ECSA REPORT

Student Number: 27141748

Contents

Introduction

The following report presents a hyper-focused analysis of data from manufacturing and service processes. The report focuses on key aspects of statistical process control and process capability analysis. The main focus of the data analysis is to assess the performance as well as the reliability of any respective process using a combination of descriptive statistics, capability indices, and SPC tools, to identify areas for improvement and optimize operational efficiency. Throughout the report a variety of data-related challenges are addressed using efficient, compact re-usable code. The results are further interpreted using various statistical techniques. The calculations and interpretation of process capability indices are used in “Part 3”, along with a discussion on the clarification of products as capable or not capable based on Upper Specification Limits (USL). Throughout the report I’ve identified signals of process instability, calculating errors and out-of-bound samples and their implications. The report also investigates the loss function in regards to process optimization and evaluates the impact of delivery times on profitability. The results from analysis have provided actionable insight that is able to guide improvements in process reliability and service delivery.

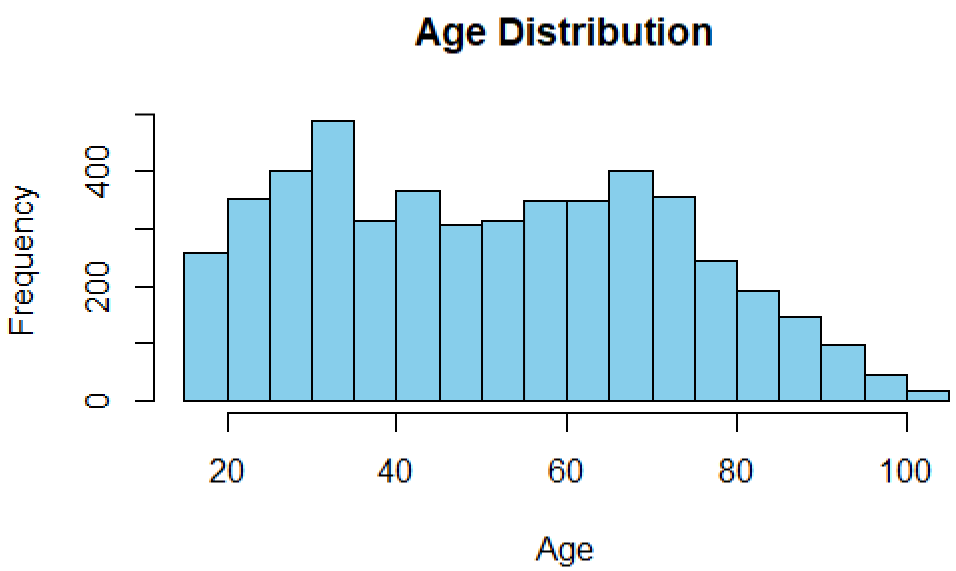
Part 1

Data

	vars <dbl>	n <dbl>	mean <dbl>	sd <dbl>	median <dbl>
CustomerID*	1	5000	2500.50	1443.52	2500.5
Gender*	2	5000	1.56	0.58	2.0
Age	3	5000	51.55	21.22	51.0
Income	4	5000	80797.00	33150.11	85000.0
City*	5	5000	3.99	2.00	4.0

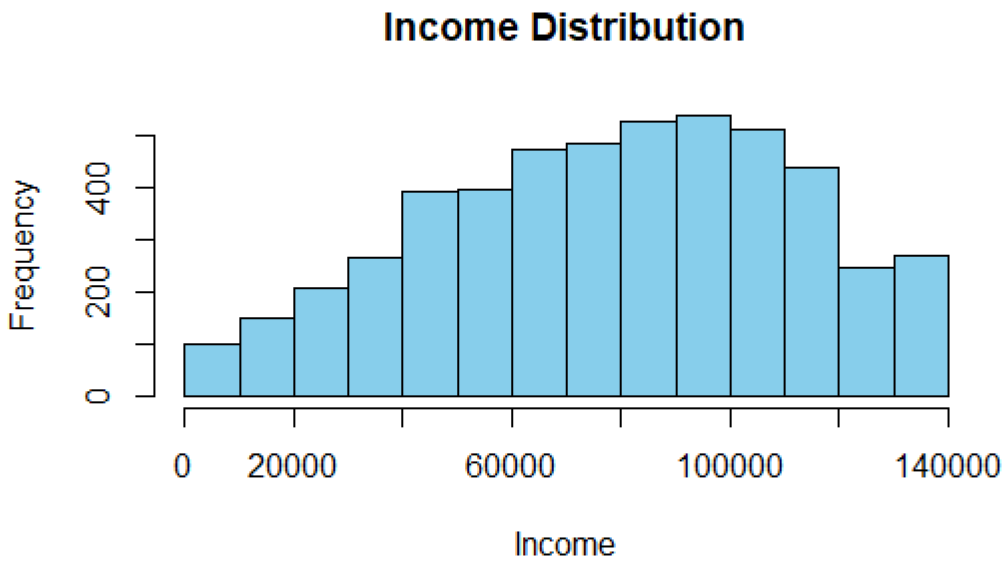
This is the outline of our data, CustomerID, Gender and City have an asterisk next to them because the data these columns represent is represented in the forms of characters not numbers.

Age

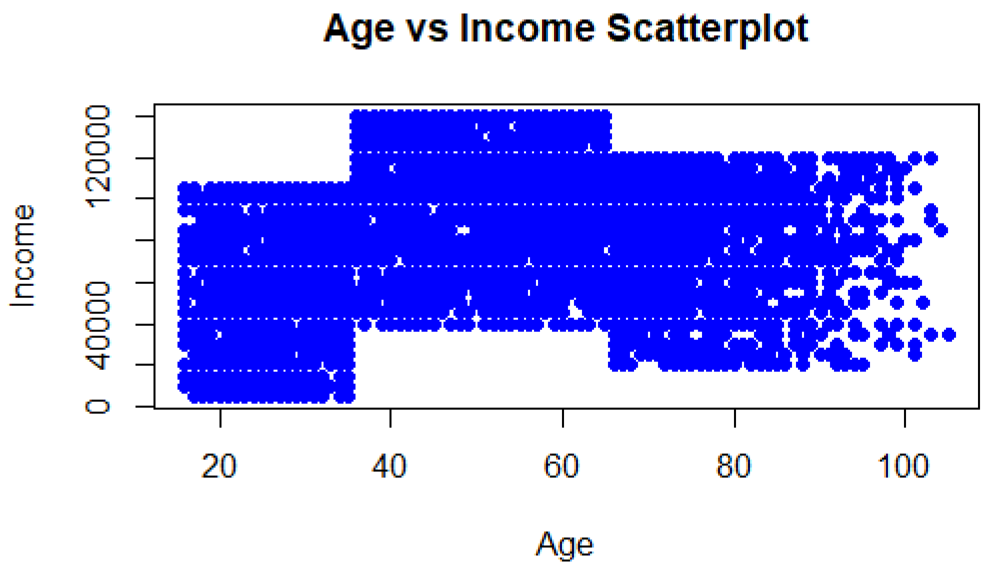


Most customers are middle aged, with a large number of customers being in their late 30s, late 60s and early 40s, this shows that the demographic that we most appeal to are working adults who are either in the later stages of their careers, approaching the end of it or they're already retired.

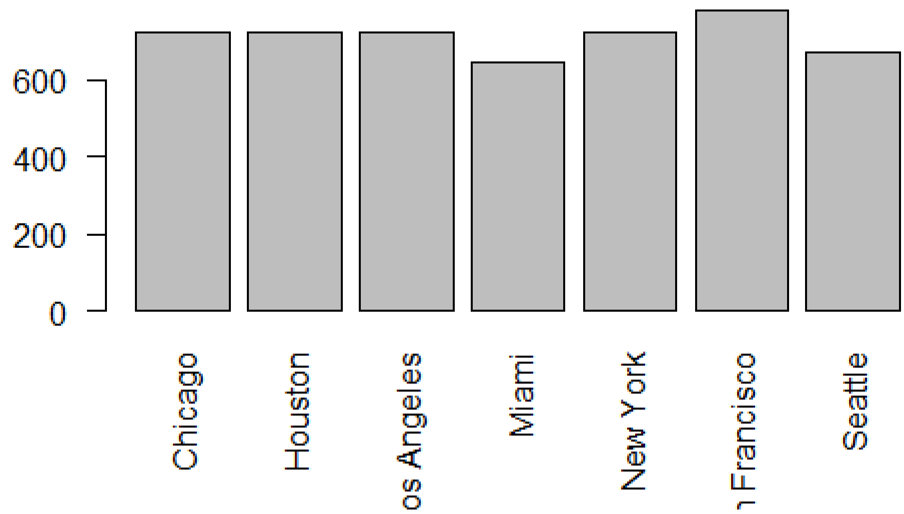
Income



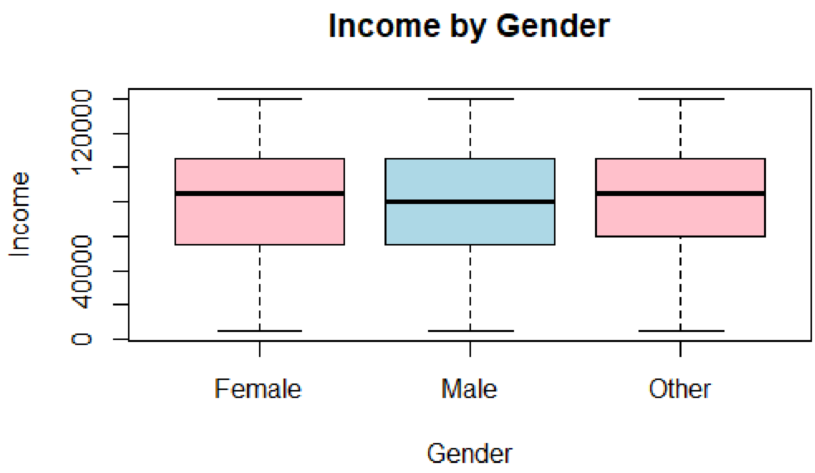
The average customer income is $80,797, with a standard deviation of $33,150. Most incomes are on the higher end, which means many of our customers are high-earners. This makes them a strong audience for premium products and services. Using this information, we can adjust our marketing to focus more on luxury offerings that fit our customers' income levels. This underscores the financial importance of calibrating marketing initiatives to cater to the spending patterns of our affluent customer base.



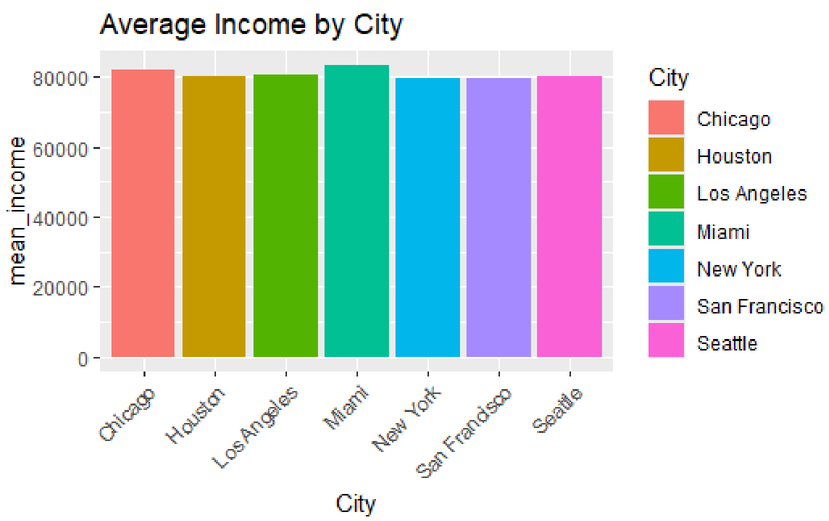
From the scatterplot we can see that roughly ages 38 – 65 have the highest income, whereas ages 16 – 35 have the lowest income.



