changed in variation.

	Probability of making a Type 1 (Manufacturer's) Error
A	0.002699796 = 0.27%
B	0.7266668 = 72.67%

Table 4.9: Probability of making a Type 1 (Manufacturer's) Error

For A there is a 0.27% chance that the company would believe the process is out of control when in reality it is in control.

For B there is a 72.67% chance that the company would believe the process is out of control when in reality the process is in control.

4.3

The current total sales slower than 26 hours are 1356. Given that we lose R329/item-late-hour the business loses R446 124. The mean delivery time for technology is 20 hours. At a rate of R2.5/item/hour to reduce the average time by one hour over 26 hours the total cost would be R636 072.5 to move the mean.

4.4

The probability of making a Type ll error is 0.4884959. This means there is a 48.85% chance that the company believes the process is in control while in reality it is out of control. This percentage is quite high and is cause for concern. We must ensure that the deliveries are made on time and not just assume that they have been delivered on time.

Part 5: MANOVA

5.1 First Hypothesis

Ho	Age and Price does not influence the sales patterns of each class.
Ha	At least one of Age or Price influences the sales patterns of each class
p-value	2.2e-16 (2.2e-16 < 0.05)
Reject Null Hypothesis	

Table 5.1: First Hypothesis

The p-value demonstrates that Age and Price are highly significantly different among each class. The Null hypothesis is thus rejected as at least one of Age or Price influences the sales patterns.

