10/21/2022

QUALITY ASSURANCE ASSIGNMENT

ECSA GA4

Mahlakoane, KR, Mr [23845538@sun.ac.za]
STELLENBOSCH UNIVERSITY

Abstract

This report is aimed to analyse and manipulate the data to acquired. The data is wrangled to eliminate unwanted values and missing values. This data filtering technique helps us produce clean graphical representations of the data. A code is written on R-studios to produce proper data for graphical representation. Descriptive statics are used to analyse the valid data set. This is made possible by using standard descriptive statistics.

Necessary statistical graphs are produced, and comments are made, to describe trends, interesting correlations, and the significant differences between the features/class. The process capabilities, Cpk, Cp, Cpu, and Cpl on the process delivery time are evaluated. Comments are made to establish the capabilities of the company regarding delivery time based on the process capabilities' values.

R program is used to produce S and X-charts based of a sample of 15, and the total sample/data points. X an S-chart on the classes' vs (UCL, U2Sigma, U1Sigma, CL, L1Sigma, L2Sigma, LCL). Optimisation of the delivery time is done. Exercise from the textbook is done

A conclusion is drawn to establish the company's direction based of the methods implemented throughout the report. Suggestions are given on how to improve the company's operation

Contents

Introduction

To properly analyse the data for this report, data wrangling will be done. Additionally, descriptive statistics will be used to analyse the provided data, and process capability indices for data under the technology class will be calculated. The report then uses statistical process control (SPC) charts for all data classes to assess whether each class's parameters are being met. Hypothesis tests are then conducted after this (multivariate analysis of variables). Exercises from the required module textbook are completed at the end

.

Part 1 Data wrangling.

Data wrangling transforms data into a format that is suitable with the final system, which enhances data usability. It allows to simply schedule and automate the data-flow process and develop data flows rapidly using an intuitive user interface. Data wrangling transforms data into a format that is suitable with the final system, which enhances data usability. It allows to simply schedule and automate the data-flow process and develop data flows rapidly using an intuitive user interface.

Part 2 Descriptive statistics

Business statistics includes descriptive statistics, which not only processes and provides data without deciding who should participate but also broadly explains the data acquired. It alludes to the statistical comprehension of how a statistical science might discover how to tabulate, organize, gather, analyse, interpret, and show data.

A measure of central tendency is a good indication of the businesses position. To measure variation, standard measures of variation are used to evaluate data points that are not coinciding with the normal pattern of the processes taking place. It is ideal to have the lowest possible variation (controllable parameters). Below is a table with some of the descriptive features (5 number summary).

Feature	Min	1st Q	median	3rd Q	max	mean
Price	35.65	482.31	2259.63	15270.97	116618.97	12294.1
Age	18	38	53	70	108.00	54.75
Delivery time	0.5	3	10	15.4	18.5	75

Summary of the categorial features

Feature	Count	card	Mode	ModeFreq	Mode %	2n_Mode	2nd_ModeFreq	2nd_Mode %
Class	179978	0	Gifts	39149	21.75	Technology	36347	20.2
Year	179978	9	2021	33443	18.58	2029	22475	12.49
Month	17997	12	12	15225	8.46	10	15221	8.46

	8							
Day	17997 8	30	17	6126	3.4	25	6122	3.4
Why.Boug ht	17997 8	0	Recommend ed	106985	59.44	website	29447	16.36

In addition to the table above, graphs are produced to visualise the trend and distribution of the different features of the given sales data.

Box plots are presented to elaborate on the skewness of the data set.

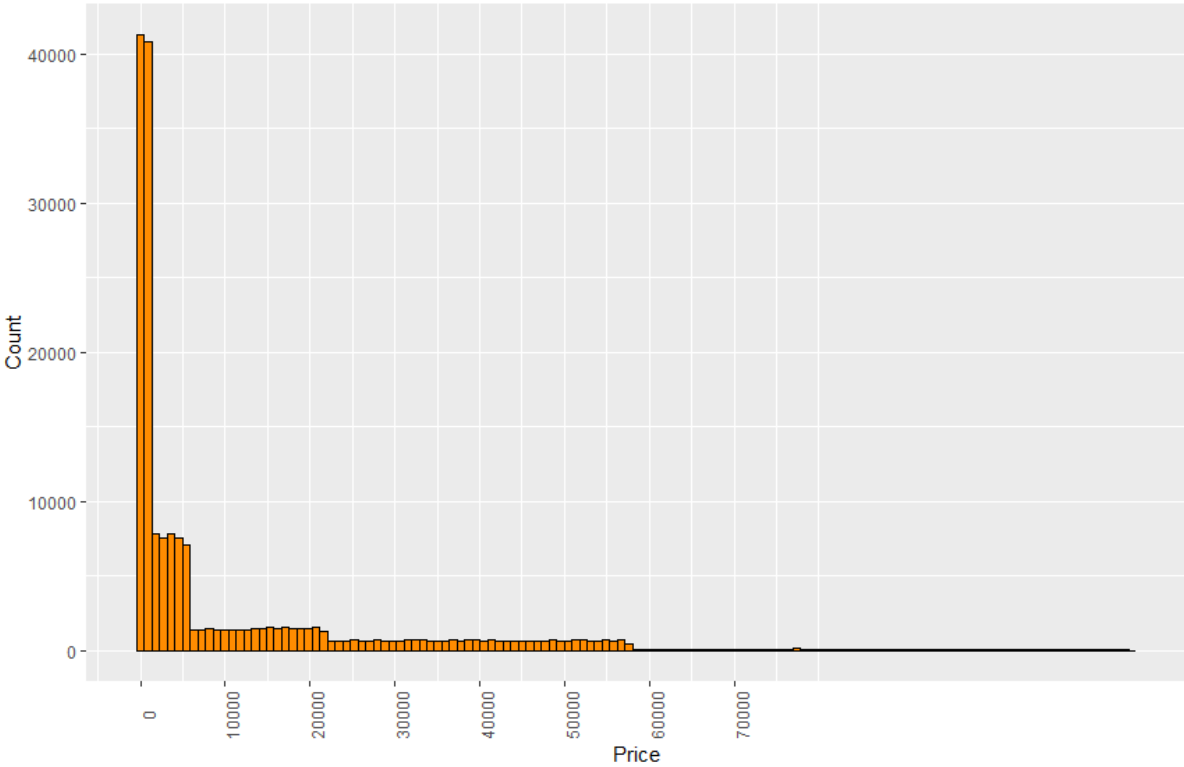


Figure 1:Price distribution

Figure 1 above describes how much people spend on the interval of R0 to approximately R60 000. The distribution is right skewed where approximately 42 000 customers spend between a mean of R R0 to R1 700. Approximately 7 000 customers spend between 1700 to 5900.

The mode of the data is R567.18 which means that most of the customers buy items that cost an average of R567.18. These items are the winners of the business. The business may prioritise having more of this product in stock based of the fact that they are the most bought. Below is the code used to find the mode.

Fig2: mode code

The summary is as follows: Min =R35.65, Quartile 1=R4821.31, Median=R2259.63, Mean=1294.10, Quartile 3=R 1 5270, Max=116 618.

Min	Q1	Median	Mean	3Q	Max
35.65	482.31	2259.63	12294.1	15270.97	116618.97

Item	Clothing	Food	Gifts	Household	Luxury	Sweets	Technology
Count	26403	24582	39149	20065	11868	21564	36347

The following table shows the number of counts per item That were produced. From this table It shows that Gifts is the feature that has the highest Count, followed by Technology feature. The lowest one is Households. From this table one can predict that most of the bought items are by middle aged individuals with constant income.

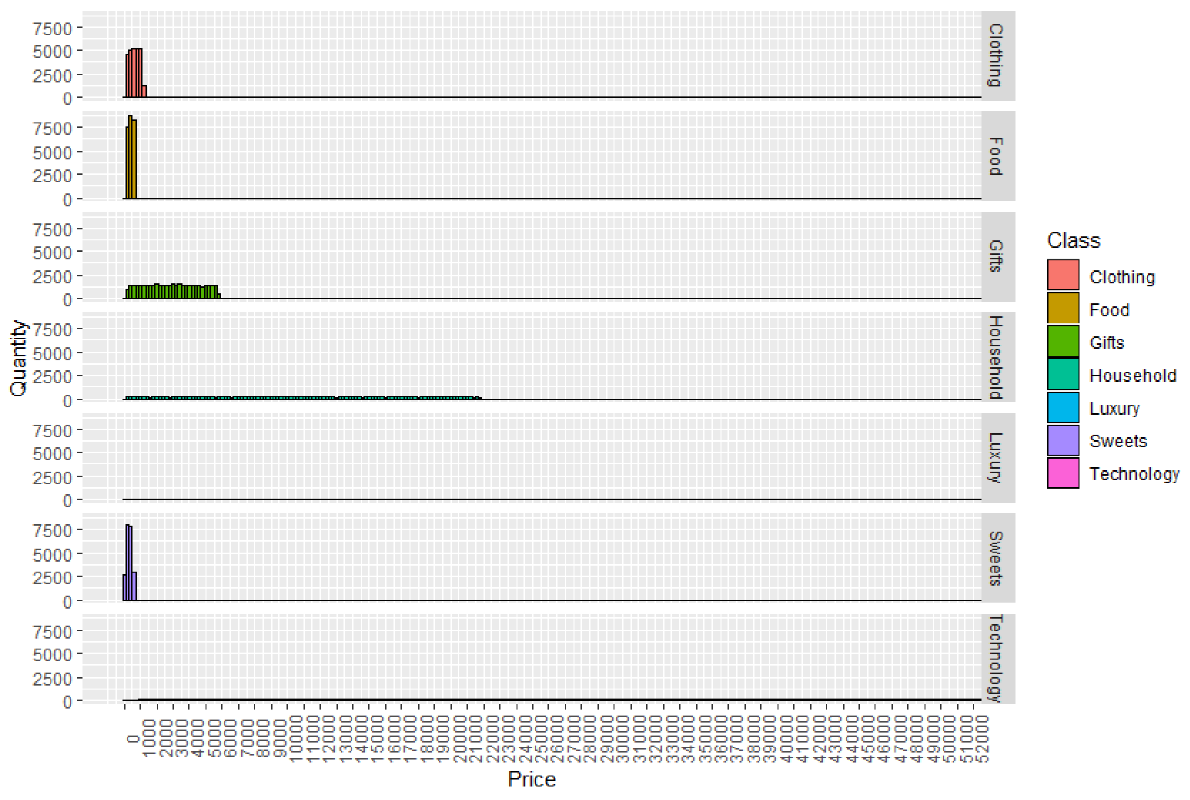


Figure 2:Combined price plot

Figure 2 shows a combined feature of the different categories of price vs quantity. Technology is the lowest purchase category, and clothing and food relatively look equal in terms of quantity. The graph is not deduced but helps in visualising aspects. Household components varies along the different price ranges. Food is at the lowest price range, followed by clothing's and sweets.

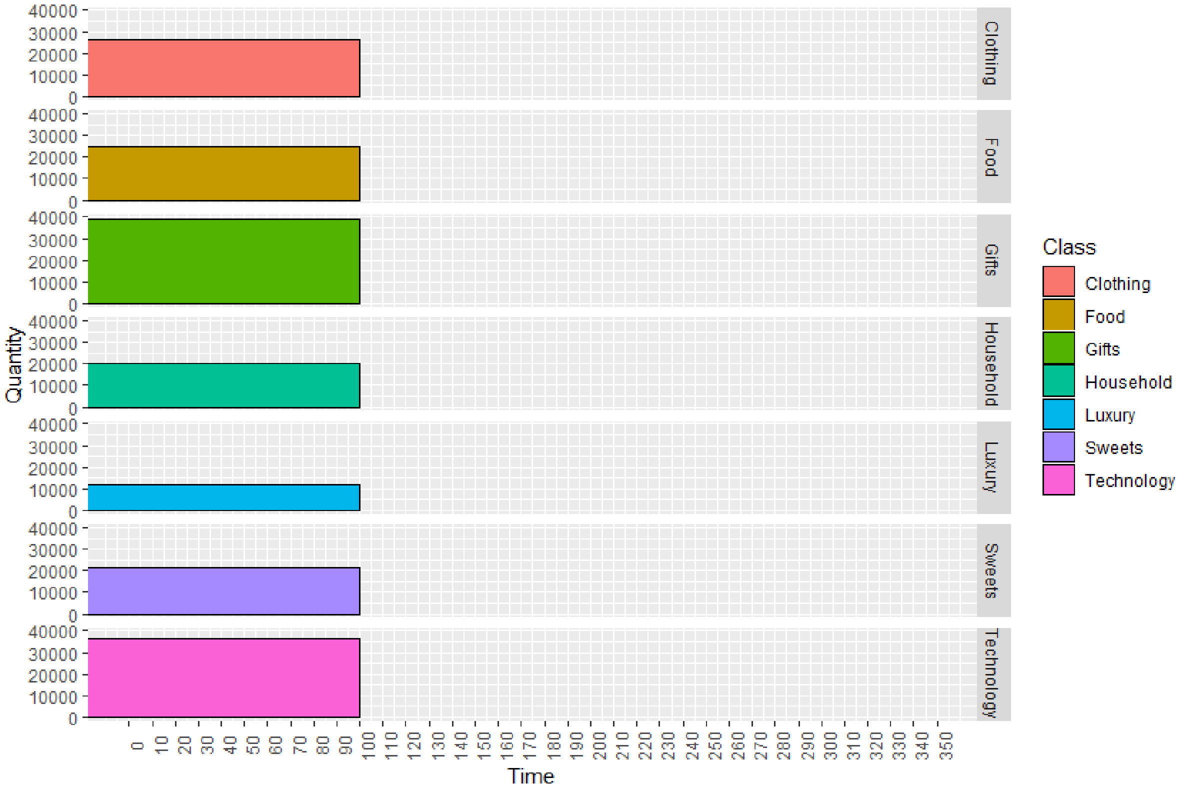


Figure 3:combined plot

Figure 3 shows that delivery times are not necessarily influenced by items purchased. It is clear in figure 3 that the delivery times were evenly distributed equally regardless of the feature at hand.