From the bar plot we can see that San Francisco has the most customers, whereas Chicago, Houston and Los Angeles have the same number of customers.



From the box plot we can see that on average, men earn the highest.



The average income amongst customers is highest in Miami followed by Chicago and the rest are relatively the same.

Product Data

```
Cloud Subscription           Keyboard          Laptop
                10                 10              10
           Monitor              Mouse        Software
                10                 10              10
```

Each product type has 10 different variations

Category <chr>	Total_Revenue <dbl>
Laptop	52175.45
Monitor	50141.70
Keyboard	46381.72
Mouse	45854.65
Software	38143.44
Cloud Subscription	36918.61

Product Category “Laptop” made the most amount of money from sales, however, product category “Monitor” showed to make the most profit from sales.

Category <chr>	Total_Profit <dbl>
Monitor	207.27
Mouse	206.68
Laptop	206.23
Cloud Subscription	205.53
Keyboard	201.61
Software	200.38

Software sales need to be more closely analyzed; it makes low sales and low profit in comparison to the rest of the products, but this could simply mean that it is not as popular of a product choice as the others, that doesn’t necessarily mean that that’s a bad thing.

Product Head Office Data

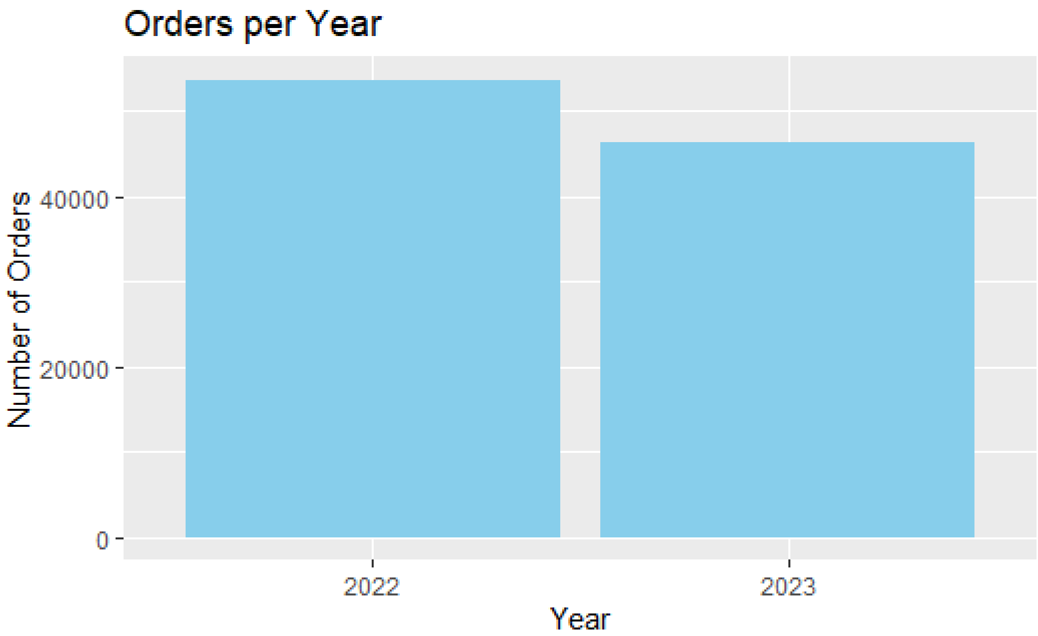
```
Cloud Subscription            Keyboard              Laptop
                60                  60                  60
           Monitor               Mouse            Software
                60                  60                  60
```

Each product type has 60 different sub types. This is much more than the data we had from the sampled product data.

Category <chr>	Total_Revenue <dbl>
Mouse	268734.0
Software	267431.6
Monitor	267404.7
Cloud Subscription	263202.6
Keyboard	262829.1
Laptop	258344.3

Product Category “Mouse” made the most amount of money from sales, whereas product category “Cloud Subscription” made the most profit from sales.

Sales 2022 & 2023



2022 Had the most orders in a year.

orderYear <dbl>	orderMonth <dbl>	Order_Count <int>
2022	1	3305
2022	2	4816
2022	3	4744
2022	4	4831
2022	5	4819
2022	6	4665
2022	7	4762
2022	8	4755
2022	9	4738
2022	10	4710
2022	11	4734
2022	12	2848
2023	1	2829
2023	2	4096
2023	3	4107
2023	4	4128
2023	5	4077
2023	6	4059
2023	7	4137
2023	8	4045
2023	9	4083
2023	10	4078
2023	11	4178
2023	12	2456

For 2022 April had the most orders, and for 2023 November had the most orders.

orderYear <dbl>	Avg_Picking_Hours <dbl>	Avg_Delivery_Hours <dbl>
2022	14.68171	17.51089
2023	14.71146	17.43649

2022 Had a higher average of delivery hours whereas 2023 had a higher average of picking hours.

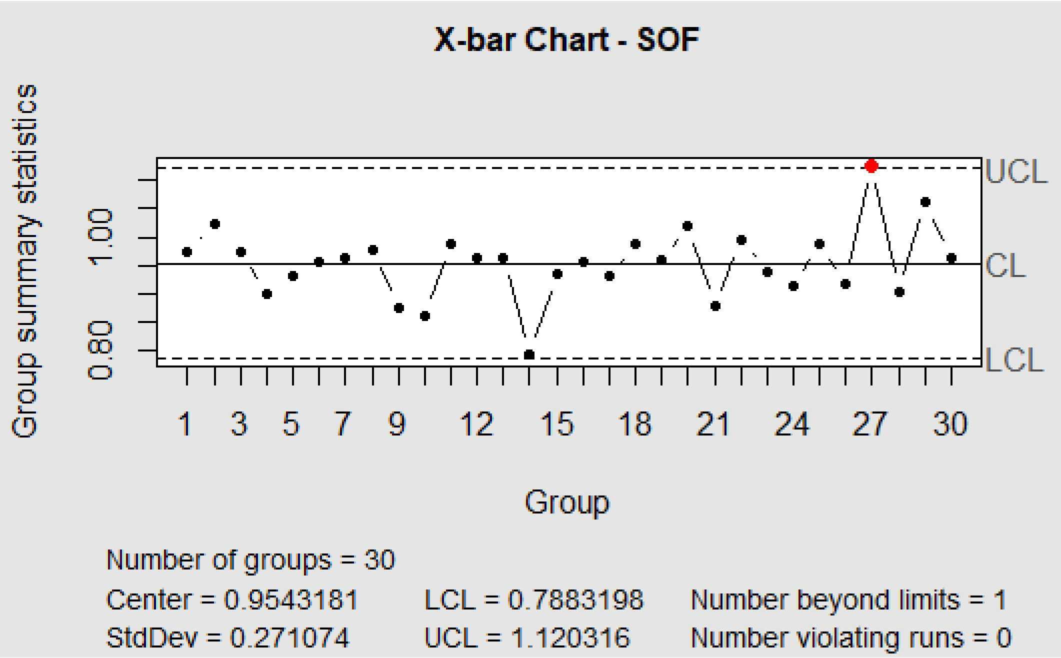
data$orderTime <dbl>	n <int>
12	6483
15	6436
11	6390
14	6350
18	6338
10	6313
17	6296
9	6252
16	6218
7	4408

Most orders were made during the day.

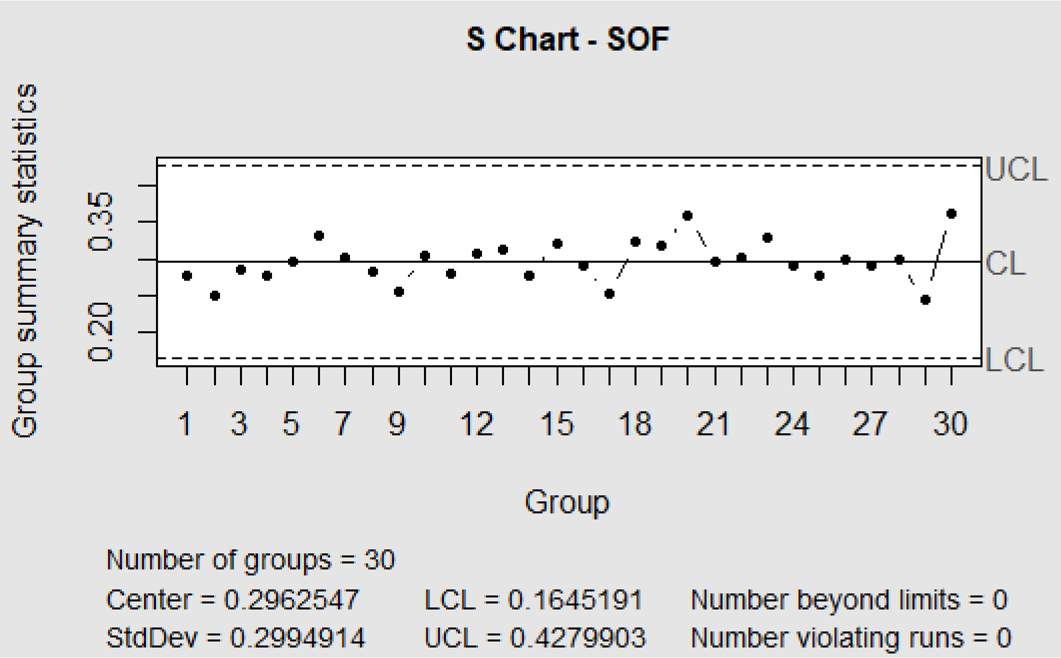
Part 3

3.1

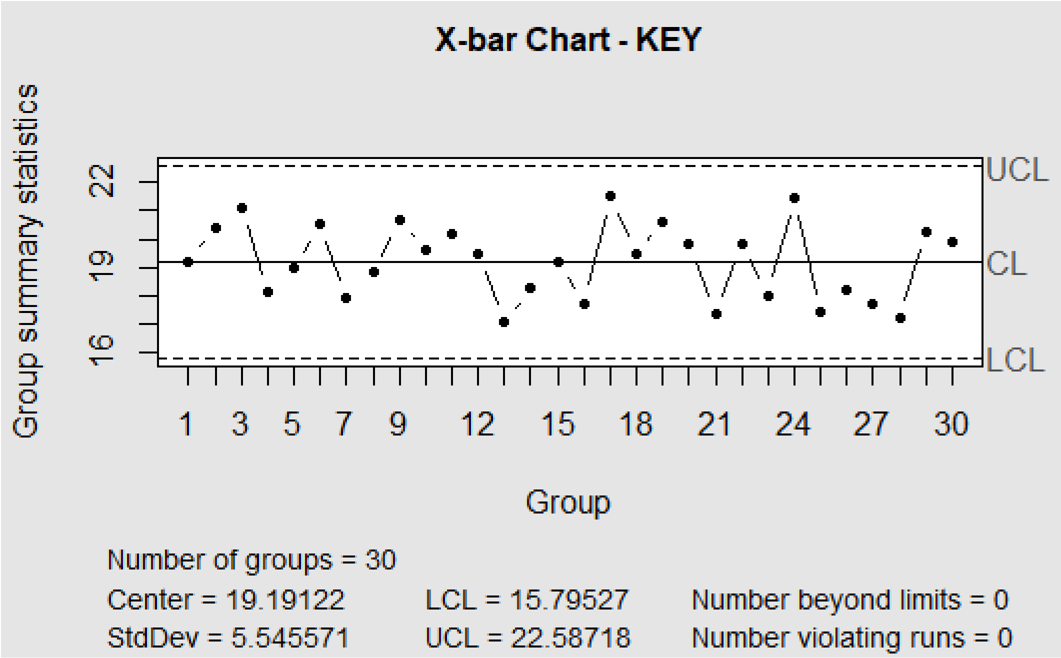
Below are the s and x-bar charts for the first 30 samples of each type of product:



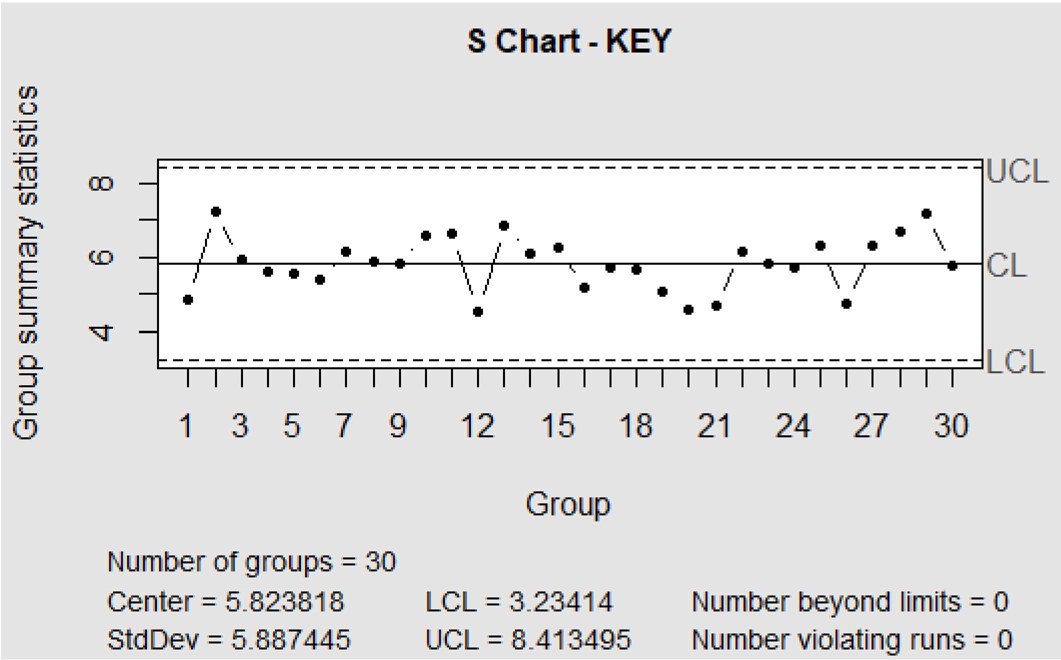
Here is an X-bar chart for the first 30 samples of the Software product, the LCL and UCL values are shown above with all except for one sample lying outside of the UCL and LCL bounds with no violating runs.



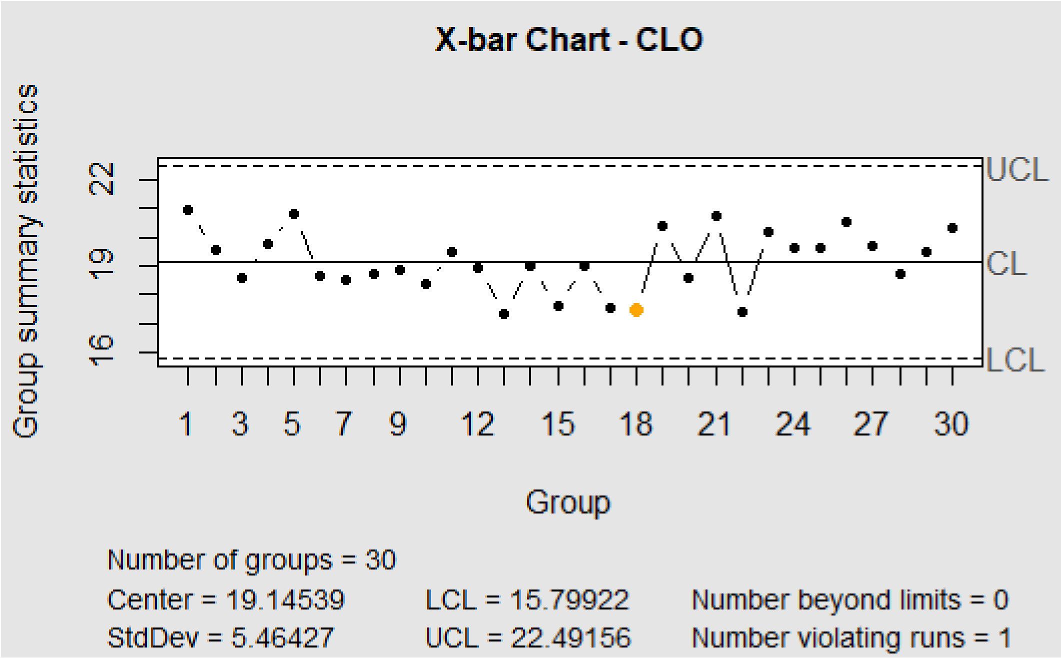
Here is an S chart for the first 30 samples of the Software product, the LCL and UCL values are shown above with all samples lying within the UCL and LCL bounds with no violating runs.



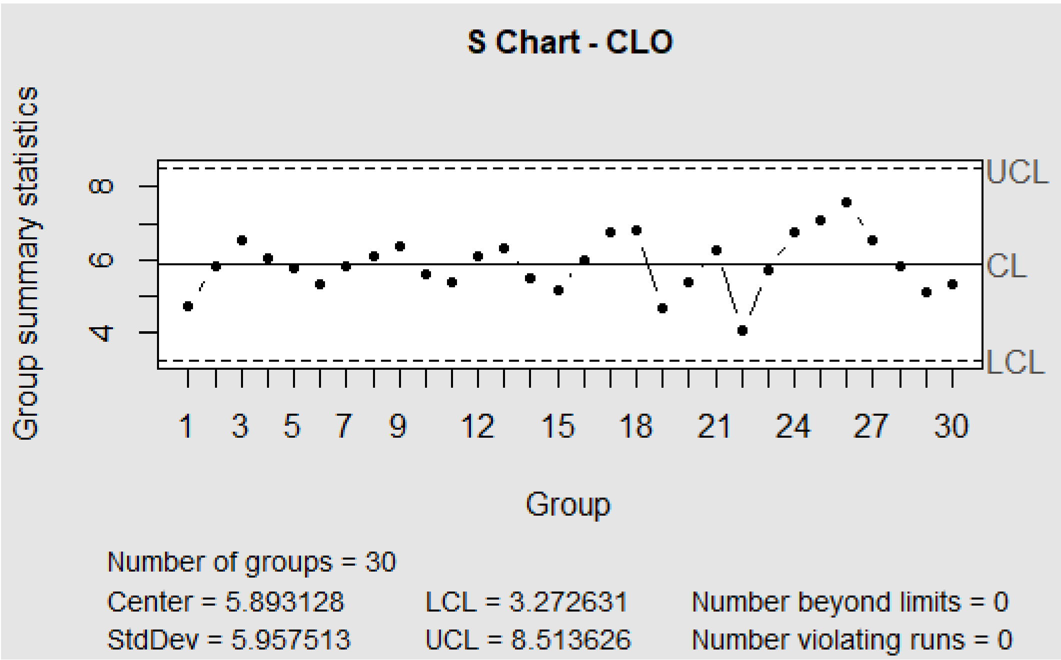
Here is an X-bar chart for the first 30 samples of the Keyboard product, the LCL and UCL values are shown above with all samples lying within the UCL and LCL bounds with no violating runs.



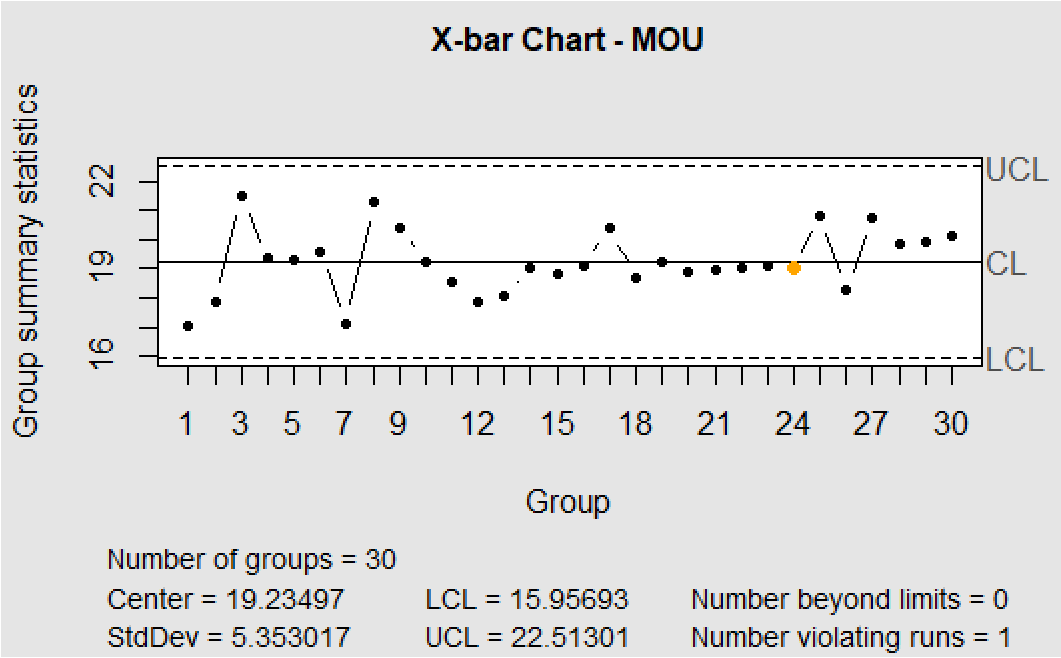
Here is an S chart for the first 30 samples of the Keyboard product, the LCL and UCL values are shown above with all samples lying within the UCL and LCL bounds with no violating runs.



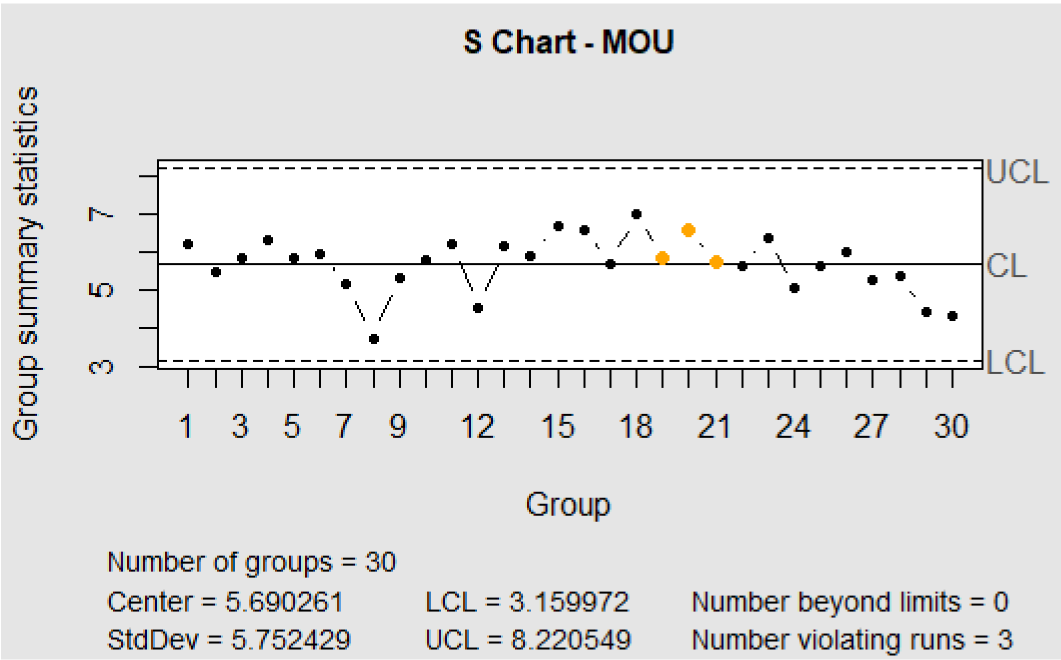
Here is an X-bar chart for the first 30 samples of the Cloud product, the LCL and UCL values are shown above with all samples lying within the UCL and LCL bounds with 1 violating run. This could possibly indicate process instability. Investigative action must be taken.