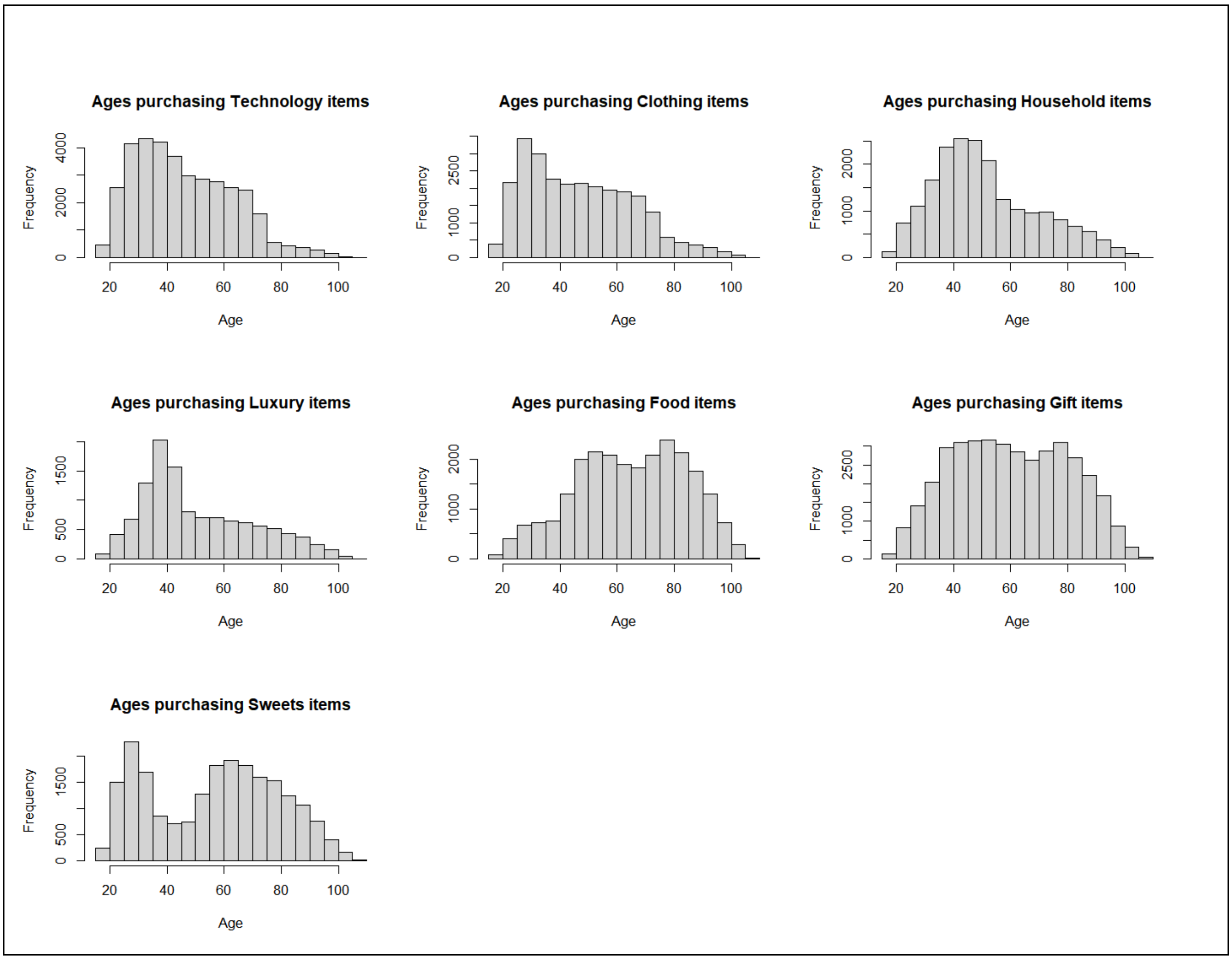

Figure 5.1: Purchase records of different ages for each class

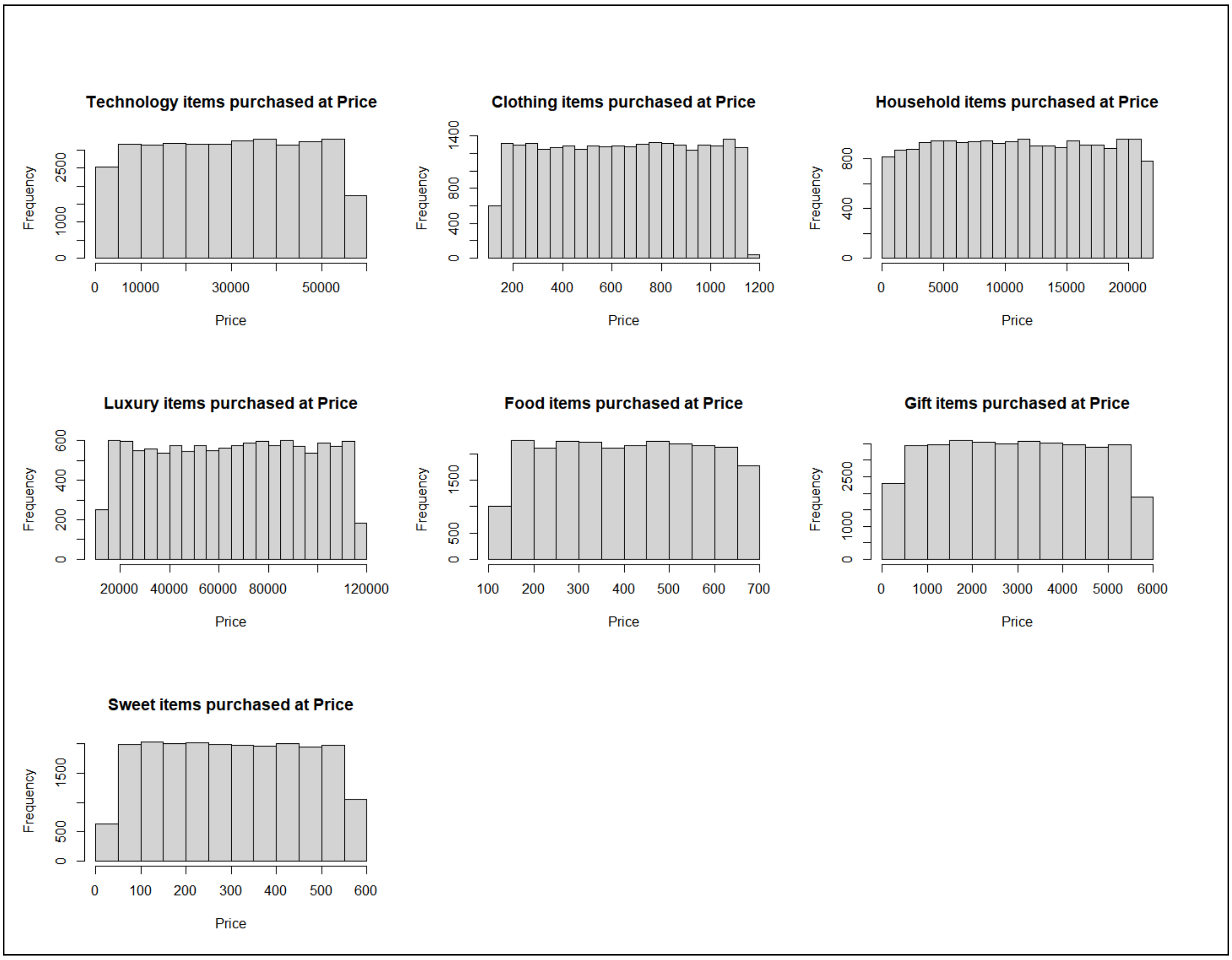

Figure 5.2: Number of items of each class bought at a certain price

From Figure 5.1 we can see that the ages between classes vary considerably, indicating that different age groups are interested in different class of items. The ages of consumers thus have an effect on the buying patterns of each class.

Figure 5.2 shows that price has some effect on the buying patterns however it is slight and much less influential than age. In all the cases the cheapest and most expensive items were purchased less than the other prices. This could indicate consumers avoiding the less expensive items out of fear for its lacking quality and avoiding the most expensive items simply because it is too expensive. This indicates that price has an effect on buying patterns.

5.2 Second Hypothesis

Ho	Month and Day does not influence the sales patterns of each class.
Ha	At least one of Month or Day influences the sales patterns
p-value	0.2076 (0.2076 > 0.05)
Fail to reject Null Hypothesis	

Table 5.2: Second Hypothesis

The p value indicates that Month and Day are not significantly different among each class, and we thus fail to reject the Null Hypothesis. Therefore, Month and Day do not influence the buying patterns of each class.