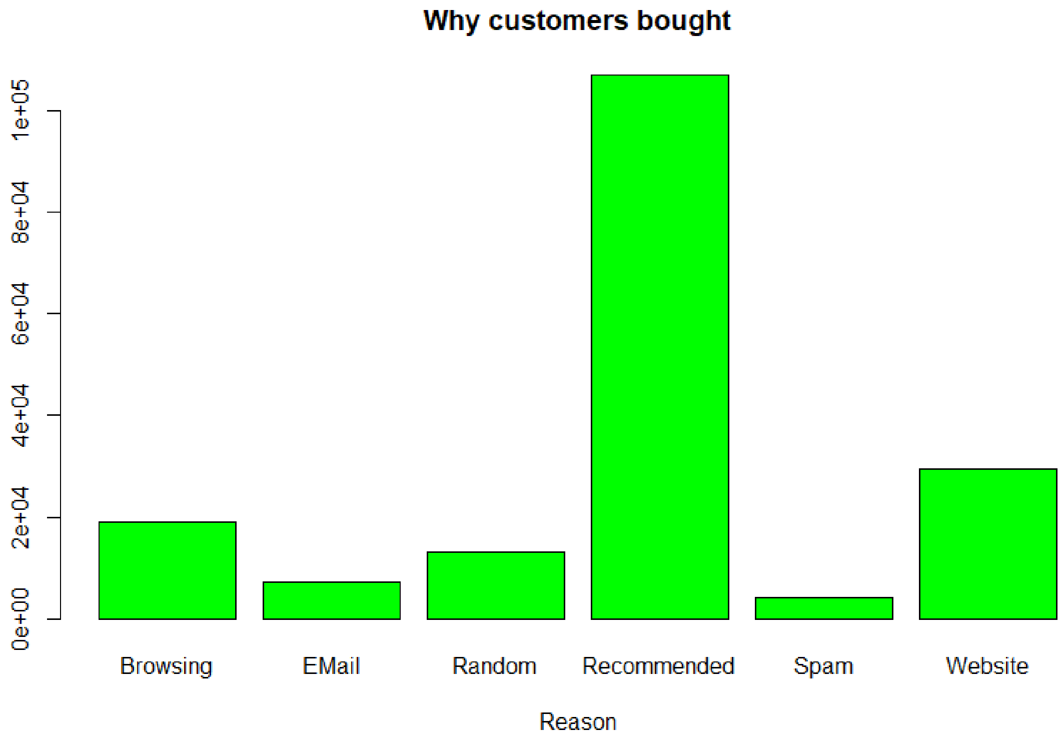


Figure 4:REASON

In figure 4 majority of the sales are due to recommendations, the second highest is websites, and the minimum is Spam. The data in terms of reason for purchase is not evenly distributed. Although the other features seem evenly distributed, excluding the 'Recommended'.

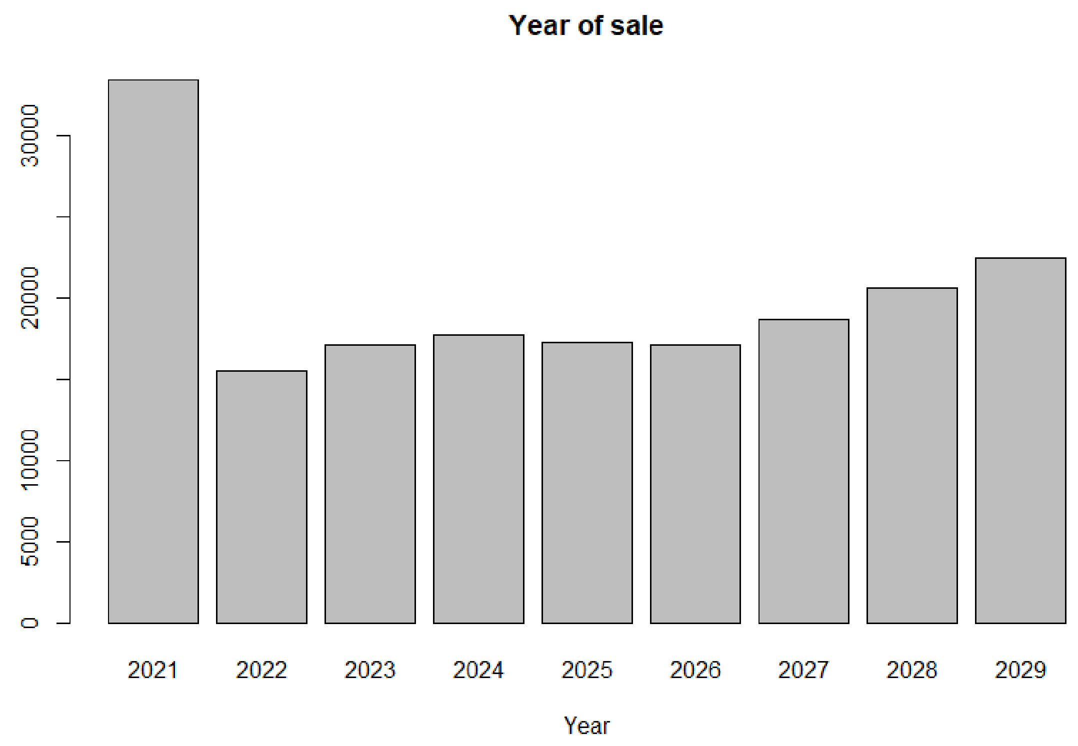


Figure 5:big year sale

The year that had the most sales is 2021, The company was not able to reach those sales after that year. The improvement they may have made in 2021 may not be sustainable.

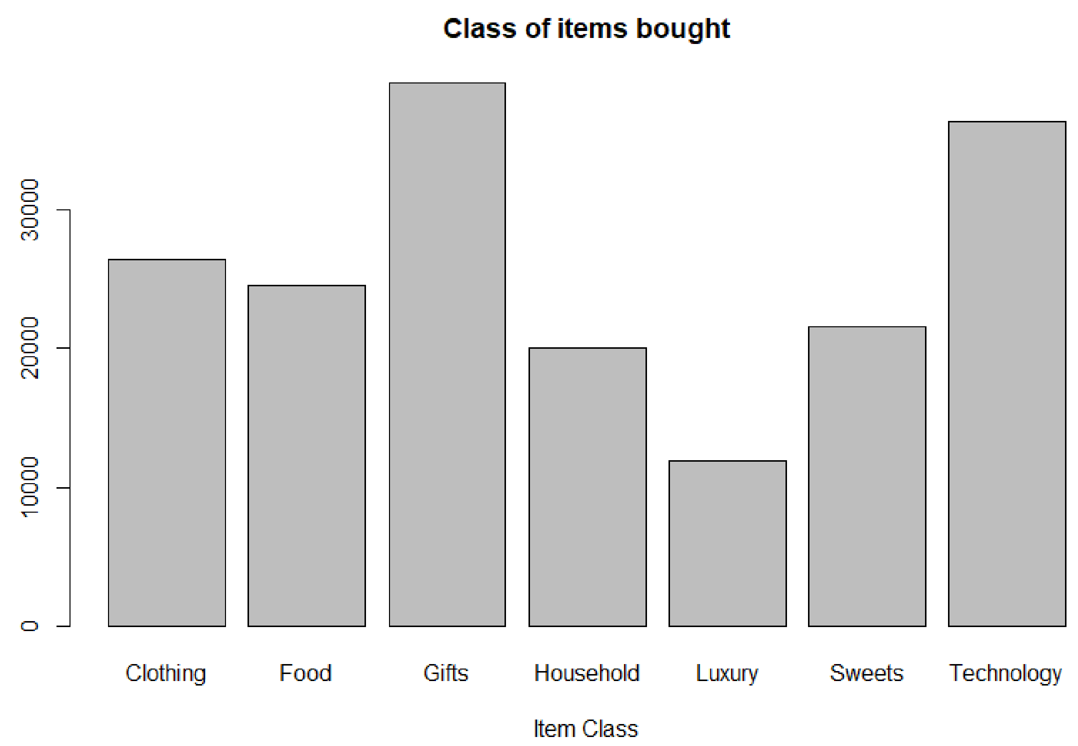


Figure 6:Items bought

The most class of items bought are gifts, followed by technology, and the lowest class is luxury.

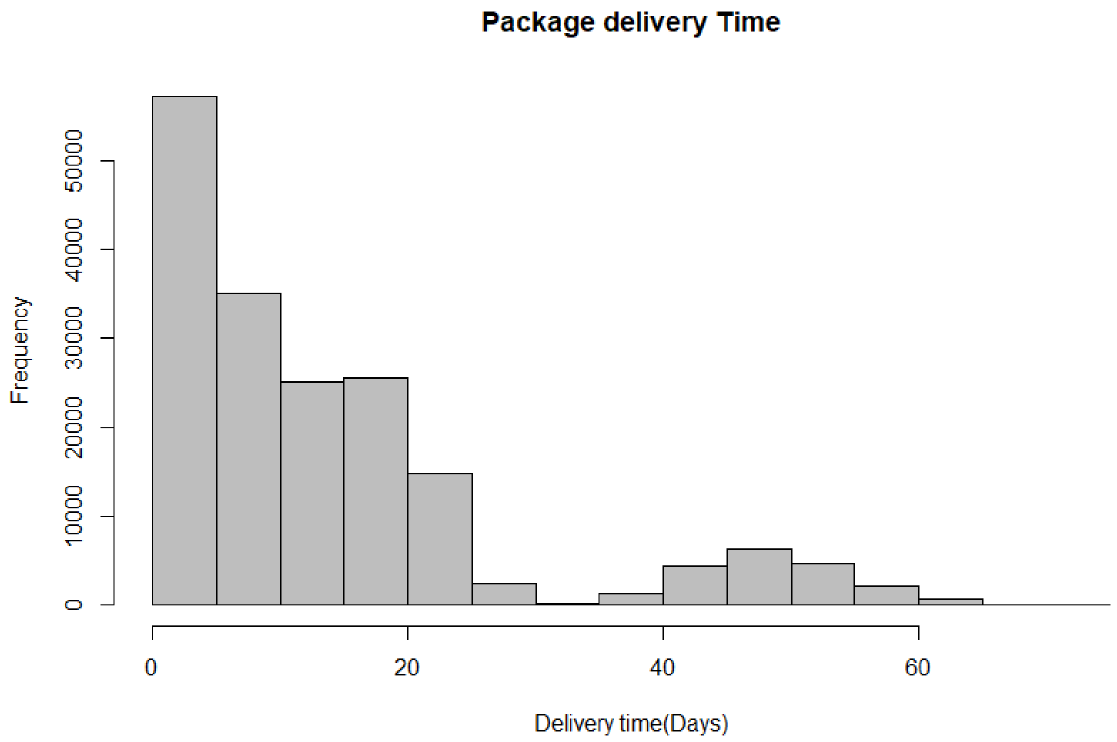


Figure 7:Delivery time

Within a period of 60 days most deliveries are done in the first 5 days, and the least deliveries are done around the 30th and 35th day. The data of delivery times vs frequency of the time intervals is not evenly distributed. This is a clear representation of purchases that happen due to a new product.