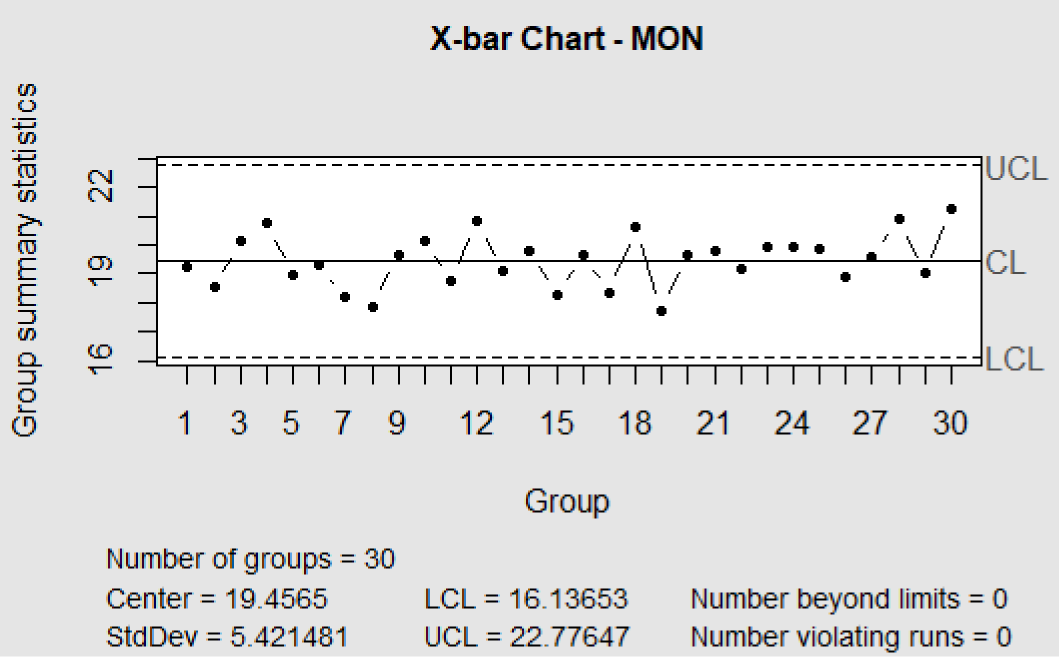
Here is an S chart for the first 30 samples of the Cloud product, the LCL and UCL values are shown above with all samples lying within the UCL and LCL bounds with no violating runs. This could possibly indicate process instability. Investigative action must be taken.



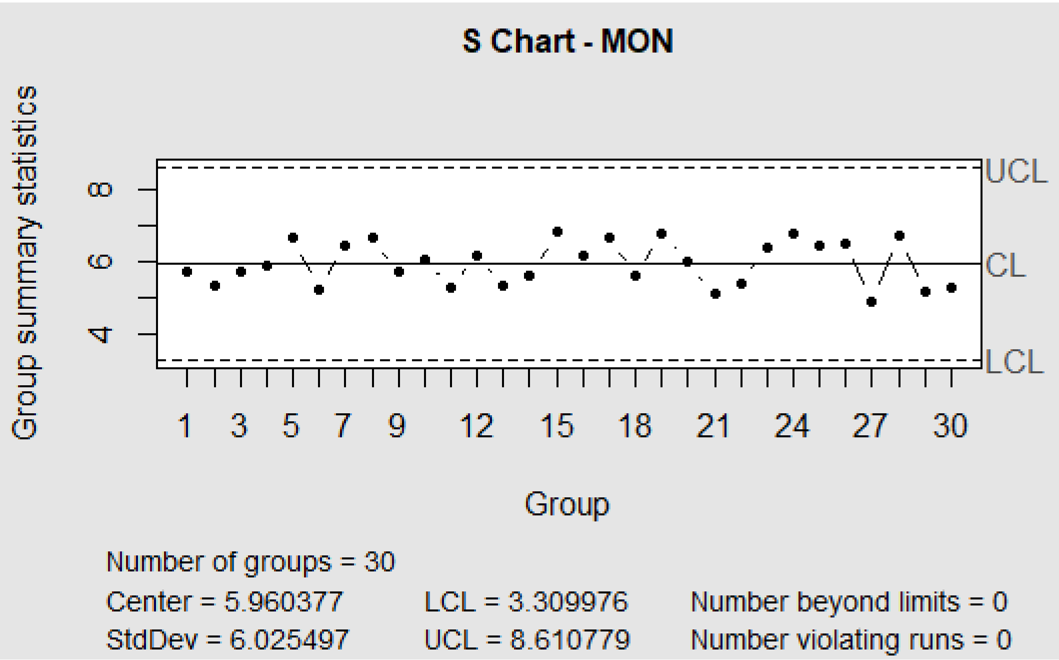
Here is an X-bar chart for the first 30 samples of the Mouse product, the LCL and UCL values are shown above with all samples lying within the UCL and LCL bounds with 1 violating run. This could possibly indicate process instability. Investigative action must be taken.



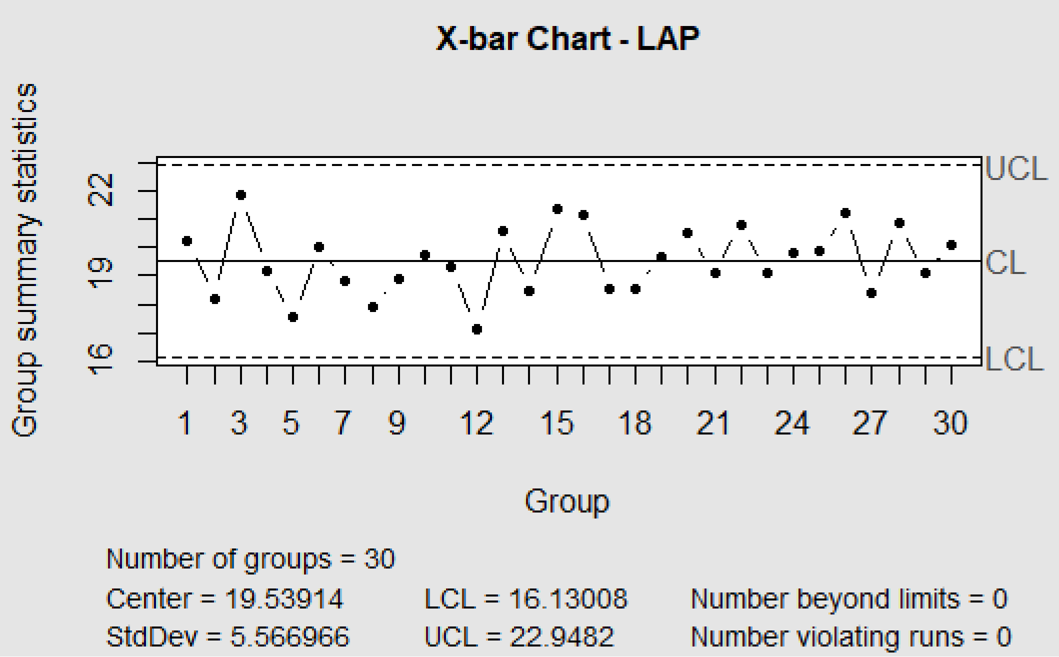
Here is an X-bar chart for the first 30 samples of the Mouse product, the LCL and UCL values are shown above with all samples lying within the UCL and LCL bounds with 3 violating runs. This is more than 2 violating runs meaning that the process is unstable and corrective action must be taken.



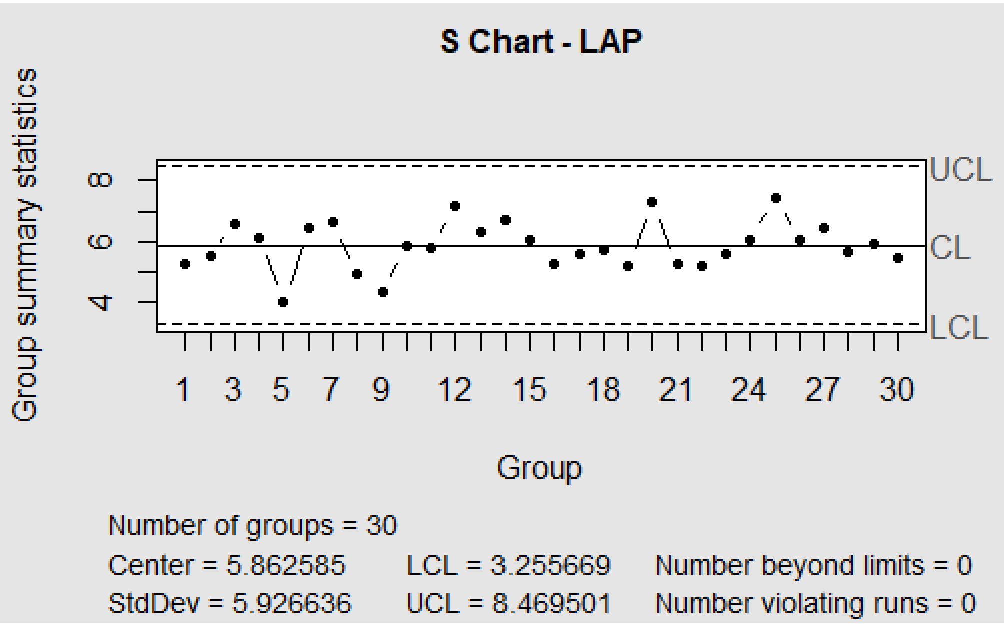
Here is an X-bar chart for the first 30 samples of the Monitor product, the LCL and UCL values are shown above with all samples lying within the UCL and LCL bounds with no violating runs.



Here is an S chart for the first 30 samples of the Monitor product, the LCL and UCL values are shown above with all samples lying within the UCL and LCL bounds with no violating runs.