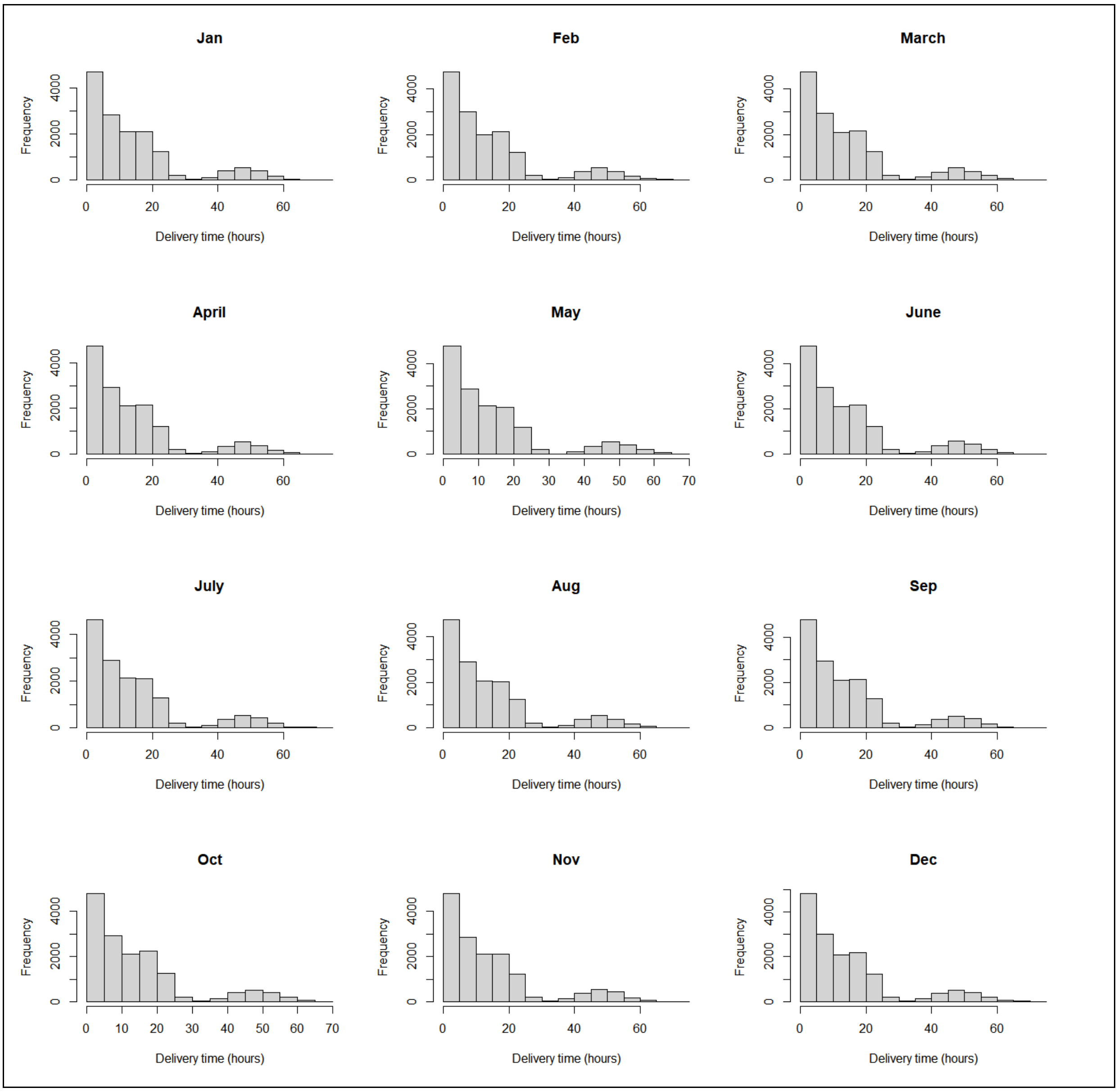

Figure 5.3: Delivery times for different months

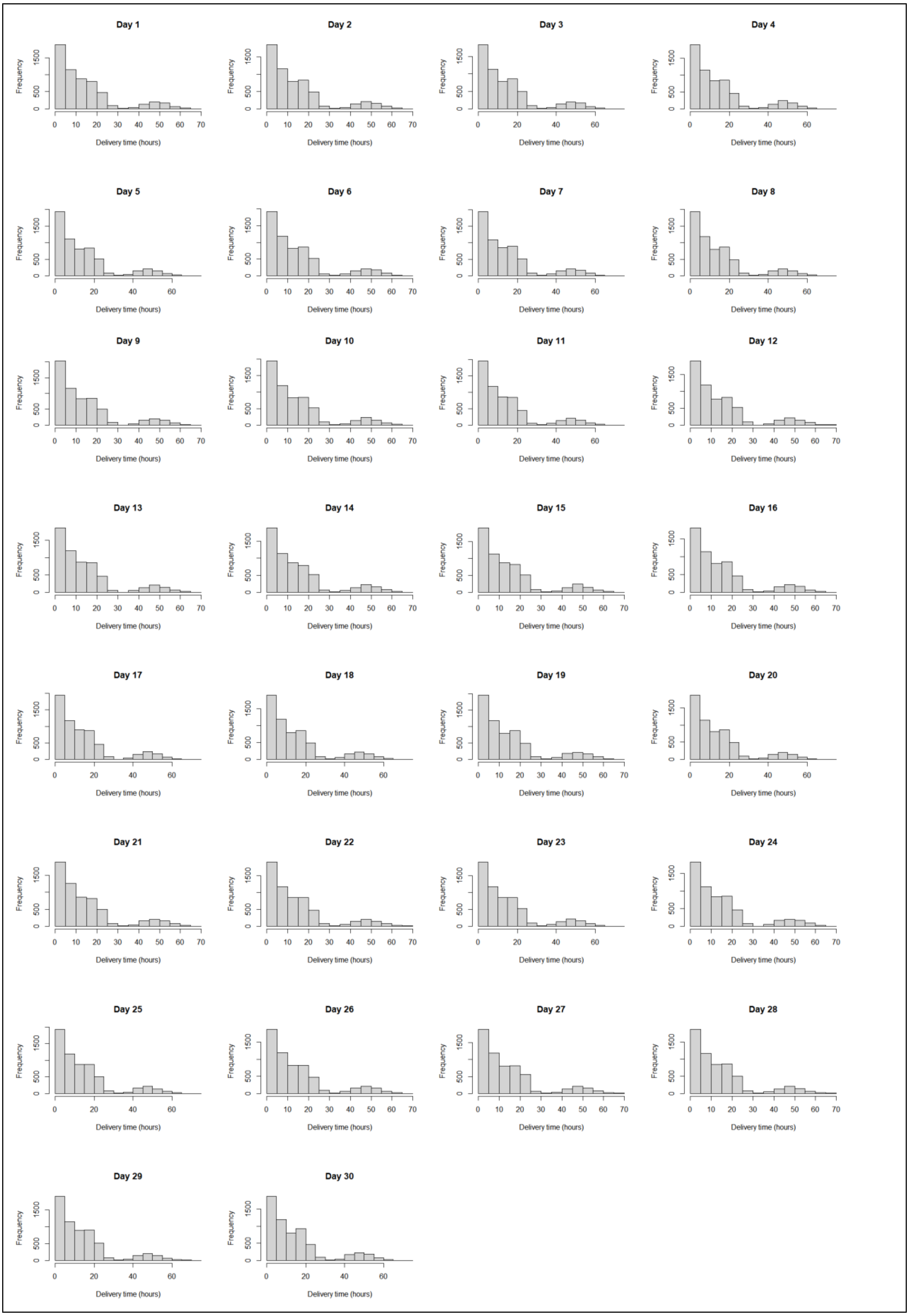

Figure 5.4: Delivery times for different days of the month

From Figure 5.3 and 5.4 above we can see there is almost no variability between the graphs, and all look the same. This is because neither the Month nor the Day has an influence on the Delivery times which is why we fail to reject the Null hypothesis.

Part 6: Reliability of service and products

6.1 A

k = 45/(0.04²) = 28125
L = 28125*(y-0.06)²

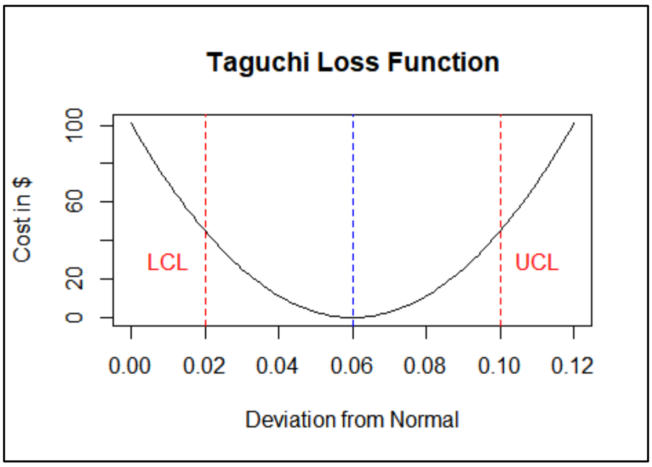

Figure 6.1: Taguchi loss function .

6.1 B (Question 7a)

k = 35/(0.04²) = 21875
L = 21875*(y-0.06)²

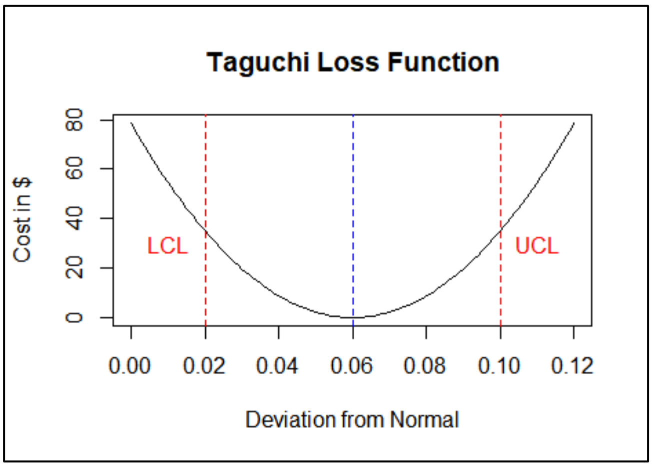

Figure 6.2: Taguchi loss function

According to both Figure 6.1 and 6.2 the target value is 0.06. The further a product deviates from 0.06 the less reliable it becomes and the greater the loss and customer dissatisfaction. This is not an immediate drop in quality rather a gradual decline.

6.1 C (Question 7b)

L = 21875*0.027^2 = 15.95

The business loses $15.95/item if the process deviation is reduced to 0.027cm

6.2 A (Question 27a)

Ra*Rb*Rc = 0.85*0.92*0.9 = 0.7038

By using only one machine at each stage the system is reliable 70.38% of the time.

6.2 B (Question 27b)

	A	B	C
Both fail	$(1-0.85)^2$ = 0.0225	$(1-0.92)^2$ = 0.0064	$(1-0.9)^2$ = 0.01
Combined reliability	1-0.0225 = 0.9775	1-0.0064 = 0.9936	1-0.01 = 0.99

Raa*Rbb*Rcc = 0.9775*0.9936*0.99 = 0.9615

By using two identical machines at each stage the reliability increases by 25.77% and the system is now reliable 96.15% of the time.