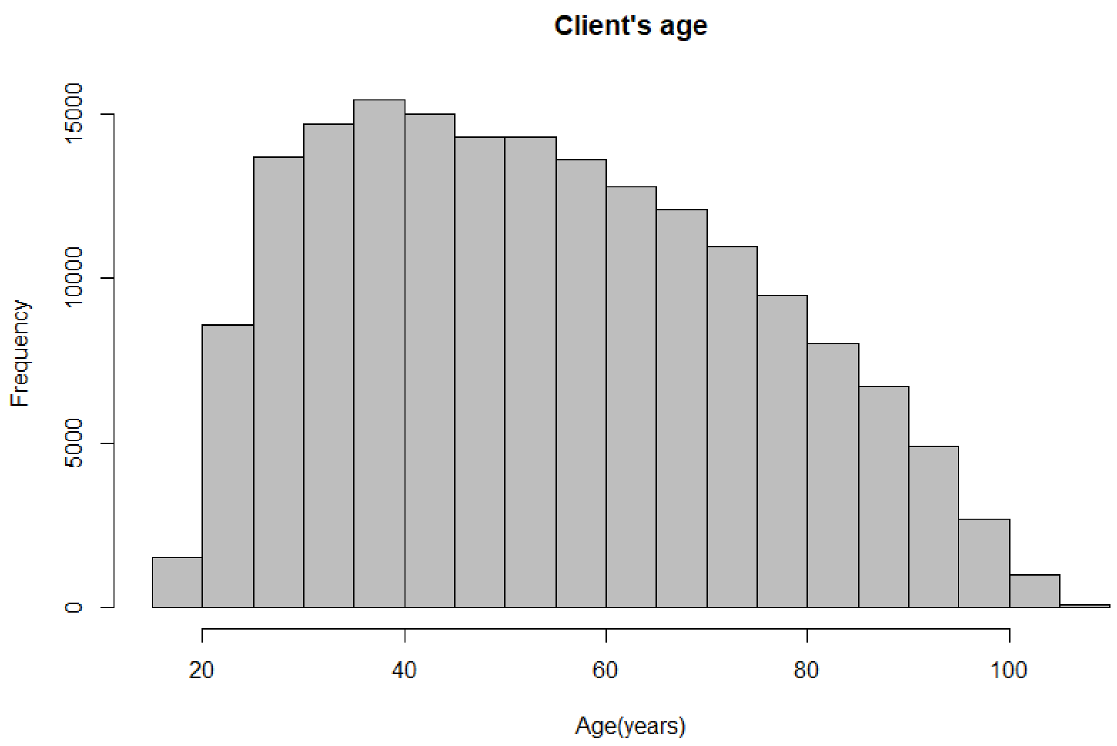


Figure 8;AGE group

The age group is evenly distributed, the company accommodates different age groups. Its is normal from the mid age group to be frequent buyers because its anormal societal trend. The data is skewed more to the left. This means that more activity in this category occurs around the range of 25 to 70.

Process Capabilities Indices

The process capability index gauges how much variance a process encounters in comparison to its specification parameters. We might compare various procedures in terms of the ideal circumstance or whether they live up to our expectations. The Cp, Cpu, Cpl, and Cpk are calculated to analyse and determine relationships(gauge). An upper specification limit (USL) of 24 days and a lower specification limit (LSL) of 0 days. An LSL of 0 is logical in the sense that an item cannot be delivered before being ordered, although it can be delivered the same day that will be in hours (between day 0 and day 1).

Calculations for the process capability indices.

CP = (USL- LSL)/6σ

CPU = (USL - μ)/3σ

CPL = (μ – LSL)/3σ

CPK = min (CPL, CPU)

R-code results:

Cp = 1.142207

Cpl = 1.90472

Cpu = 0.3796933

Cpk = 0.3796933

The analysis of process capability indices

To see if the process can meet the specifications, the variation must be than the specific range according to Evans and Lindsay (2017). The Cp of the process delivery time is calculated to be 1.142207. The Cp is lower than the specification range, hence the process can meet the specifications of delivery times.

To evaluate the process cantering, we compare the Cpu and Cpl values. From the Calculations above we can conclude that the Cpu value is smaller than the Cpl value. According to Evans and Lindsay (2017), this means that the process mean is closer to the upper specific limit (UCL).

3.Control charts

Part3.1

A graph used to examine how a process evolves over time is the control chart. Data are plotted according to time. An average line in the middle, an upper line for the upper control limit, and a lower line for the lower control limit are always present on a control chart. These curves were created using historical data.