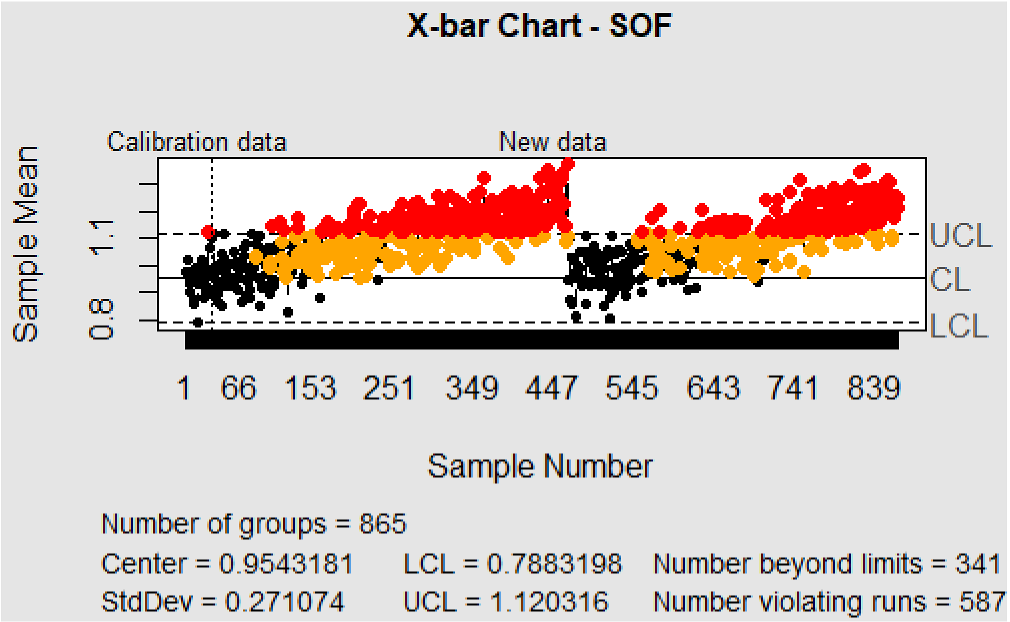


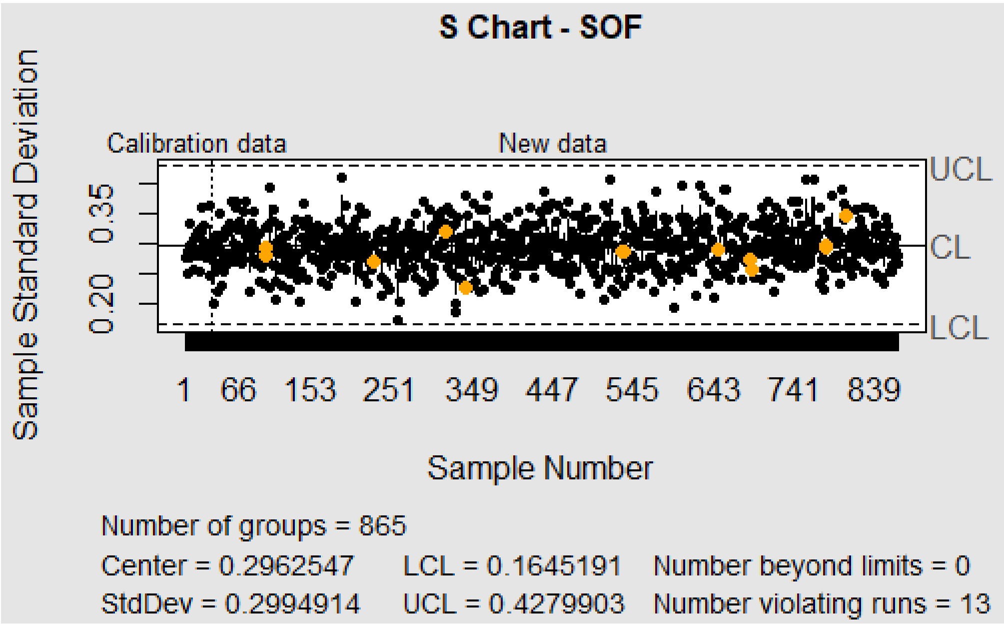
Here is an X-bar chart for the first 30 samples of the Laptop product, the LCL and UCL values are shown above with all samples lying within the UCL and LCL bounds with no violating runs.

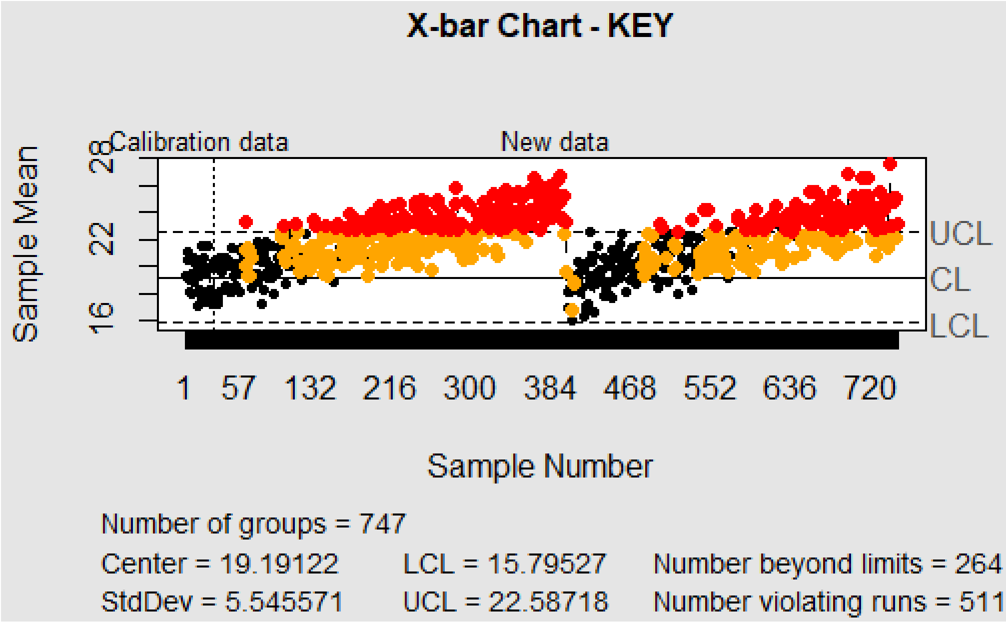


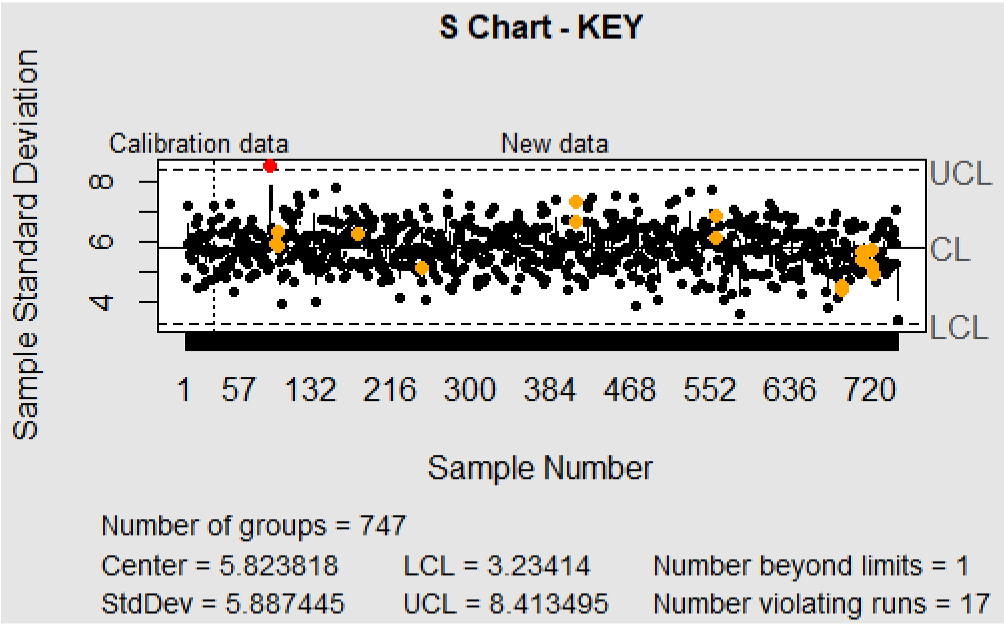
Here is an S chart for the first 30 samples of the Laptop product, the LCL and UCL values are shown above with all samples lying within the UCL and LCL bounds with no violating runs.

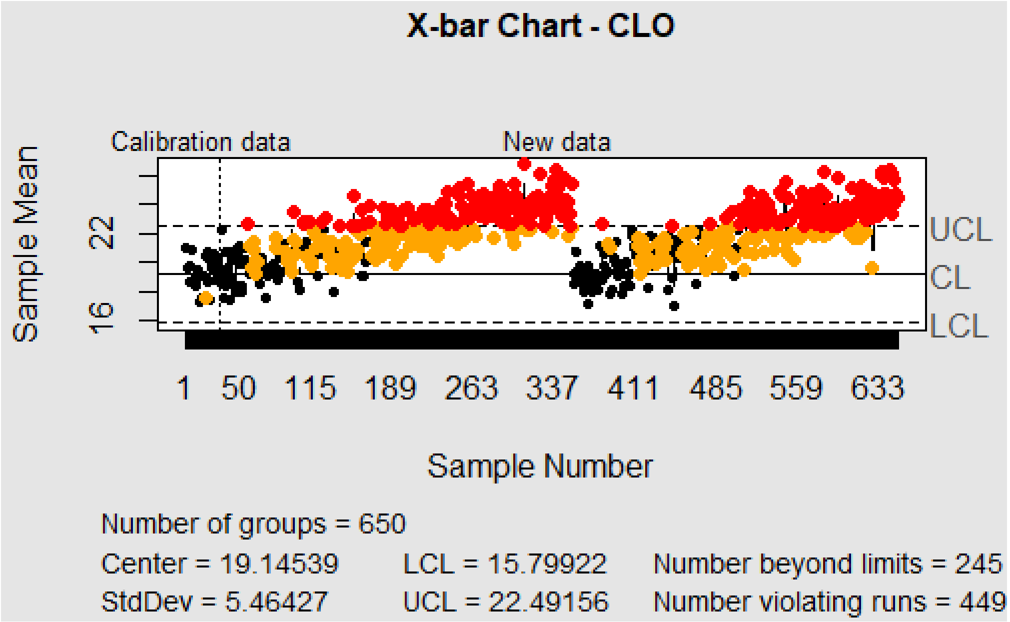
3.2

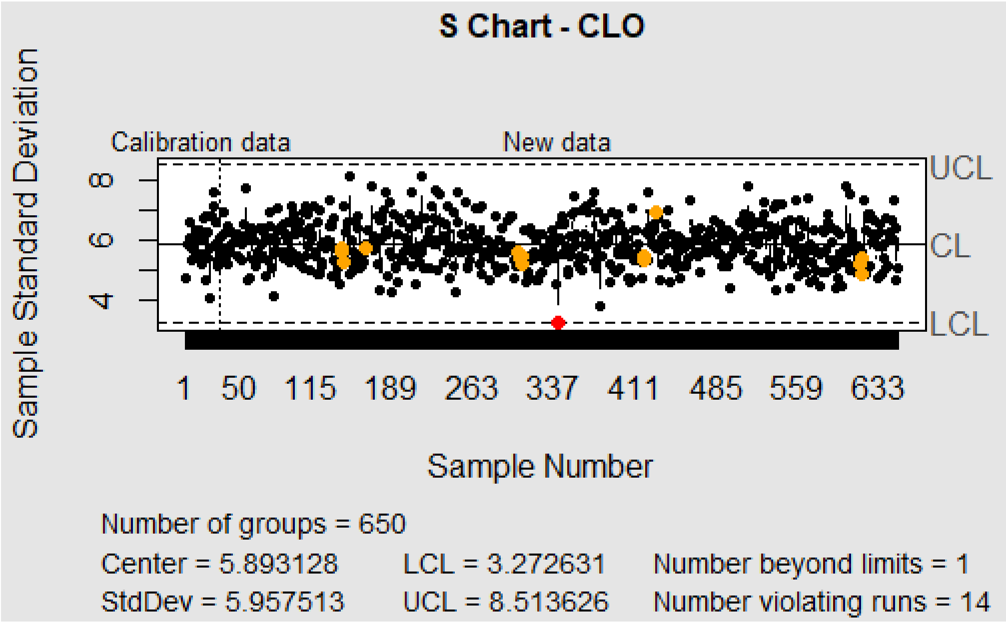

X-bar Chart - SOF
Calibration data
New data
Sample Mean
0.8
1.1
UCL
CL
LCL
1 66 153 251 349 447 545 643 741 839
Sample Number
Number of groups = 865
Center = 0.9543181
StdDev = 0.271074
LCL = 0.7883198
UCL = 1.120316
Number beyond limits = 341
Number violating runs = 587