6.3 Part 1

Vehicles:

P(0) = (21C0 * p * $(1-p)^{21}$) * 1560 = 1344
p = 0.0070711

P(1) = (21C1 * p^1 * $(1-p)^{20}$) * 1560 = 190
p = 0.0066243

P(2) = (21C2 * p^2 * $(1-p)^{19}$) * 1560 = 22
p = 0.0089232

P(3) = (21C3 * p^3 * $(1-p)^{18}$) * 1560 = 3
p = 0.0121700

P(4) = (21C4 * p^4 * $(1-p)^{17}$) * 1560 = 1
p= 0.0196857

Weighted Average (p): 0.007060693

Reliable Delivery Times based on Vehicles:

P(X<=2) = 0.9995744

P(X<=2) *365 = 364.8447

Drivers:

$P(0) = (21C0 * p^0 * (1-p)^{21}) * 1560 = 1458$
p = 0.0032148

$P(1) = (21C1 * p^1 * (1-p)^{20}) * 1560 = 95$
p = 0.0030848

$P(2) = (21C2 * p^2 * (1-p)^{19}) * 1560 = 6$
p = 0.0044655

$P(3) = (21C3 * p^3 * (1-p)^{18}) * 1560 = 1$
p = 0.0082395

Weighted Average (p): 0.003214915

Reliable Delivery Times based on Drivers:

P(X<=2) = 0.9999577

P(X<=2) * 365 = 364.9846

Reliable Days:

P(Vehicles) * P(Drivers) = 0.9995744*0.9999577 = 0.9995321

365 * 0.9995321 = 364.8292 days

This means that out of 365 days the business has 365 days where delivery would be reliable.

6.3 Part 2

p for all vehicles remains the same. Therefore, the weighted average for the vehicles will also remain the same.

$P(X<=3) = (22C3 * p^3 * (1-p)^{19})*365 + (22C2 * p^2 * (1-p)^{20})*365 + (22C1 * p^1 * (1-p)^{21})*365 + (22C0 * p^0 * (1-p)^{22})*365$

= 364.994 = 365 days

Increasing the number of vehicles with one will not have an effect on the reliability. The reliability out of 365 days would still be 365 days.

Conclusion

The in-depth analysis offered in this report shows how the interpretation of the appropriate data can help make strategic evaluations and decisions.

Using Descriptive Statistics, the target market has been identified as customers between the ages of 35-40. It was also determined that the month in which most sales occurred was December , gifts were the most frequently bought item, sales did not decrease significantly as the month progressed and most customers bought at the business because it was recommended to them. Finally, it was determined that the process is capable with tight control however improvement is required.

The X&s charts from Statistical Process Control section identified that the delivery times of the class of Technology, Clothing, Food and Sweets are in control while Household, Luxury and Gifts are not in control and improvement is required.

In the Optimising of the Delivery Process section the conclusion about which classes are deemed in control and out of control in the Statistical Process Control section were confirmed to be accurate. The chance of making a Type l error for sample means outside the outer control limits and sample standard deviations between -0.3 and 0.4 sigma control limits are 0.27% and 72.67% respectively. The chance of making a Type ll error was calculated as 48.85%.

The MANOVA tables were used to confirm that at least one of Age or Price has an effect on sales patterns while neither Month nor Day influence the buying patterns of customers.

In the Reliability of Service and Products section it was determined that the target value for product quality within the Taguchi Loss Function is 0.06. It was also determined that adding two identical machines at each stage of the process will increase the reliability by 25.77%. Finally, considering the probability that our vehicles and drivers would not be available, it was determined that in a year, 365 days of reliable delivery times can be expected. Increasing the number of vehicles by one does not change the 100% reliable delivery times.

References

Bhandari, P., 2022. Type I & Type II Errors | Differences, Examples, Visualizations. https://www.scribbr.com/statistics/type-i-and-type-ii-errors/

Hessing, T., 2014. Process Capability (Cp & Cpk). https://sixsigmastudyguide.com/process-capability-cp-cpk/

Hessing, T., 2019. X Bar S Control Chart. https://sixsigmastudyguide.com/x-bar-s-chart/

'MANOVA Test in R Programming' (2020). GeeksforGeeks. https://www.geeksforgeeks.org/manova-test-in-r-programming/

'Taguchi Loss Function' (2017). Lean Six Sigma Definition. https://www.leansixsigmadefinition.com/glossary/taguchi-loss-function/