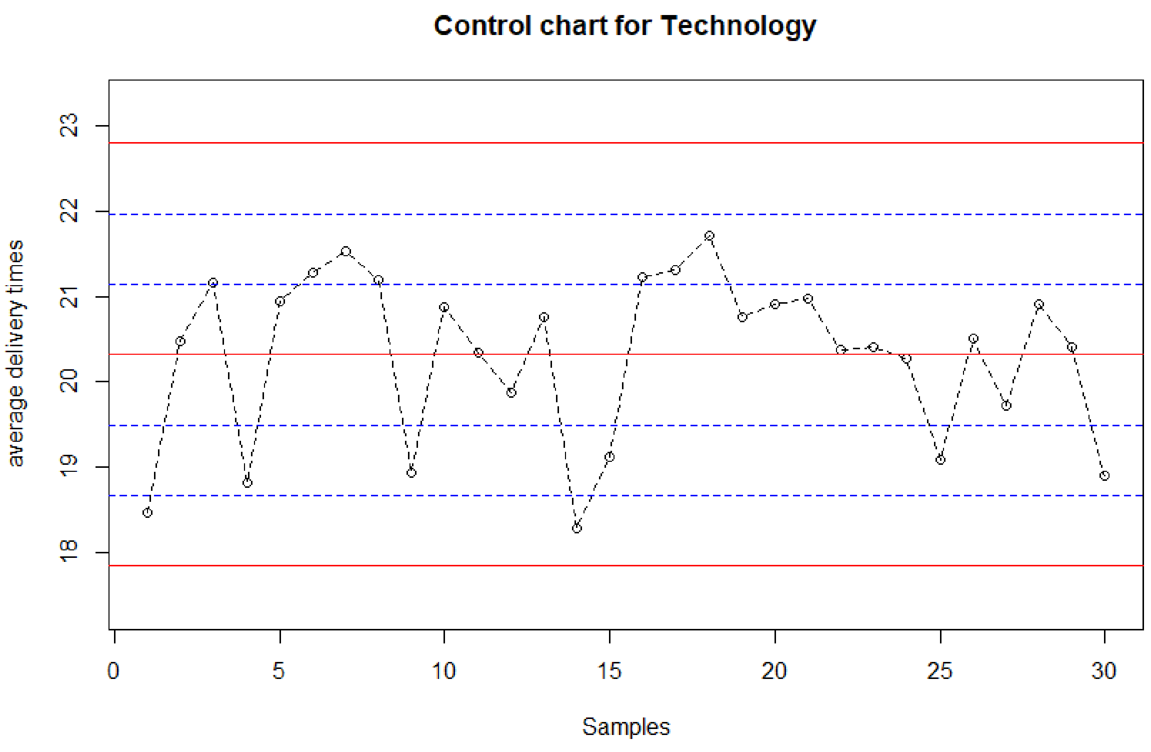


Figure 9:X-bar control -Technology

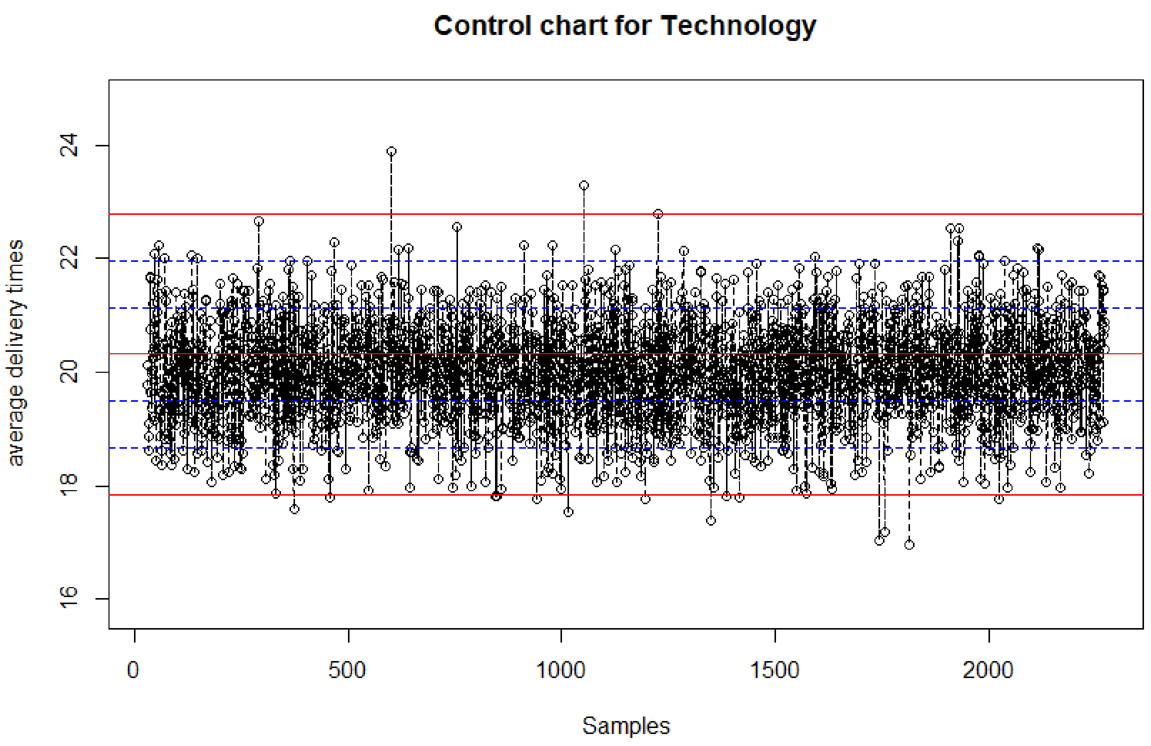


Figure 10:S-chart-Technology

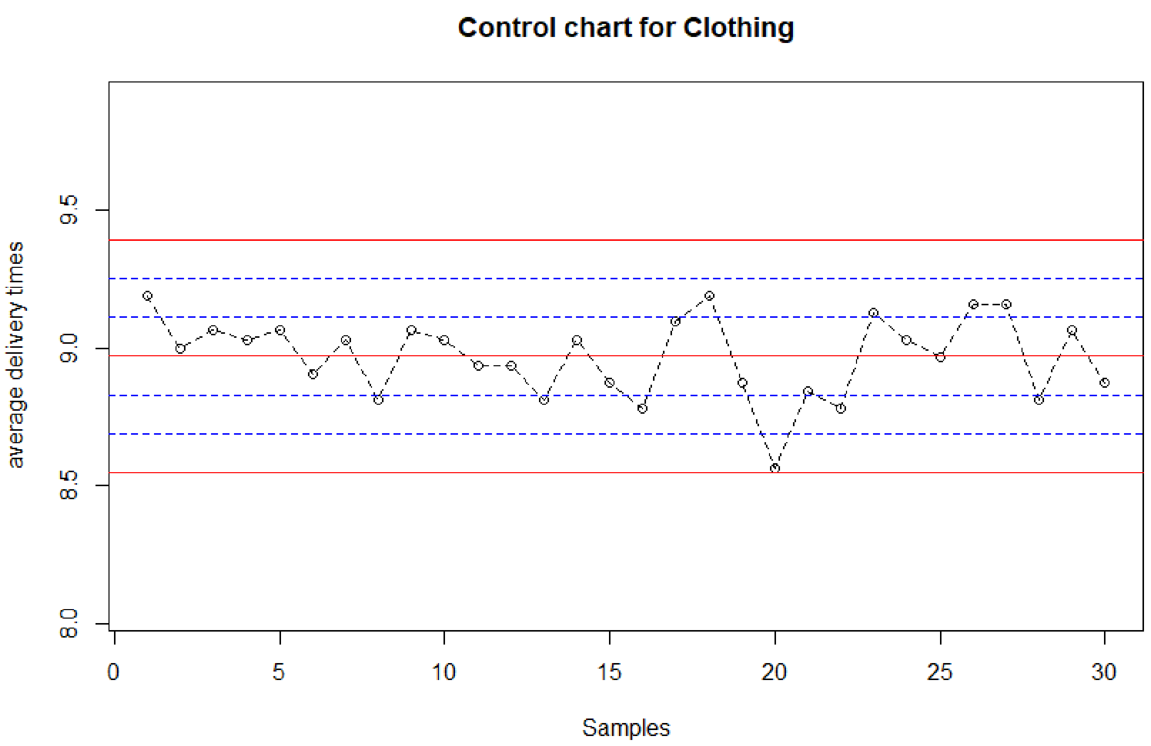


Figure 11:X- chart-Clothing

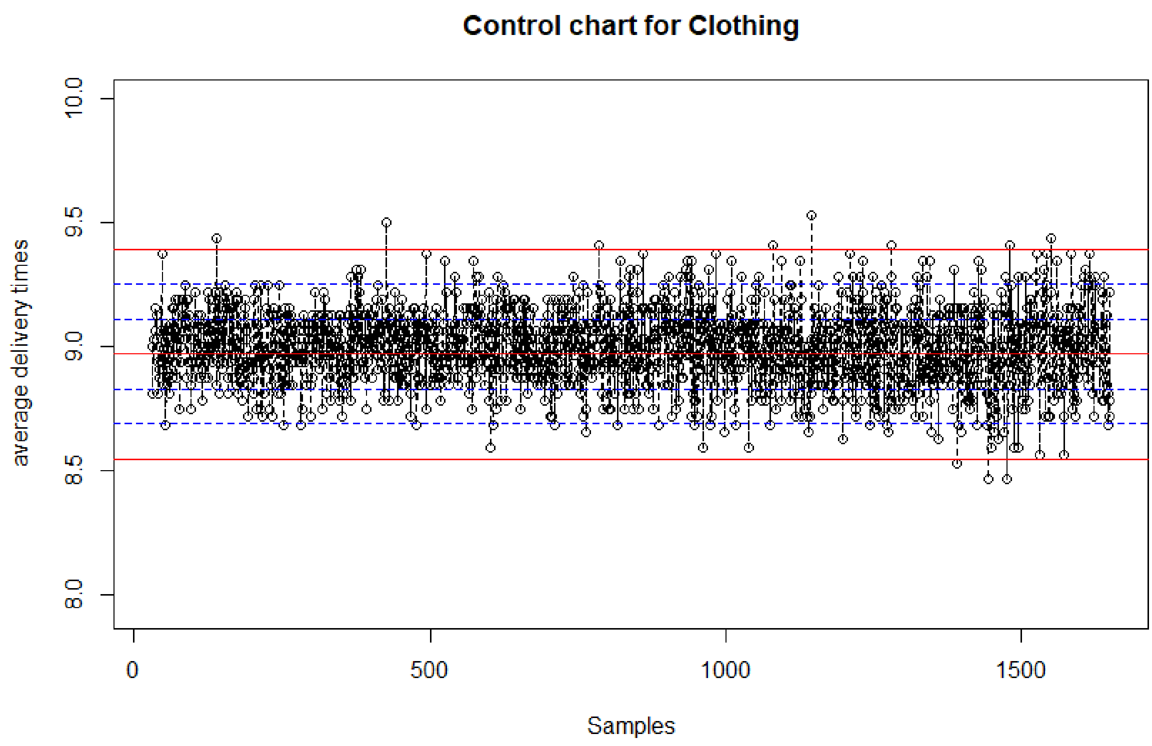


Figure 12:S-chart-Clothing

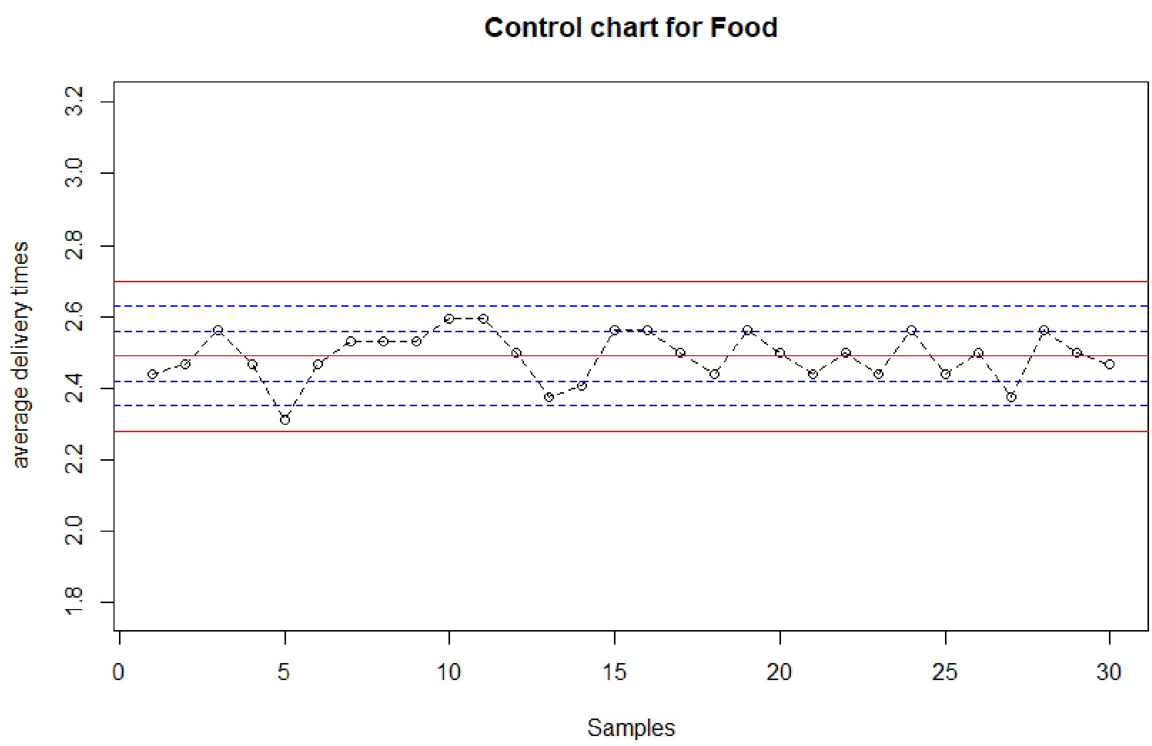


Figure 13:X-chart-Food

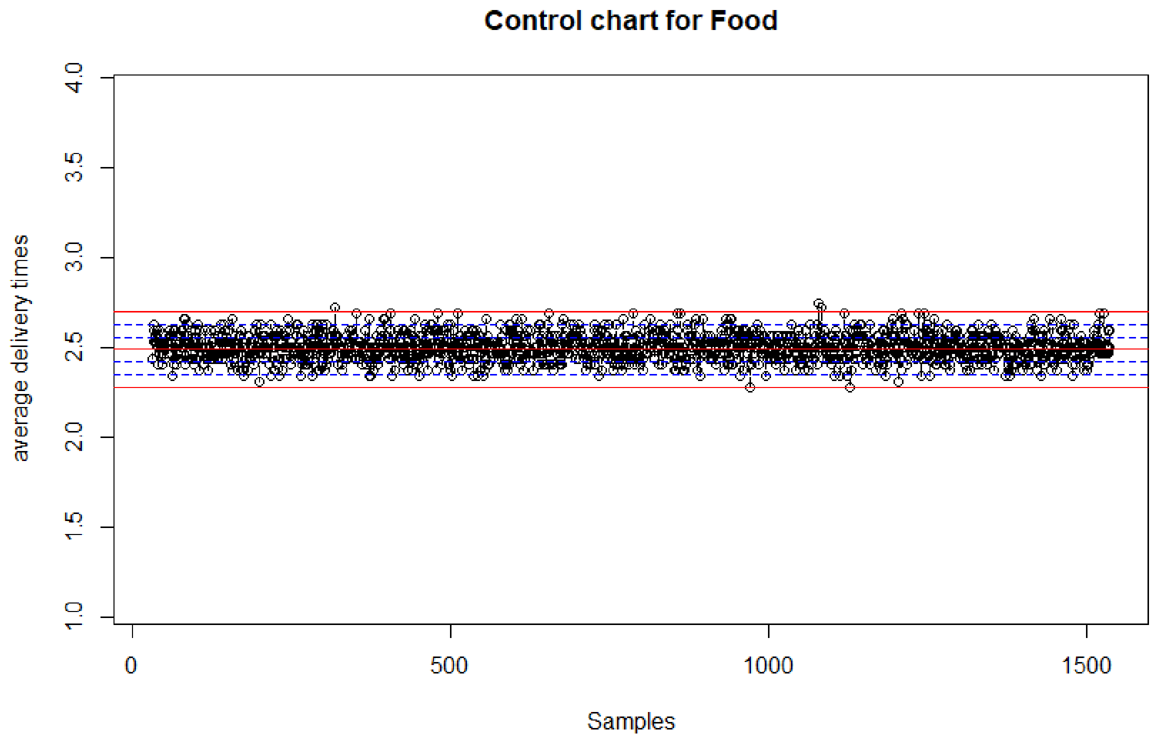


Figure 14:S-chart-Food

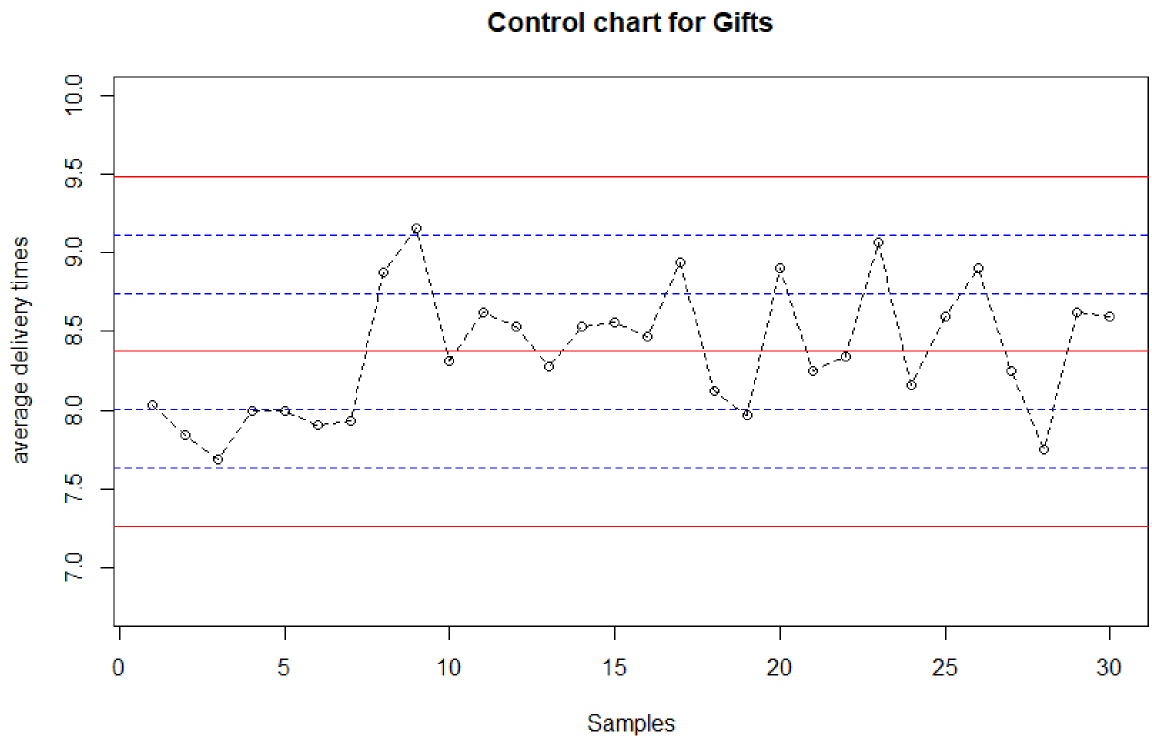


Figure 15:X-chart-Gifts

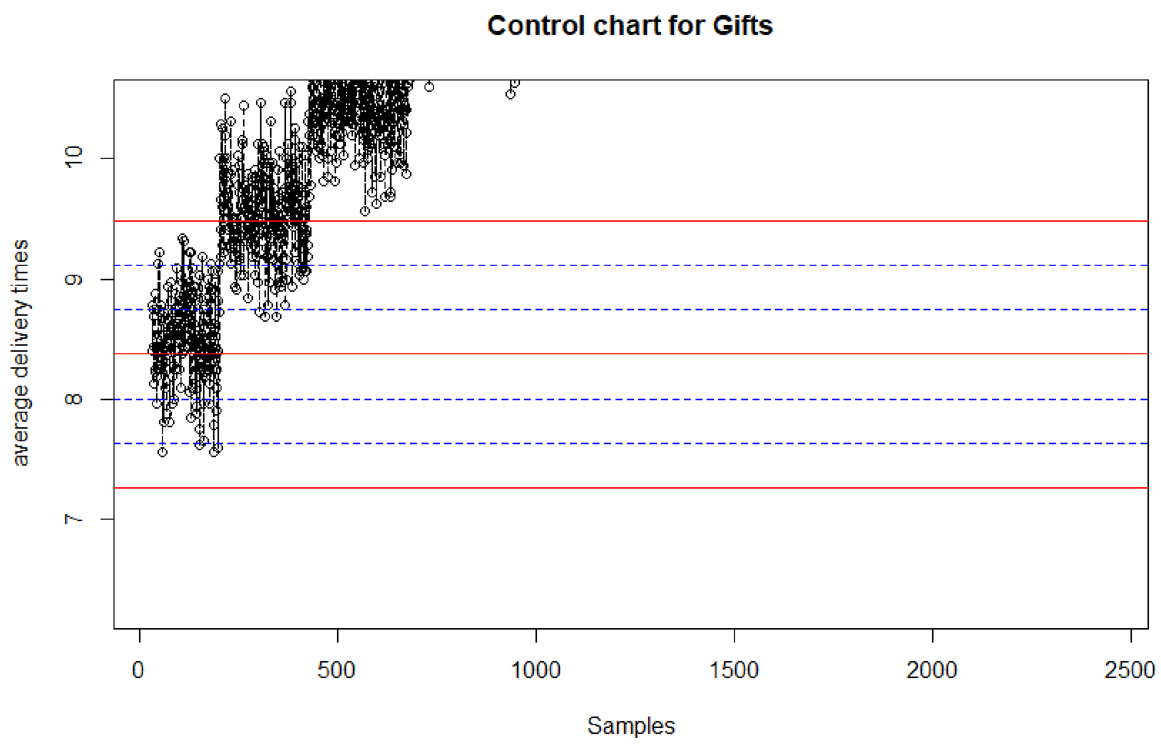


Figure 16:S-chart-Gifts

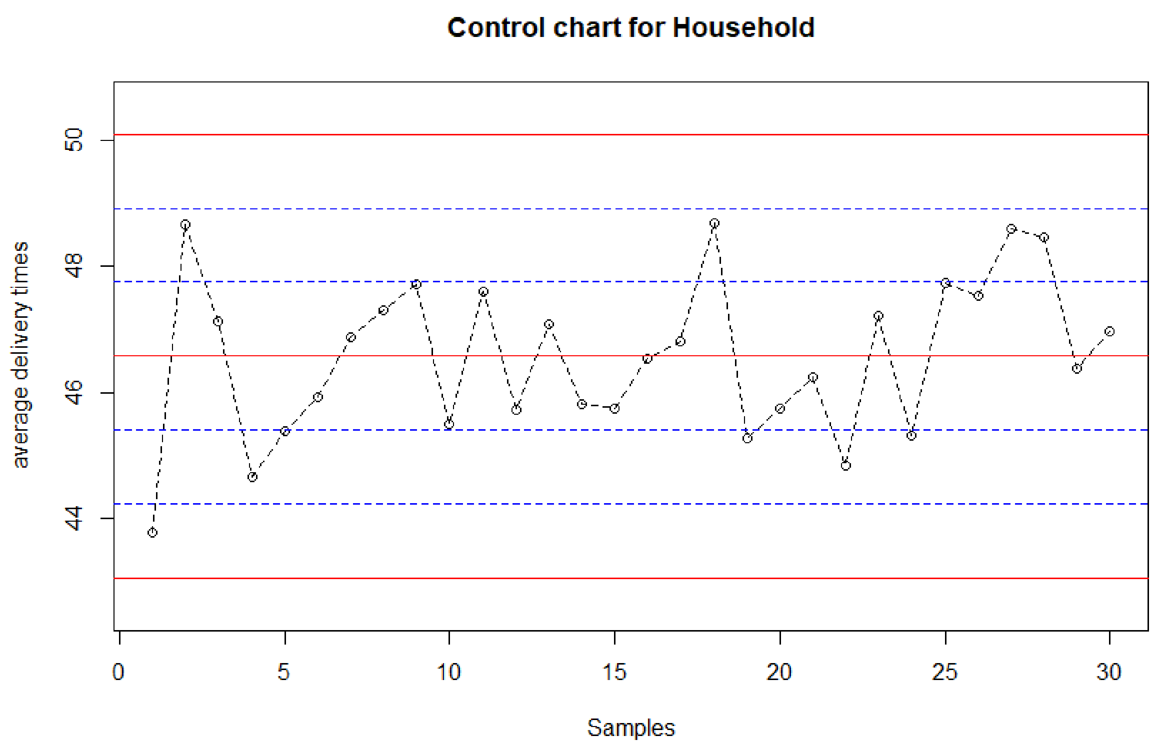


Figure 17:X-chart-Household

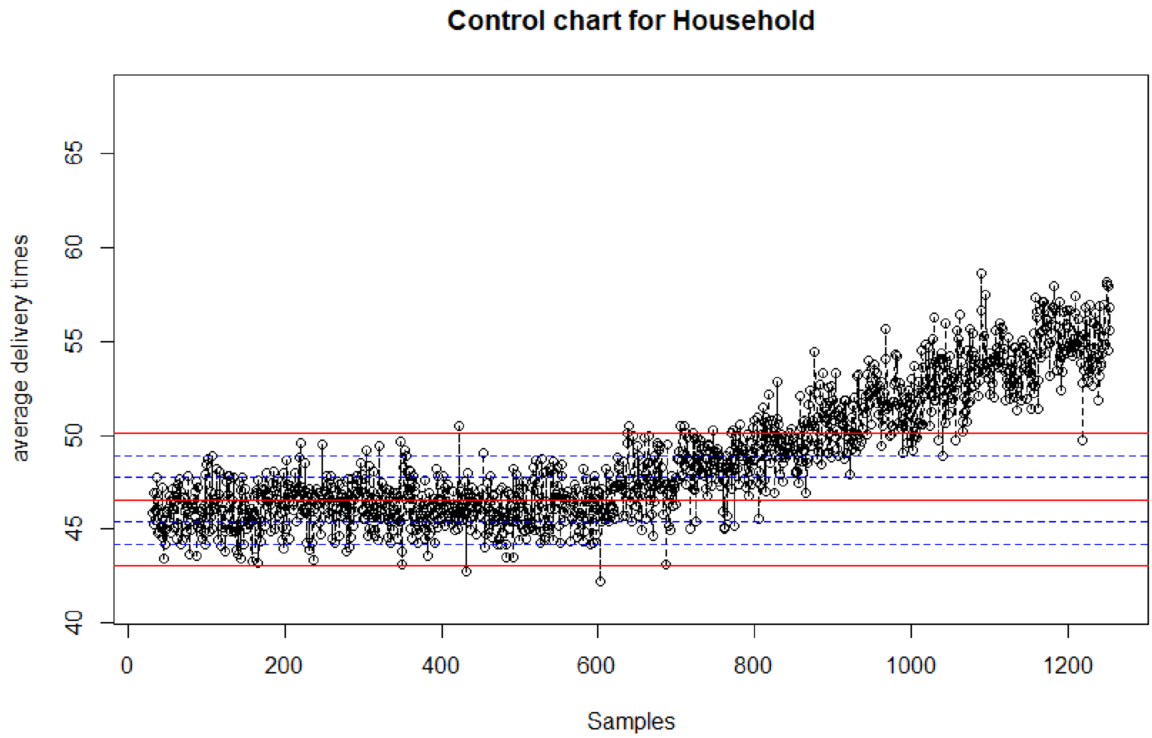


Figure 18:S-chart-Household

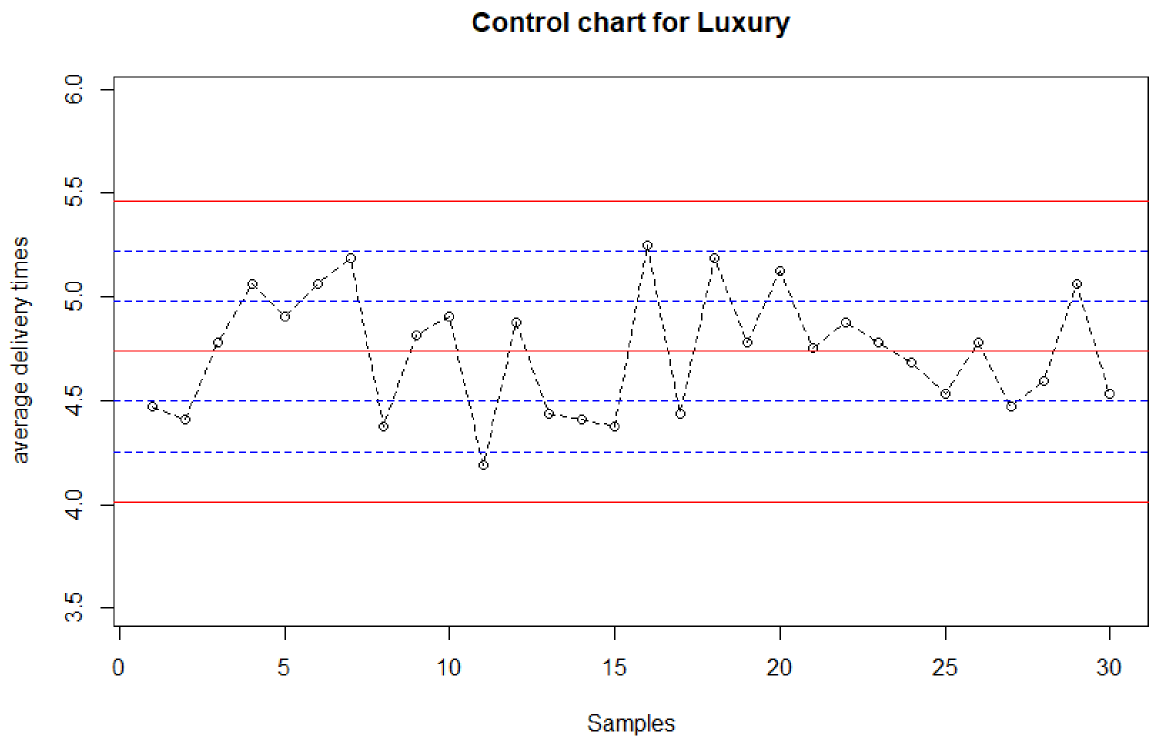


Figure 19:X-chart-Luxury