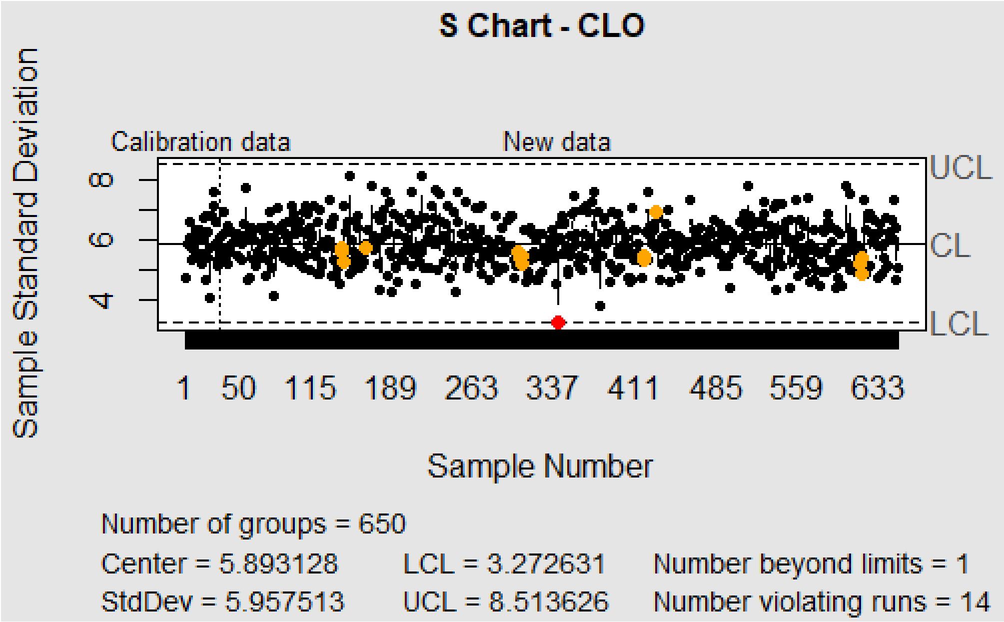

S Chart - SOF
Calibration data
New data
Sample Standard Deviation
0.20
0.35
UCL
CL
LCL
1 66 153 251 349 447 545 643 741 839
Sample Number
Number of groups = 865
Center = 0.2962547
StdDev = 0.2994914
LCL = 0.1645191
UCL = 0.4279903
Number beyond limits = 0
Number violating runs = 13


X-bar Chart - KEY
Calibration data
New data
Sample Mean
16
22
28
UCL
CL
LCL
1 57 132 216 300 384 468 552 636 720
Sample Number
Number of groups = 747
Center = 19.19122
StdDev = 5.545571
LCL = 15.79527
UCL = 22.58718
Number beyond limits = 264
Number violating runs = 511

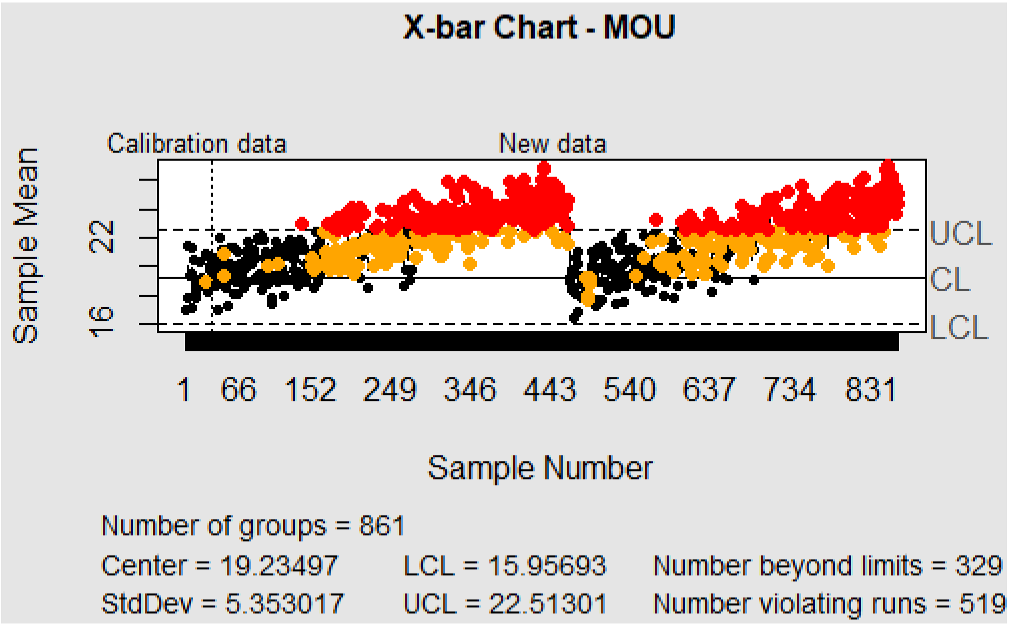

S Chart - KEY
Calibration data
New data
Sample Standard Deviation
4
6
8
UCL
CL
LCL
1 57 132 216 300 384 468 552 636 720
Sample Number
Number of groups = 747
Center = 5.823818
StdDev = 5.887445
LCL = 3.23414
UCL = 8.413495
Number beyond limits = 1
Number violating runs = 17


X-bar Chart - CLO
Calibration data
New data
Sample Mean
16
22
UCL
CL
LCL
1 50 115 189 263 337 411 485 559 633
Sample Number
Number of groups = 650
Center = 19.14539
StdDev = 5.46427
LCL = 15.79922
UCL = 22.49156
Number beyond limits = 245
Number violating runs = 449

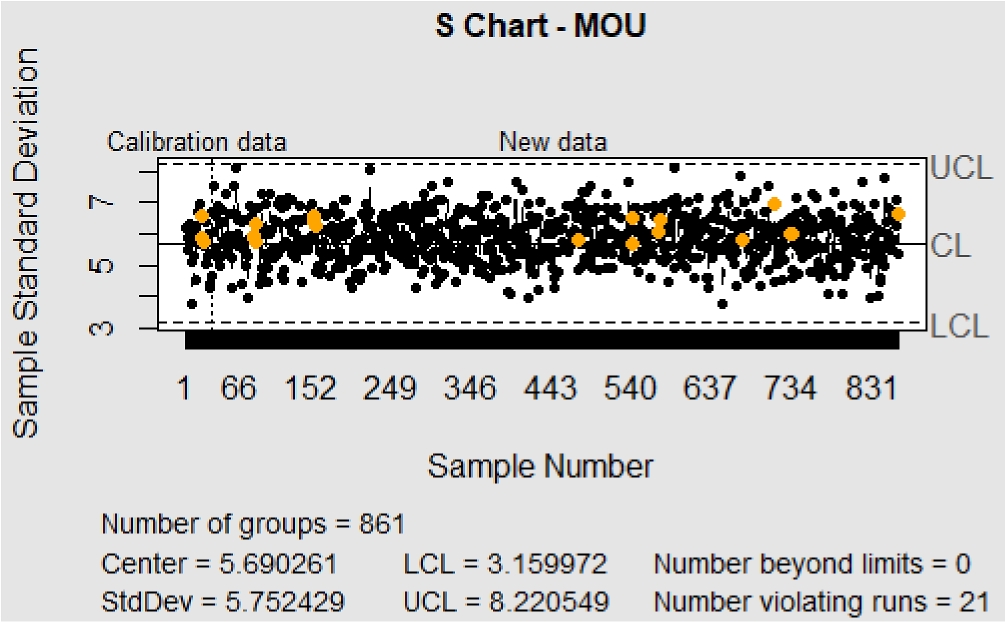

S Chart - CLO
Calibration data
New data
Sample Standard Deviation
4
6
8
UCL
CL
LCL
1 50 115 189 263 337 411 485 559 633
Sample Number
Number of groups = 650
Center = 5.893128
StdDev = 5.957513
LCL = 3.272631
UCL = 8.513626
Number beyond limits = 1
Number violating runs = 14

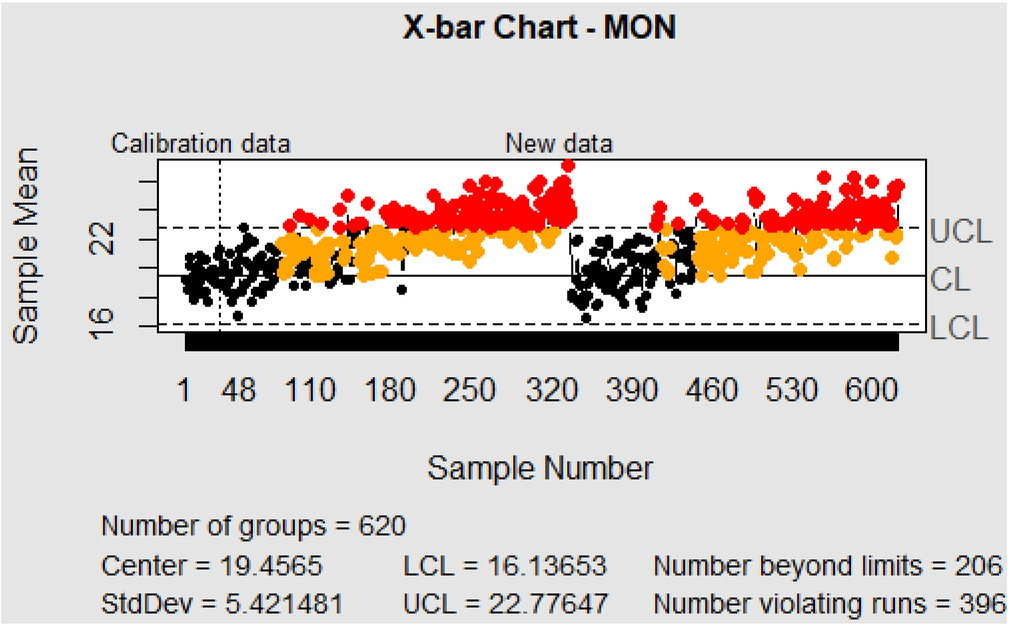


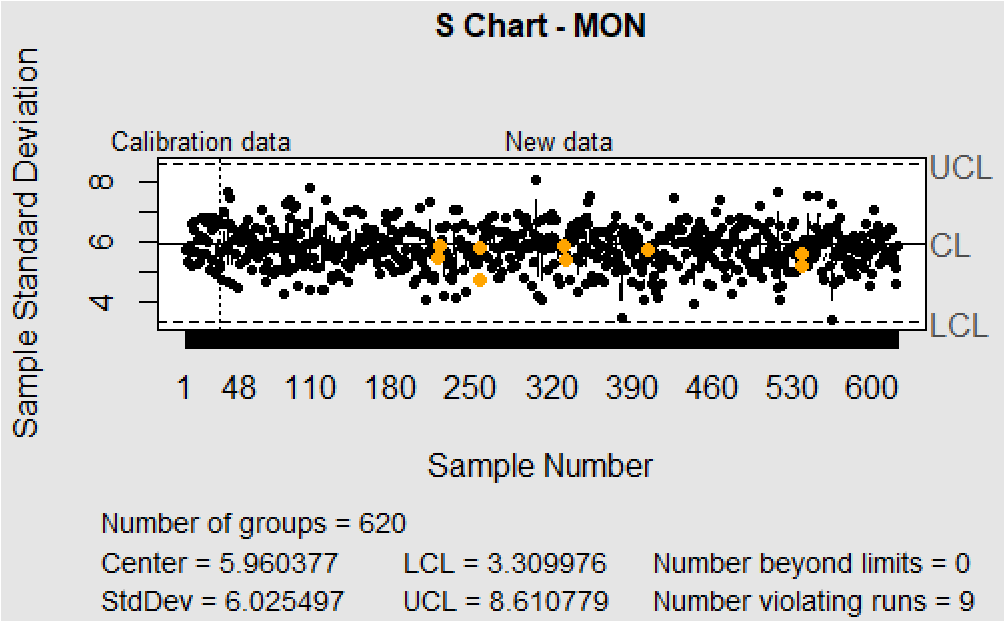
S Chart - CLO
Calibration data
New data
Sample Standard Deviation
4
6
8
UCL
CL
LCL
1 50 115 189 263 337 411 485 559 633
Sample Number
Number of groups = 650
Center = 5.893128
StdDev = 5.957513
LCL = 3.272631
UCL = 8.513626
Number beyond limits = 1
Number violating runs = 14