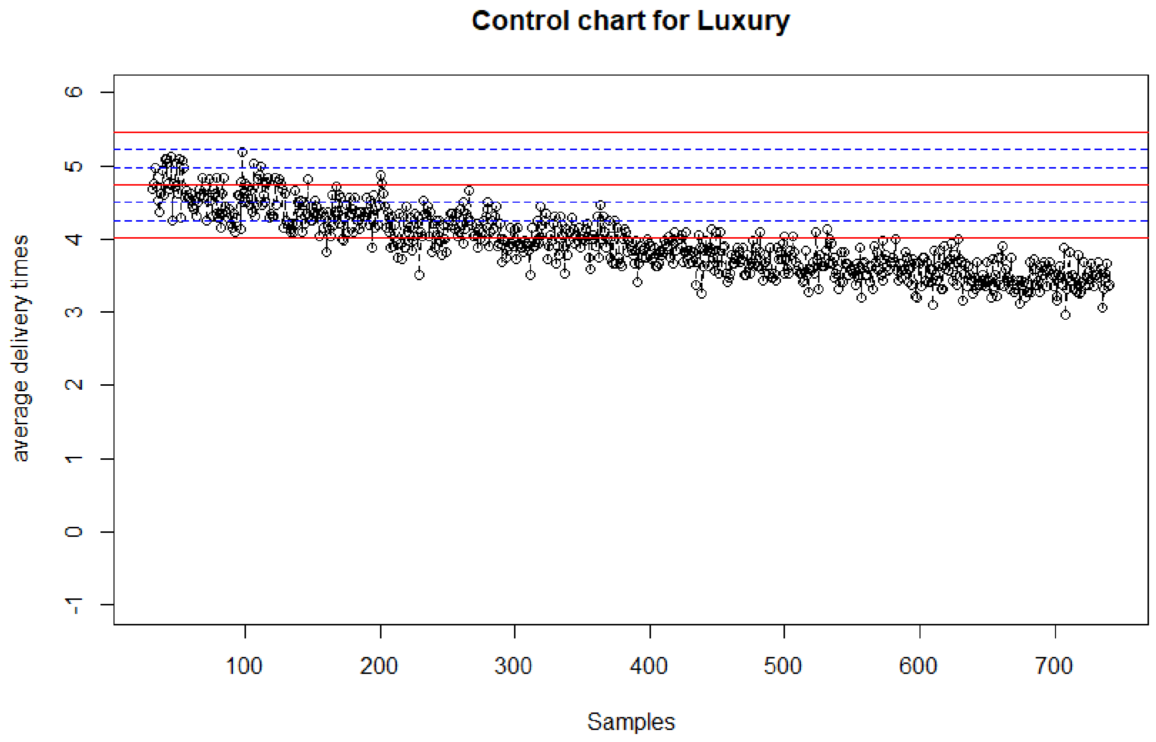


Figure 20:S-chart-Luxury

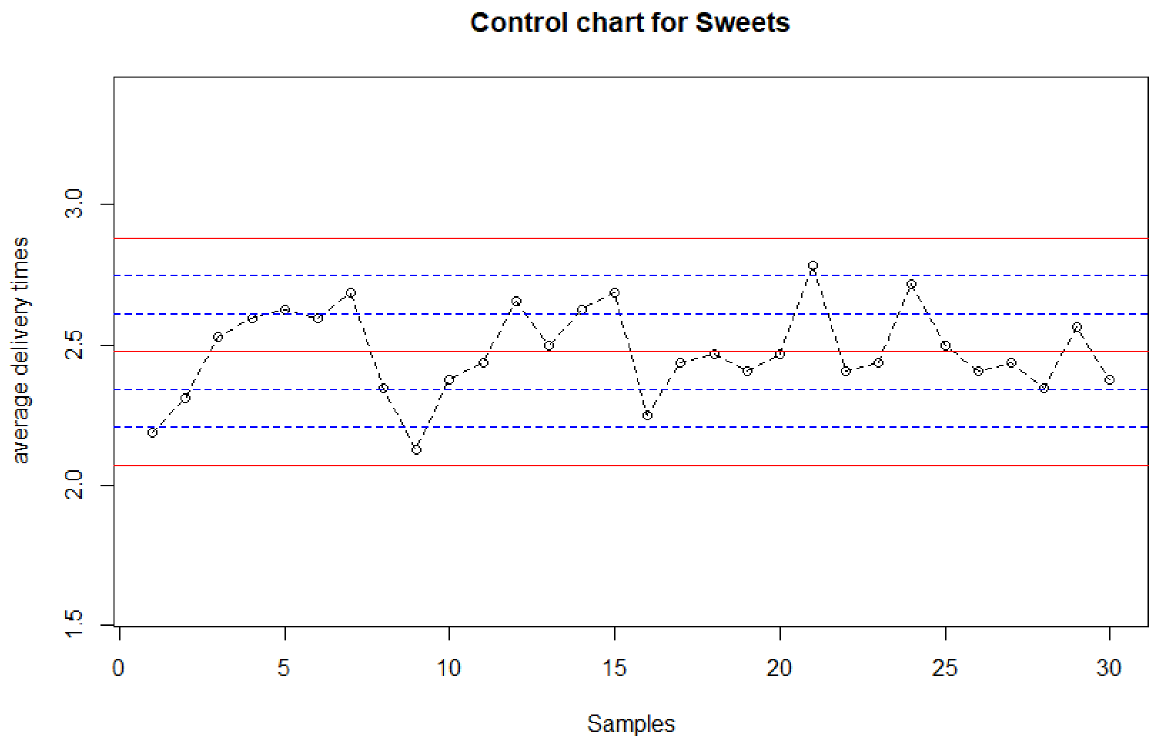


Figure 21:X-chart-Sweets

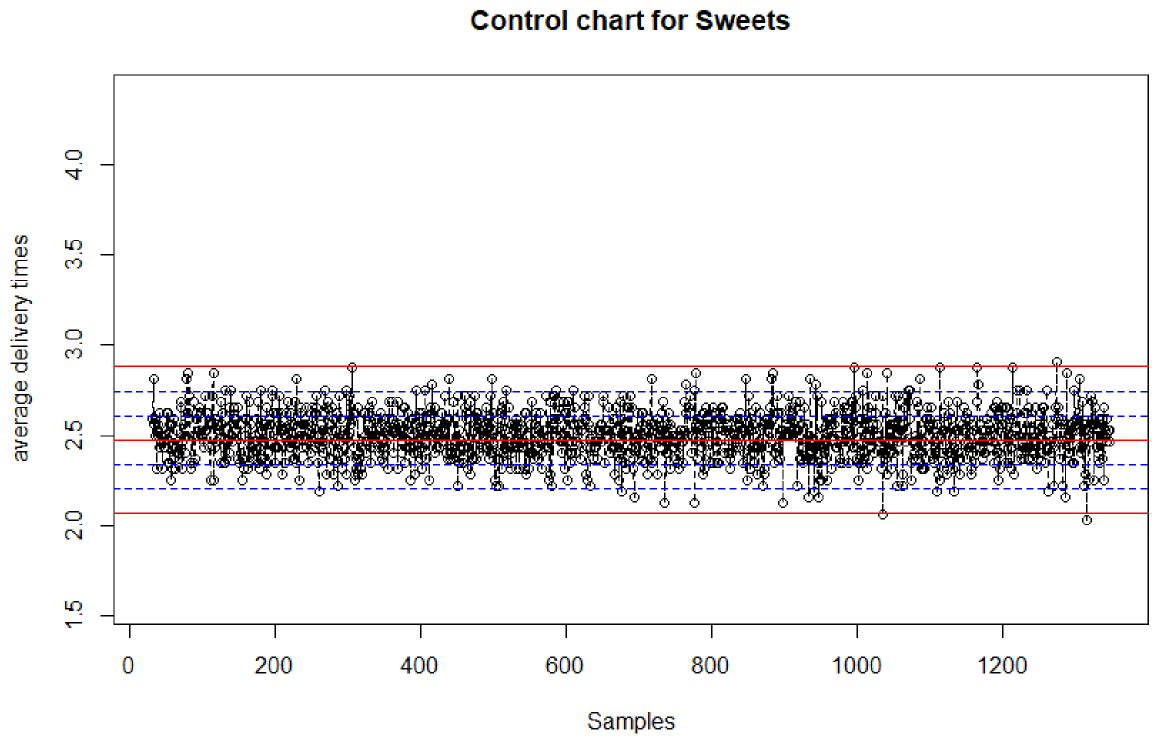


Figure 22:S-chart-Sweets

X-chart table

Class	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
Technology	22.790762	21.966064	21.141365	20.316667	19.491968	18.667269	17.843
Clothing	9.391	9.251	9.11	8.969	8.829	8.688	8.548
Food	2.700	2.630	2.560	2.490	2.420	2.350	2.280
Gifts	9.483	9.113	8.744	8.374	8.000	7.635	7.264
Household	50.101	48.926	47.751	46.576	45.400	44.226	43.051
Luxury	5.459	5.210	4.977	4.736	4.495	4.254	4.013
Sweets	2.881	2.746	2.611	2.446	2.341	2.206	2.071

S-chart table

Class	SUCL	SU2Sigma	SU1Sigma	SCL	SL1Sigma	SL2Sigma	SLCL
Technolgy	5.032	10.127	15.221	20.317	25.411	30.506	1.452
Clothing	0.857	3.561	6.266	8.970	11.674	14.378	0.247
Food	0.427	1.115	1.802	2.489	3.177	3.864	0.123
Gifts	2.256	4.295	6.335	8.374	10.413	12.452	0.651
Household	7.170	20.305	33.441	46.576	59.711	72.846	2.069
Luxury	1.471	2.559	3.647	4.7364	5.825	6.913	0.424
Sweets	0.823	1.3744	1.925	2.476	3.027	3.578	0.237

Part 3.2

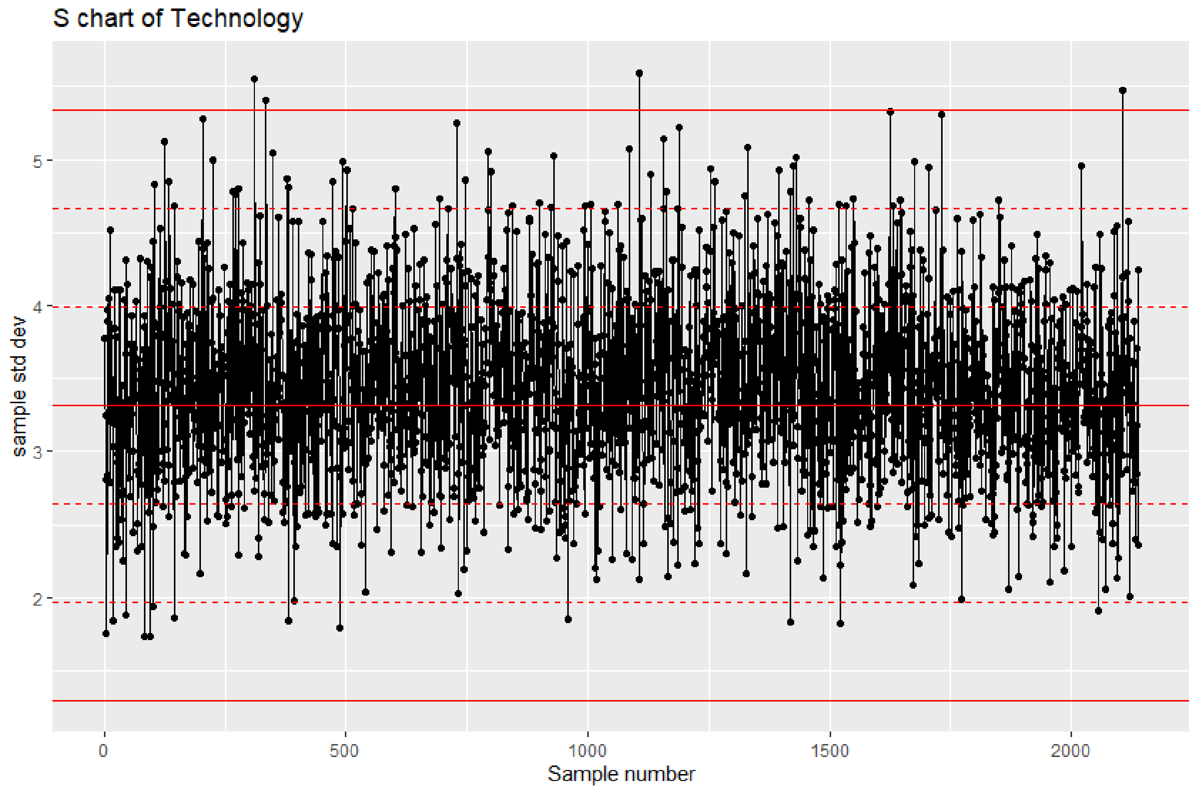


Figure 23:complete S-chart-Technology

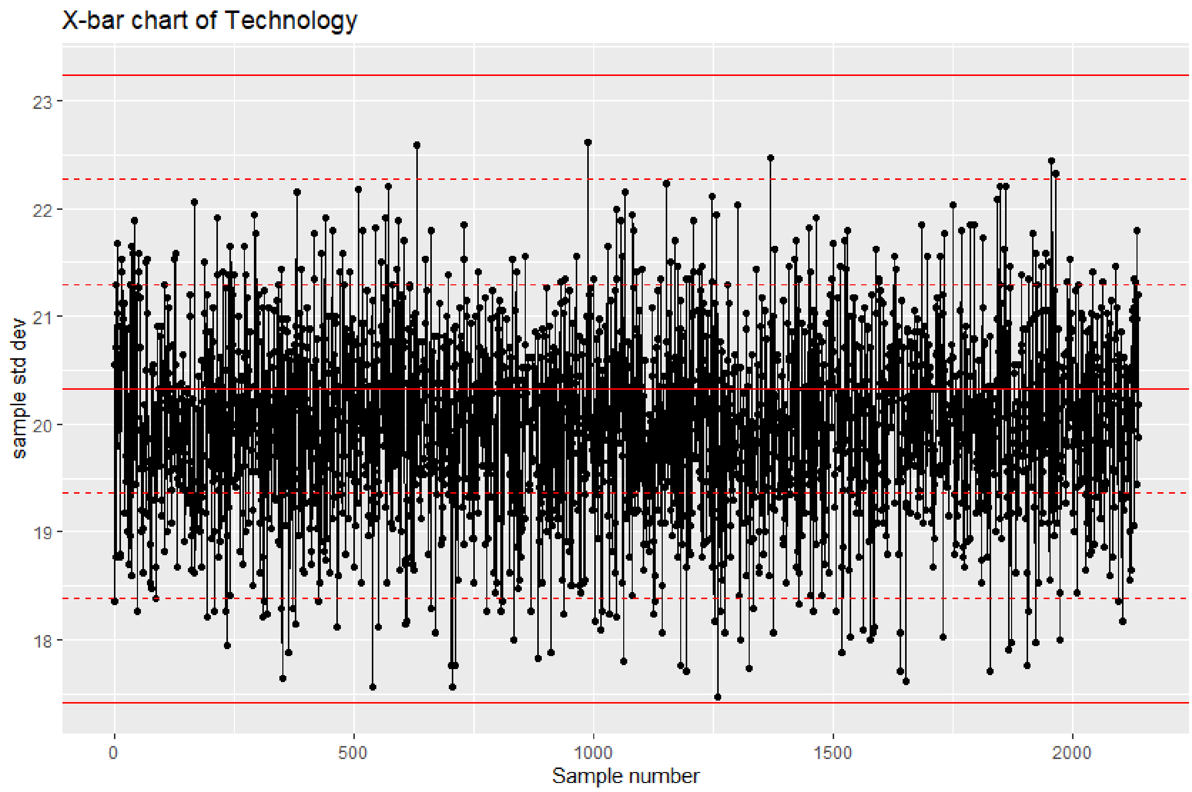


Figure 24:full X-chart-Technology

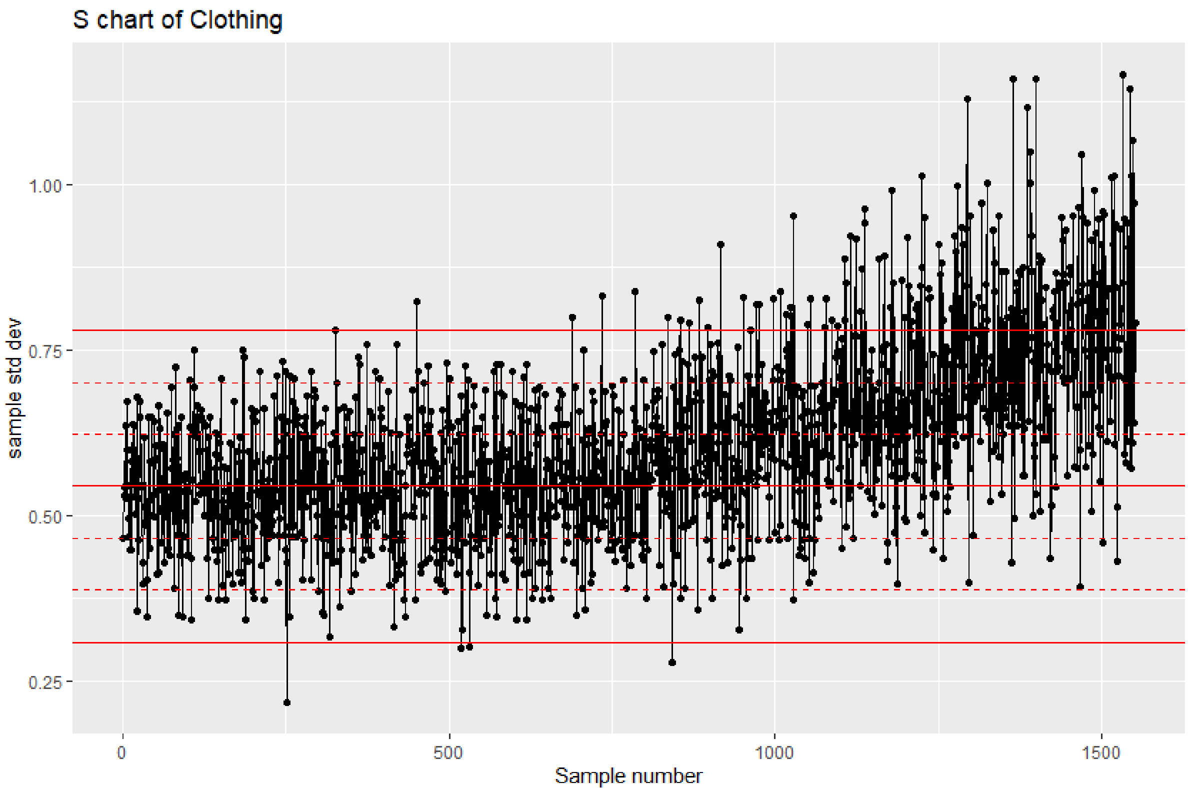


Figure 25:Complete S-chart-Clothing

Various data points in the clothing s-chart are outside the limits (More than 50 sample points). Due to the fact that this are clothes , this could be due to seasonal change.