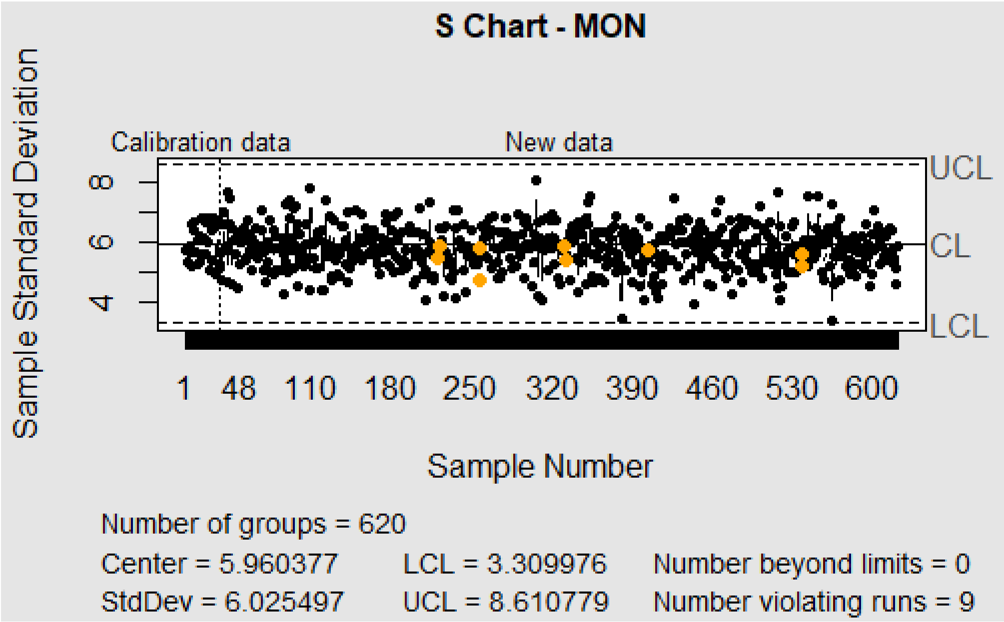


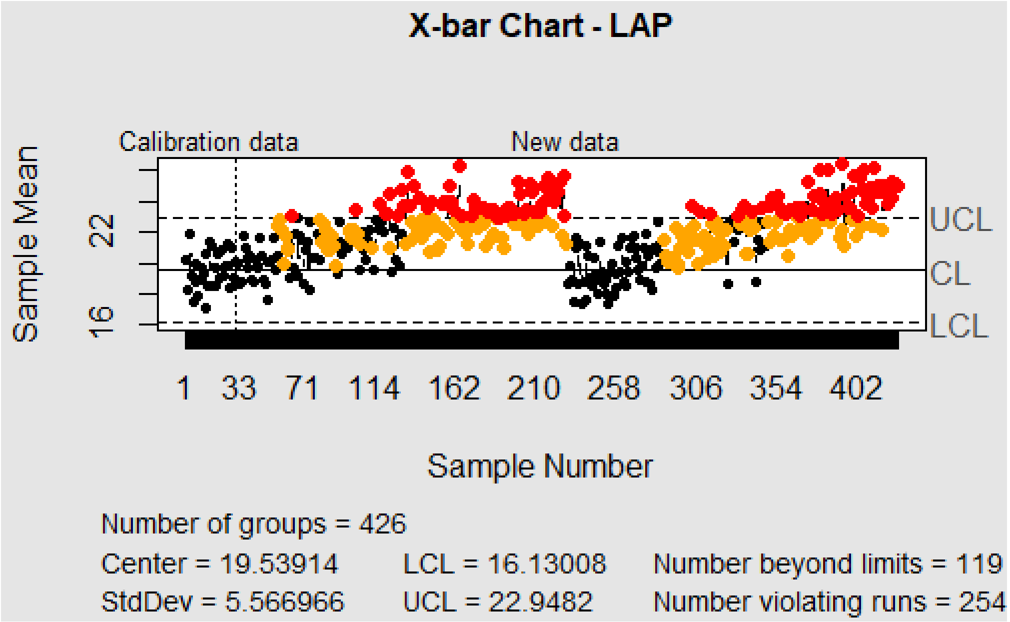
X-bar Chart - MOU
Calibration data
New data
Sample Mean
16
22
UCL
CL
LCL
1 66 152 249 346 443 540 637 734 831
Sample Number
Number of groups = 861
Center = 19.23497
StdDev = 5.353017
LCL = 15.95693
UCL = 22.51301
Number beyond limits = 329
Number violating runs = 519

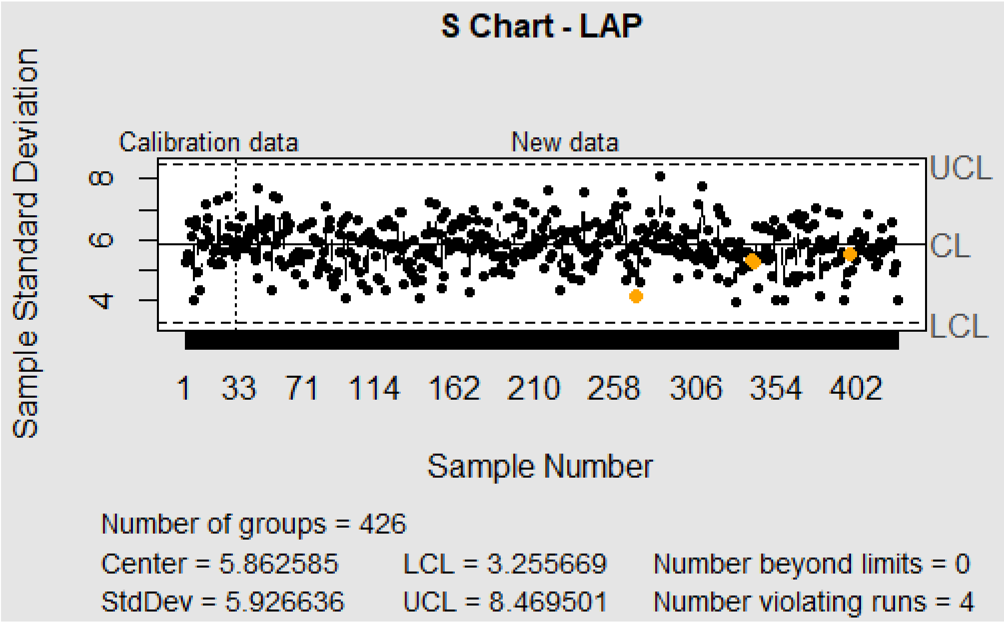

S Chart - MOU
Sample Standard Deviation
Calibration data
New data
UCL
CL
LCL
3
5
7
1
66
152
249
346
443
540
637
734
831
Sample Number
Number of groups = 861
Center = 5.690261
StdDev = 5.752429
LCL = 3.159972
UCL = 8.220549
Number beyond limits = 0
Number violating runs = 21


X-bar Chart - MON
Sample Mean
Calibration data
New data
UCL
CL
LCL
16
22
1
48
110
180
250
320
390
460
530
600
Sample Number
Number of groups = 620
Center = 19.4565
StdDev = 5.421481
LCL = 16.13653
UCL = 22.77647
Number beyond limits = 206
Number violating runs = 396


S Chart - MON
Calibration data
New data
Sample Standard Deviation
UCL
CL
LCL
4
6
8
1
48
110
180
250
320
390
460
530
600
Sample Number
Number of groups = 620
Center = 5.960377
StdDev = 6.025497
LCL = 3.309976
UCL = 8.610779
Number beyond limits = 0
Number violating runs = 9


S Chart - MON
Calibration data
New data
Sample Standard Deviation
UCL
CL
LCL
4
6
8
1
48
110
180
250
320
390
460
530
600
Sample Number
Number of groups = 620
Center = 5.960377
StdDev = 6.025497
LCL = 3.309976
UCL = 8.610779
Number beyond limits = 0
Number violating runs = 9





By observation, when looking at all of the data at once for each respective product type it would seem that almost all of these processes are out of control, they are all unstable and need to be corrected.

We start by using Rule A&C as violation indicators to quarantine the affected samples.

Quantity <int>	orderTime <int>	orderDay <int>	orderMonth <int>	orderYear <int>	pickingHours <dbl>	deliveryHours <dbl>	ProductType <chr>	Sample <int>
7	2	24	2	2022	10.7216667	31.0440	MOU	61
48	11	24	2	2022	11.7216667	17.0440	MOU	61
7	20	24	2	2022	11.7216667	30.0440	MOU	61
6	10	24	2	2022	12.7216667	17.0440	MOU	61
19	20	24	2	2022	12.7216667	13.0440	MOU	61
2	9	24	2	2022	13.7216667	12.0440	MOU	61
2	14	24	2	2022	13.7216667	13.0440	MOU	61
4	16	24	2	2022	13.7216667	31.0440	MOU	61
3	1	24	2	2022	13.7216667	23.0440	MOU	61
23	7	24	2	2022	13.7216667	20.0440	MOU	61
3	13	24	2	2022	13.7216667	14.0440	MOU	61
1	7	24	2	2022	13.7216667	10.0440	MOU	61
26	21	24	2	2022	13.7216667	27.0440	MOU	61
1	17	24	2	2022	14.7216667	25.0440	MOU	61
9	11	24	2	2022	14.7216667	16.0440	MOU	61
2	11	24	2	2022	14.7216667	19.0440	MOU	61

Now that we have the quarantined rows, we can move onto rapid diagnostics. Our goal here is to find obvious fixable causes quickly. We run the following checklist Measurement system, Material/input, Machine, Operator, Environment/power. We first repeat the measurement on a sample, if there's an output drift, we stop using that measurement system and re-calibrate.

Rule A – Sample Standard Deviation (S) Above the Upper Control Limit (UCL)

Rule A indicates that the delivery-hour variation within a specific sample or batch is abnormally high.

Possible causes in a delivery system include:

- Traffic related delays, road closures, or accidents on that route.

- Driver inconsistency (new or substitute drivers who are unfamiliar with routes)

- Batching errors, such as combining distant customers into one route

- System or device errors that recorded delivery times incorrectly

- Customer-sided issues, like unavailable recipients or access delays at delivery sites.

There are a number of short-term corrective actions that can be taken. The affected sample could be investigated immediately. Data accuracy could be validated (ensure that timestamps and system uploads were correct). External disruptions could be identified such as traffic incidents or severe weather. We could re-schedule or re-allocate deliveries temporarily to reduce pressure on the delayed routes.