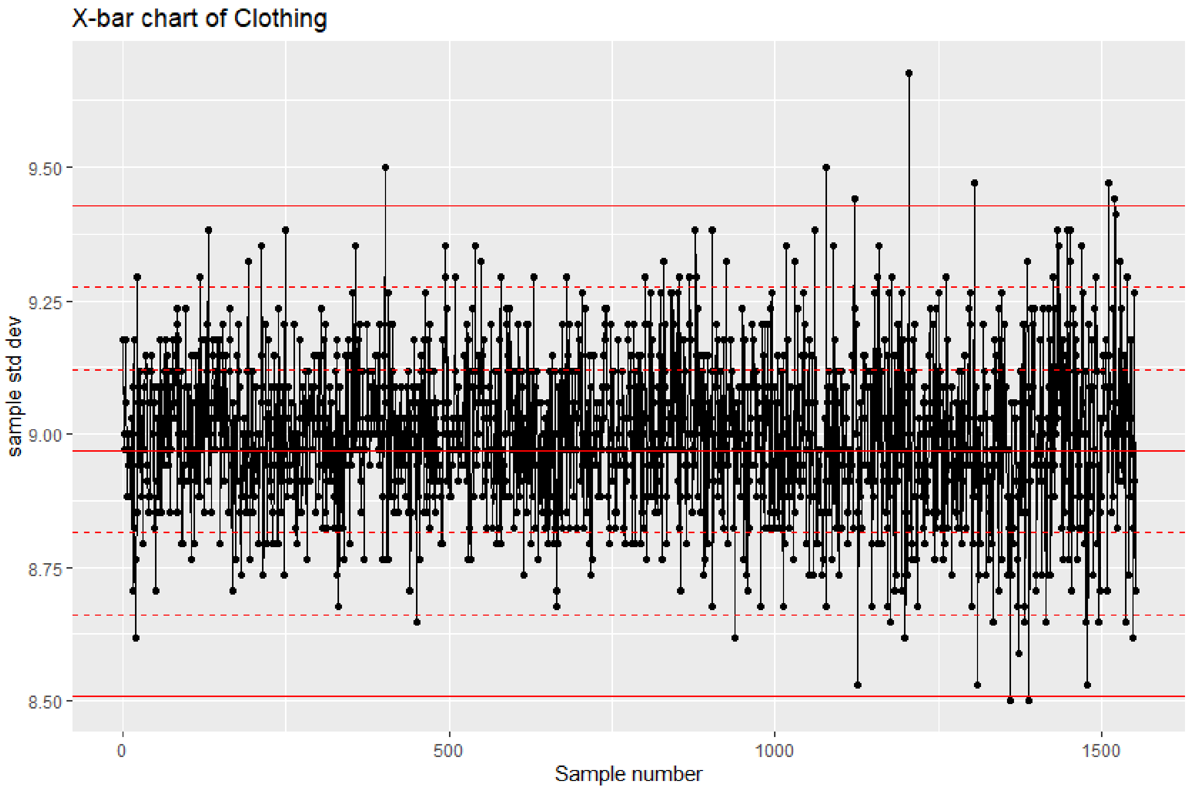


Figure 26:Complete x-chart-Clothing.

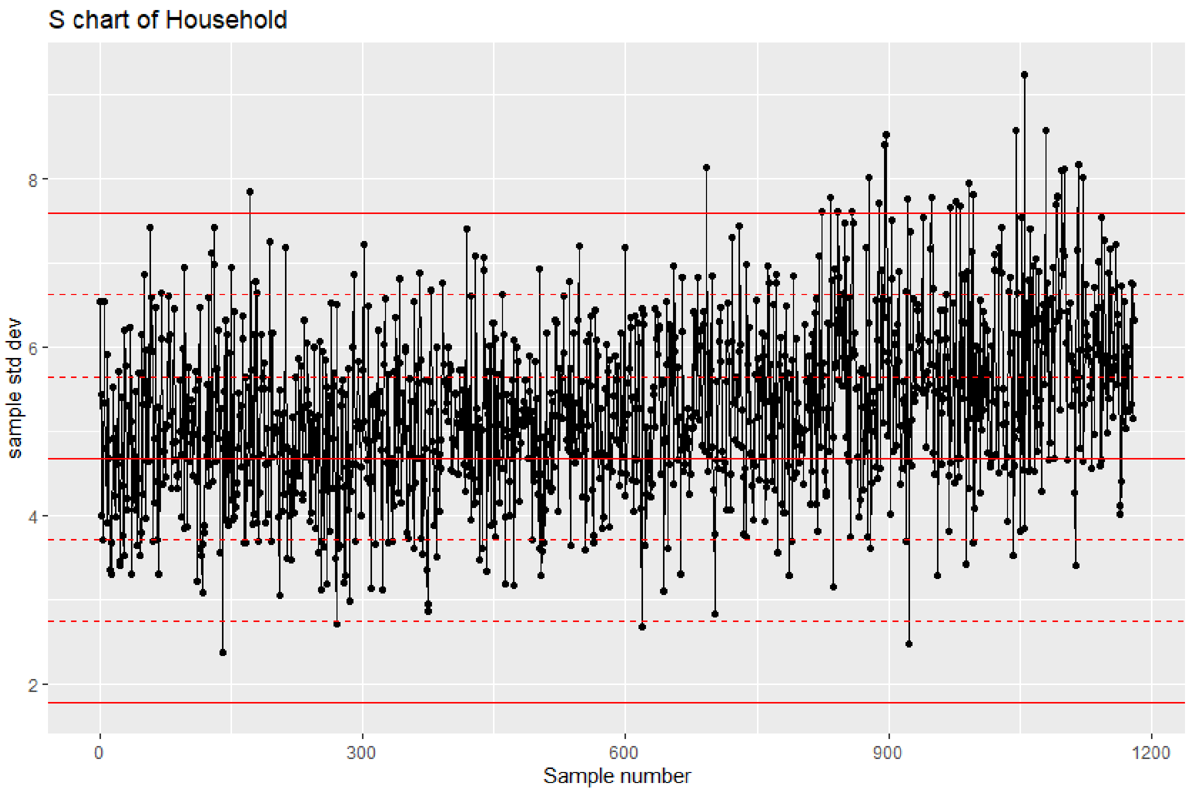


Figure 27:Complete S-chart-Household

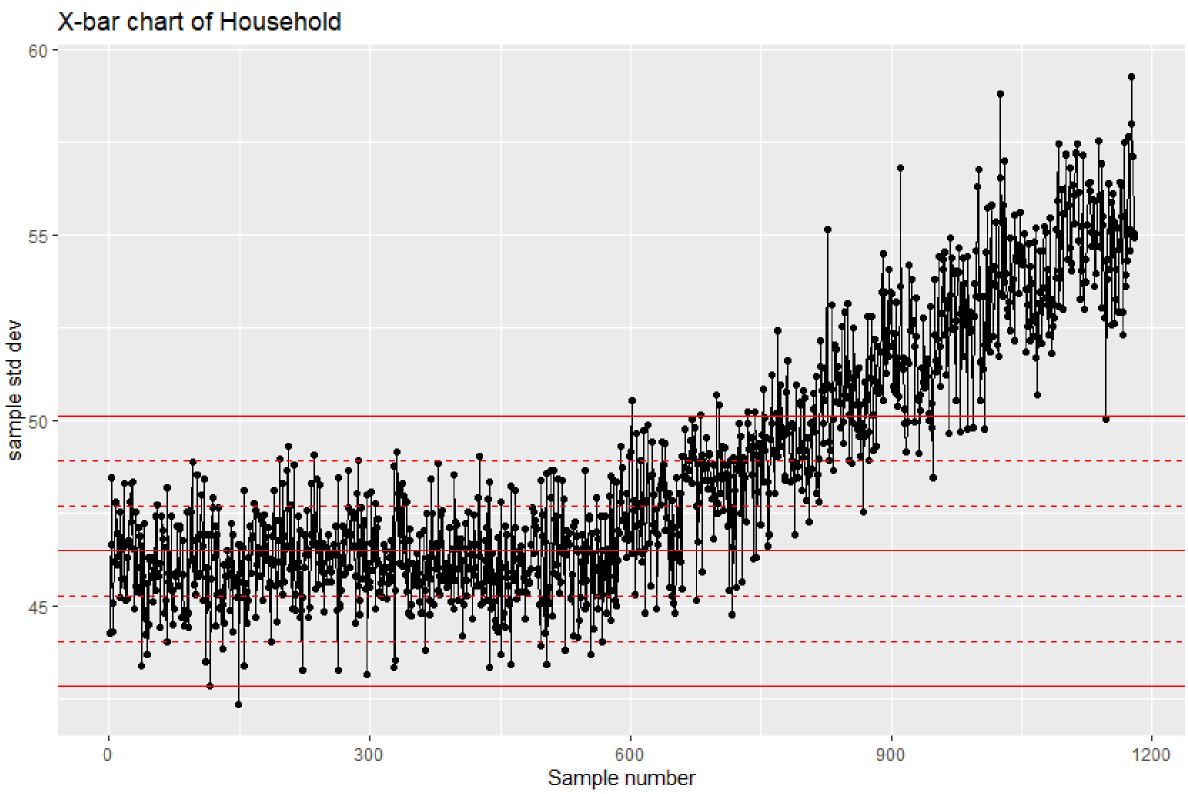


Figure 28:Complete x-chart-Household

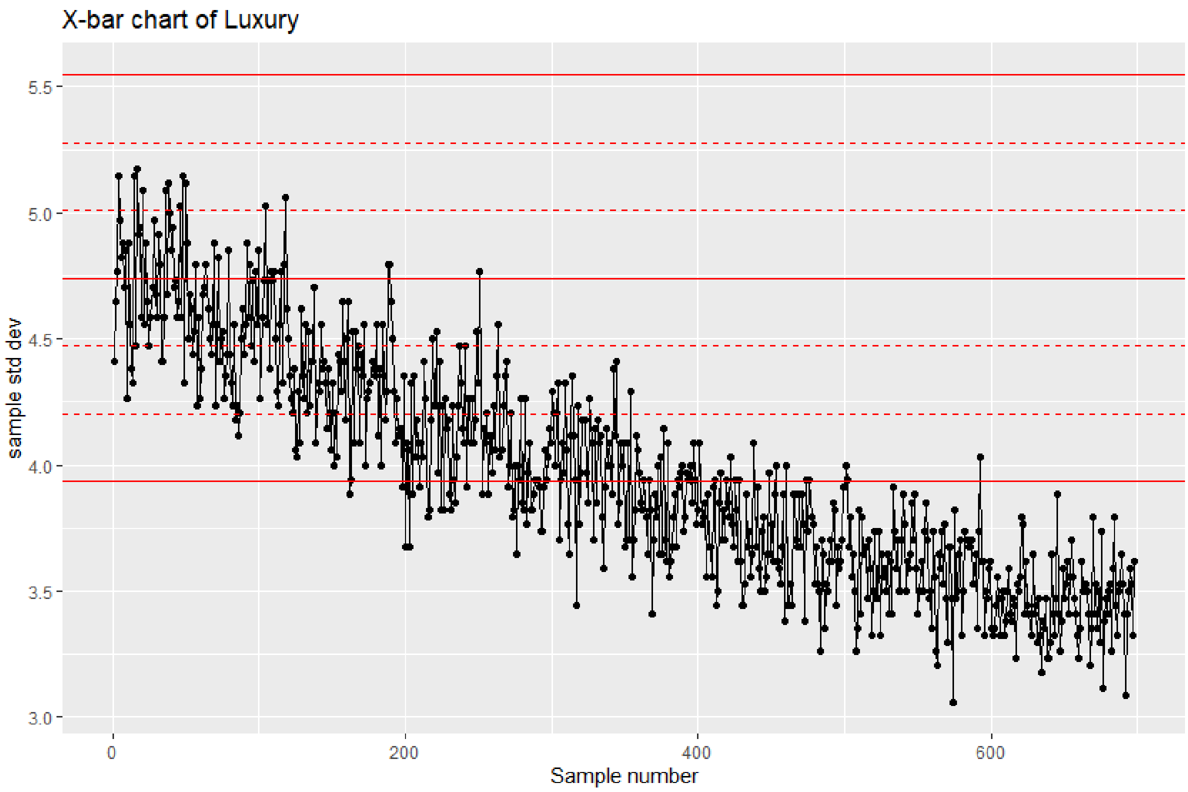


Figure 29:Complete X-chart-Luxury

Optimising the Delivery process

In this section of the report, the data that is provided is thoroughly examined using statistical process control. Apart from the data between the first and twenty-fifth data sets, all data were utilised.

A code on r-studios is used as tool for analysis.

For clothes

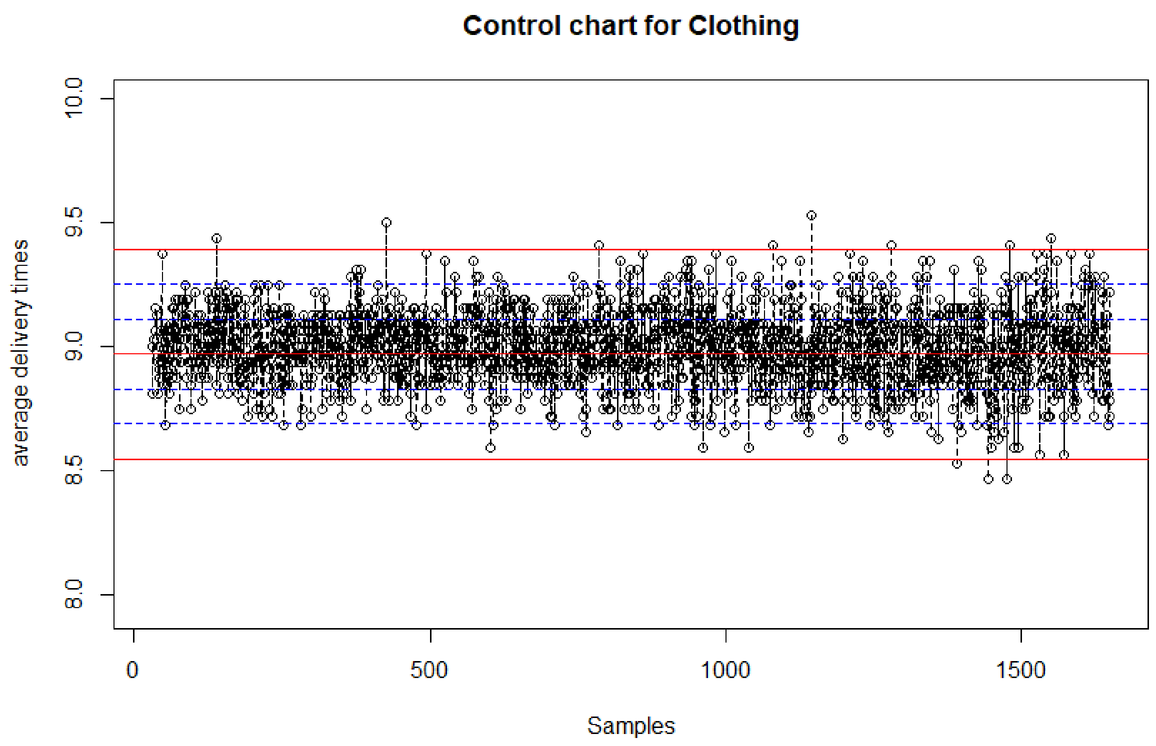


Figure 30:rest of Clothing data

Optimising the delivery process

For technology data

The are 16 points outside the limits. The 3 smallest are (373, 457,603) and the largest are (1755,1812,2023). The u1 sigma is 21.14

Considering Household data

There are 372 points that are outside. The U1Sigma is 47.75. The 3 smallest values are (422, 432,603) and the 3 largest value are (1252, 1253, 1254)

11 data points are presented outside of the limits. 3 smallest are:139, 427, 785. The 3 biggest are1476, 1480, 1550.

Type 1 error likelihood

The type 1 error for A and B was estimated as part of the various analysis computations for the section above. Based on the computations, it was shown that all clusters have an equal chance of producing a Type I error. The Type 1 probability for A is 0.0027 while for B is 0.683.

When we examine the technology class's statistics, we find that there are frequently delivery times that are more than 48 days. It follows from the problem statement that 30 deliveries that take longer than 48 days result in a decrease in sales. The average delivery time must be decreased to prevent such an outcome. Even though the reduction will cost money, it will be far less than the potential loss of sales if the delivery timeframes are not changed to be shorter than 48 days.

The days that the delivery procedure can be focused on are decided in this section.

The following graph, created with R programming, shows how many days the delivery procedure takes.

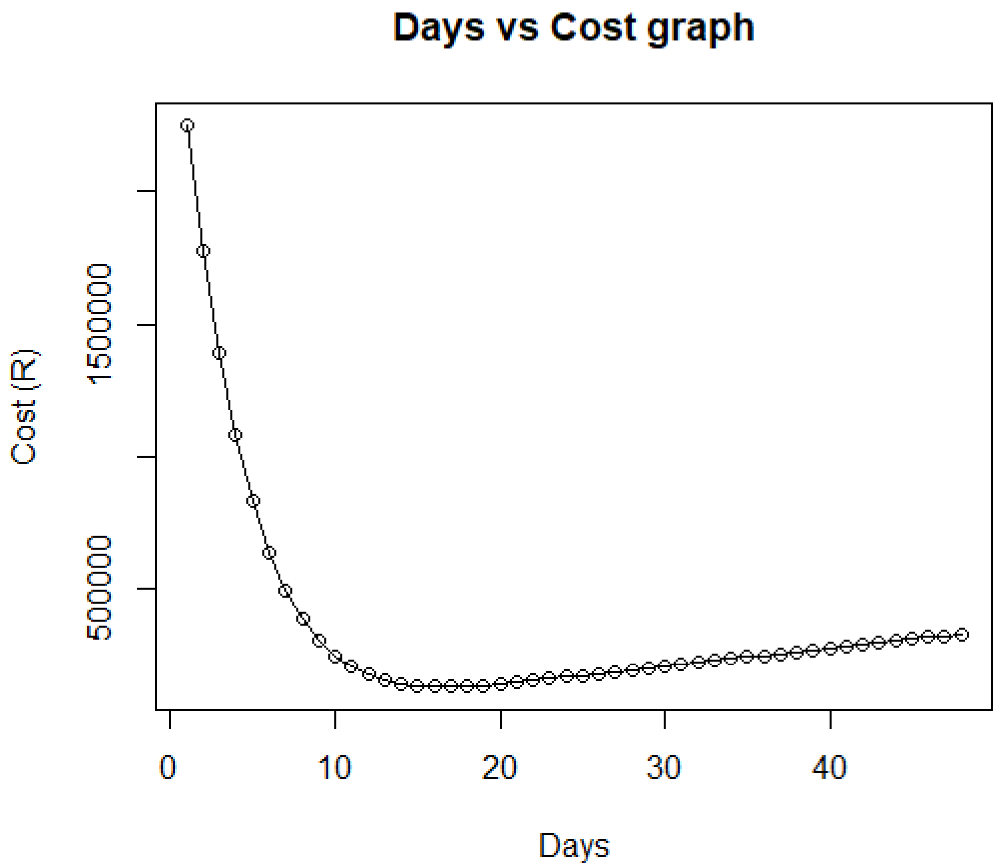


Figure 31: days vs SALES

q quantile-quantile graph

DOE and MANOVA

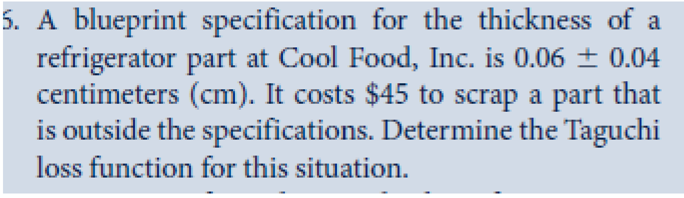
In this component of the assignment, a hypothesis test was carried out utilizing data from the technology class using Multivariate Analysis of Variance (MANOVA).

The following question was used to establish the testing for the test: Do our clients' ages and the cost of the goods they purchase have an impact on why they choose a particular technological product? So, "the age of customers and pricing of technical items determine the reason why a given technological product is acquired" is our null hypothesis (HO). Therefore, "the age of buyers does not affect the reason why a given technological product is bought" is our first alternative hypothesis (H1). The pricing of technological products has little bearing on why a particular technological product is purchased, according to our second alternative theory. The age of buyers has no bearing on the rationale behind the purchase of a particular technical product, according to our third alternative hypothesis (H3).

Discussion of MANOVA results

Part 6- Reliability of the service and products

Part 6.1



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

Figure 32:Problem 6-Textbok

$K = 45/0.04^2 = \$\ 28125$

$L = 28125(y-0.06)^2$

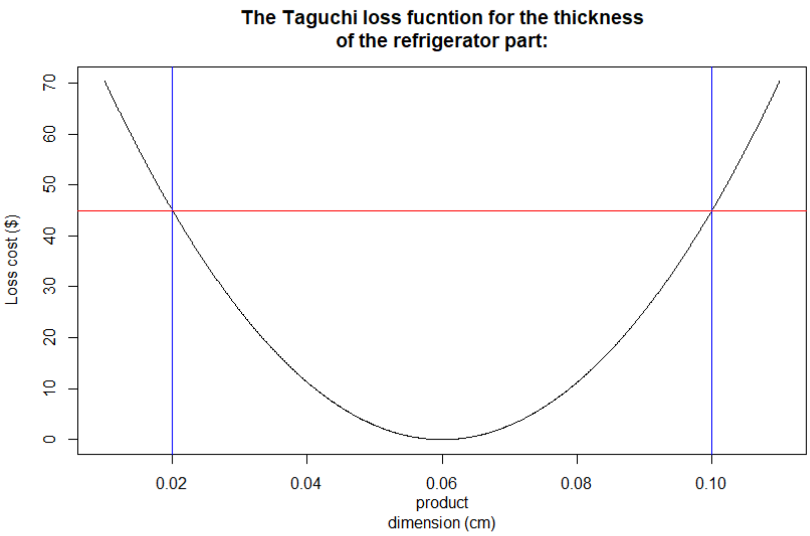


Figure 33:graphical representation of problem 6

From the graph we can see that at a dimension of 0.02 an 0.10 cm the loss in $ is 45. The graph is symmetrical about point 0.06. The Taguchi loss decreases from point 0 until point 0.06, the increases again from point 0.06 beyond 0.1.

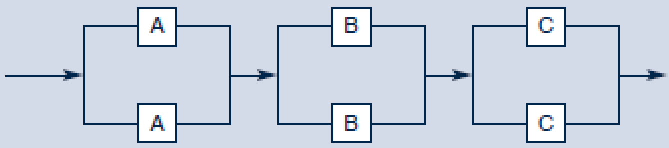
7. A team was formed to study the refrigerator part at Cool Food, Inc. described in Problem 6. While continuing to work to find the root cause of scrap, they found a way to reduce the scrap cost to $35 per part.
   a. Determine the Taguchi loss function for this situation.
   b. If the process deviation from target can be reduced to 0.027 cm, what is the Taguchi loss?

Figure 34:Problem 7-Textbook

a. $K=35/0.04^2 = \$\ 21875$
   $L = 21875*(y-0.06)^2$
b. $L = 21875 * 0.027^2 = \$\ 15.95$

Part 6.2

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



Production System

The reliabilities of the machines are as follows:

Machine	Reliability
A	0.85
B	0.92
C	0.90

a. Analyze the system reliability, assuming only one machine at each stage (all the backup machines are out of operation).
b. How much is the reliability improved by having two machines at each stage?

Figure 35:Problem 27-Textbook

a. $Rs = R_A + R_B + R_C$
   $= 0.85 \times 0.92 \times 0.9 = 0.961$
b. $= (1-(1-0.85)^2 \times 1-(1-0.92)^2) \times (1-(1-0.90)^2)$
   $= 0.9775 \times 0.9936 \times 0.99 = 0.961$

It is clear from the two results presented above that when both machines are operating at each stage, the reliability of the system is significantly increased.

Part 6.3 Binomial probability

This component of the work calculates and analyses the probability of how many days per year can be expected for reliable delivery timings.

Below is the binomial function:

$$f(x) = \frac{n!}{x!\,(n-x)!}p^x(1-p)^{n-x}$$

P = 0.2
A code is written on R to evaluate the probability the probability of a succession of 21 drivers available is 0.0134, where alpha was defined as to be equal to 5%. This means that the number of days we can expect a reliable delivery time in a year is:
(365 x 0.0134= 4.891) = 5 days

Conclusion

Data wrangling was necessary to ensure that the analysis that was deduced was of high quality and had integrity because the presented data contained a lot of missing values. The measurements of central tendencies (mean, mode, range, etc.) that each feature of the dataset included were described in a data quality report that was created after the wrangling.

X-bar charts and s-control charts were examined in the third and fourth sections of the assignment. It was found that even though the company is normally capable of achieving the requirements, requirements are occasionally not met. This is presumably caused by a change in drivers, a car breakdown, etc. Moreover, trends were examined. The study of many factors was also done. Type II errors were shown to be more likely to occur than Type I errors, according to research. With this outcome, it was found that process flaws and problems weren't always obvious.

Exercises from the textbook are done as instructed in the assignment brief.

References

Evans, J.R. & Lindsay W.M. Managing for Quality and Performance Excellence. 10th Edition. United States: Conveo

Jacob Anhoej. Control Charts with qicharts for R(2021/04/20)
https://cran.r-project.org/web/packages/qicharts/vignettes/controlcharts.html

QA344 Statistics.pdf
https://learn.sun.ac.za/pluginfile.php/3514418/mod_resource/content/1/QA344%20Statistics.pdf