There are also a number of long-term corrective actions that can be taken. Routing plans could be reviewed and optimized to reduce travel variability. Additional driver training could be provided to ensure consistent handling of routes. Real-time monitoring of delivery delays through GPS and alert systems could be introduced. We could also revise scheduling policies to prevent overloading routes or shifts.

Rule C – 4 or More Consecutive Sample Means (X) Beyond +2σ

This rule indicates a systematic upward shift in the average delivery hours. If four or more consecutive averages fall above the +2σ line, it shows that deliveries are consistently taking longer than normal – not due to random chance but due to a persistent shift in the process.

The possible causes are the same for this rule as the previous one. It shows that the process mean has shifted upward, meaning the typical delivery is taking longer than before. The system's on-time performance has declined, and customers may experience more frequent delays. This is a chronic performance issue, not just a one-time variation. Both long-term and short-term corrective actions are effective for both rules A & C.

3.3

ProductType <chr>	n_obs <int>	mean <dbl>	sd <dbl>	Cp <dbl>	Cpu <dbl>	Cpl <dbl>	Cpk <dbl>	Capable <lgl>
SOF	1000	0.957675	0.2937719	18.1546726	35.2227029	1.086642	1.0866423	TRUE
KEY	1000	19.265000	5.8165704	0.9169206	0.7298115	1.104030	0.7298115	FALSE
MOU	1000	19.317500	5.8275559	0.9151921	0.7254328	1.104951	0.7254328	FALSE
CLO	1000	19.214000	5.9446984	0.8971579	0.7169413	1.077375	0.7169413	FALSE
MON	1000	19.414000	5.9945003	0.8897044	0.6998637	1.079545	0.6998637	FALSE
LAP	1000	19.599000	5.9341123	0.8987584	0.6965939	1.100923	0.6965939	FALSE

As it stands, from observing the first 1000 deliveries per product type, only product type SOF is capable of meeting the VOC.

3.4

A.

```
ProductType: CLO
Total identified: 0
First up to 3 samples: None
Last up to 3 samples:  None

ProductType: KEY
Total identified: 1
First up to 3 samples: 89
Last up to 3 samples:  89

ProductType: LAP
Total identified: 0
First up to 3 samples: None
Last up to 3 samples:  None

ProductType: MON
Total identified: 2
First up to 3 samples: 109, 305
Last up to 3 samples:  109, 305

ProductType: MOU
Total identified: 3
First up to 3 samples: 61, 222, 591
Last up to 3 samples:  61, 222, 591

ProductType: SOF
Total identified: 12
First up to 3 samples: 103, 190, 435
Last up to 3 samples:  754, 762, 798
```

B.

ProductType <chr>	max_run_len <int>	start_sample <int>	end_sample <int>	S_lower_1 <dbl>	S_upper_1 <dbl>
CLO	27	470	496	5.1291449	6.6571117
KEY	18	647	664	5.0773058	6.5703295
LAP	15	409	423	5.0678345	6.6573360
MON	12	155	166	5.3508783	6.5698766
MOU	12	224	235	4.9409731	6.4395483
SOF	10	150	159	0.2675771	0.3249323

C.

```
ProductType: CLO
Total runs of length >=4: 20
First up to 3 runs: 126-129 (len=4); 166-169 (len=4); 175-178 (len=4)
Last up to 3 runs:  551-555 (len=5); 557-626 (len=70); 628-650 (len=23)

ProductType: KEY
Total runs of length >=4: 24
First up to 3 runs: 187-191 (len=5); 200-204 (len=5); 210-216 (len=7)
Last up to 3 runs:  726-729 (len=4); 731-736 (len=6); 738-747 (len=10)

ProductType: LAP
Total runs of length >=4: 15
First up to 3 runs: 114-122 (len=9); 136-140 (len=5); 142-145 (len=4)
Last up to 3 runs:  374-390 (len=17); 395-400 (len=6); 402-426 (len=25)

ProductType: MON
Total runs of length >=4: 18
First up to 3 runs: 145-150 (len=6); 164-167 (len=4); 171-176 (len=6)
Last up to 3 runs:  536-563 (len=28); 566-613 (len=48); 615-620 (len=6)

ProductType: MOU
Total runs of length >=4: 21
First up to 3 runs: 200-203 (len=4); 222-226 (len=5); 249-256 (len=8)
Last up to 3 runs:  765-775 (len=11); 777-808 (len=32); 810-861 (len=52)

ProductType: SOF
Total runs of length >=4: 26
```

Part 4

4.1

A Type I error means that the rule flags a problem (A, B, or C) even though the process was actually in control.

ProductType <chr>	n_samples <int>	P_A_pct <dbl>	P_B_pct <dbl>	P_C_pct <dbl>
CLO	650	58.4398	100	11.66360
KEY	747	63.5437	100	16.30380
LAP	426	43.7546	100	3.66516
MON	620	56.7210	100	10.35080
MOU	861	68.7468	100	22.33620
SOF	865	68.9152	100	22.55590

P(A) Percentage - Chance per subgroup, probability that one sample from that respective product type is above the UCL(+3σ).

P(B) Percentage – Rule B runs are expected to occur often, since 68% of data lies within +/- 1σ by chance, hence the probabilities of 100%

P(C) Percentage – The probability of finding 4 consecutive samples above the +2σ line – P(A)^4

4.2

CL(μ) = 25.05 L

UCL = 25.089 L

LCL = 25.011 L

CL (μ new) = 25.028 L

σx(old) = 0.013

σx(new) = 0.017

A Type II error is only possible when the process is not in control and has deviated.

β = P (Type II error)

$H_0 : \mu = \mu_0$

$H_a : \mu \neq \mu_0$

$\beta = P(25{,}011 \leq \bar{X} \leq 25.089), \text{ when } \mu = 25.028)$

$$z_1 = \frac{25.011 - 25{,}028}{0{,}013} = -1{,}31$$

$$z_2 = \frac{25{,}089 - 25{,}028}{0{,}013} = 4.69$$

$\beta = P(-1.31 \leq z \leq 4{,}69) = P(z \leq 4.69) - P(z \leq -1{,}31)$

$= 0{,}9049$

4.3

ProductID <chr>	Category <chr>	SellingPrice <dbl>	Markup <dbl>
SOF001	Software	505.26	10.43
SOF002	Software	549.02	11.95
SOF003	Software	1083.11	21.25
SOF004	Software	1128.98	25.48
SOF005	Software	18711.72	13.51
SOF006	Software	6634.13	27.80
SOF007	Software	6478.10	17.46
SOF008	Software	627.92	26.15
SOF009	Software	835.62	26.80
SOF010	Software	364.75	24.70
SOF011	Software	505.26	10.43
SOF012	Software	549.02	11.95
SOF013	Software	1083.11	21.25
SOF014	Software	1128.98	25.48
SOF015	Software	18711.72	13.51
SOF016	Software	6634.13	27.80
SOF017	Software	6478.10	17.46
SOF018	Software	627.92	26.15
SOF019	Software	835.62	26.80
SOF020	Software	364.75	24.70
SOF021	Software	505.26	10.43
SOF022	Software	549.02	11.95
SOF023	Software	1083.11	21.25
SOF024	Software	1128.98	25.48
SOF025	Software	18711.72	13.51

CLO001	Cloud Subscription	516.15	11.01
CLO002	Cloud Subscription	1070.54	16.41
CLO003	Cloud Subscription	991.81	18.87
CLO004	Cloud Subscription	19452.72	19.80
CLO005	Cloud Subscription	15851.74	13.92
CLO006	Cloud Subscription	6396.18	26.20
CLO007	Cloud Subscription	530.51	25.56
CLO008	Cloud Subscription	693.24	29.53
CLO009	Cloud Subscription	424.79	21.36
CLO010	Cloud Subscription	454.04	18.95
CLO011	Cloud Subscription	516.15	11.01
CLO012	Cloud Subscription	1070.54	16.41
CLO013	Cloud Subscription	991.81	18.87
CLO014	Cloud Subscription	19452.72	19.80
CLO015	Cloud Subscription	15851.74	13.92

ProductID	Category	SellingPrice	Markup
<chr>	<chr>	<dbl>	<dbl>
LAP006	Laptop	5572.82	29.72
LAP007	Laptop	6711.03	29.50
LAP008	Laptop	662.16	14.11
LAP009	Laptop	375.59	22.22
LAP010	Laptop	394.30	15.81
LAP011	Laptop	493.69	16.18
LAP012	Laptop	540.41	11.34
LAP013	Laptop	728.26	27.70
LAP014	Laptop	19494.91	20.54
LAP015	Laptop	17202.28	19.11
LAP016	Laptop	5572.82	29.72
LAP017	Laptop	6711.03	29.50
LAP018	Laptop	662.16	14.11
LAP019	Laptop	375.59	22.22
LAP020	Laptop	394.30	15.81
LAP021	Laptop	493.69	16.18
LAP022	Laptop	540.41	11.34
LAP023	Laptop	728.26	27.70
LAP024	Laptop	19494.91	20.54
LAP025	Laptop	17202.28	19.11
LAP026	Laptop	5572.82	29.72
LAP027	Laptop	6711.03	29.50
LAP028	Laptop	662.16	14.11
LAP029	Laptop	375.59	22.22
LAP030	Laptop	394.30	15.81
MON001	Monitor	542.56	17.19
MON002	Monitor	396.72	23.47
MON003	Monitor	959.51	19.55
MON004	Monitor	16644.21	29.84
MON005	Monitor	18554.28	11.49
MON006	Monitor	6191.14	16.02
MON007	Monitor	5346.14	29.74
MON008	Monitor	708.18	17.72
MON009	Monitor	425.14	24.84
MON010	Monitor	373.82	17.41
MON011	Monitor	542.56	17.19
MON012	Monitor	396.72	23.47
MON013	Monitor	959.51	19.55
MON014	Monitor	16644.21	29.84
MON015	Monitor	18554.28	11.49
MON016	Monitor	6191.14	16.02
MON017	Monitor	5346.14	29.74
MON018	Monitor	708.18	17.72
MON019	Monitor	425.14	24.84
MON020	Monitor	373.82	17.41

ProductID	Category	SellingPrice	Markup
<chr>	<chr>	<dbl>	<dbl>
KEY011	Keyboard	478.93	16.99
KEY012	Keyboard	963.14	10.13
KEY013	Keyboard	1105.66	20.23
KEY014	Keyboard	18366.92	29.35
KEY015	Keyboard	16860.33	15.04
KEY016	Keyboard	6192.01	27.92
KEY017	Keyboard	607.41	20.44
KEY018	Keyboard	512.40	27.56
KEY019	Keyboard	417.40	11.88
KEY020	Keyboard	350.45	27.14
KEY021	Keyboard	478.93	16.99
KEY022	Keyboard	963.14	10.13
KEY023	Keyboard	1105.66	20.23
KEY024	Keyboard	18366.92	29.35
KEY025	Keyboard	16860.33	15.04
KEY026	Keyboard	6192.01	27.92
KEY027	Keyboard	607.41	20.44
KEY028	Keyboard	512.40	27.56
KEY029	Keyboard	417.40	11.88
KEY030	Keyboard	350.45	27.14
KEY031	Keyboard	478.93	16.99
KEY032	Keyboard	963.14	10.13
KEY033	Keyboard	1105.66	20.23
KEY034	Keyboard	18366.92	29.35
KEY035	Keyboard	16860.33	15.04

ProductID <chr>	Category <chr>	SellingPrice <dbl>	Markup <dbl>
MOU001	Mouse	511.53	25.05
MOU002	Mouse	527.56	16.79
MOU003	Mouse	1067.54	16.80
MOU004	Mouse	1092.07	23.14
MOU005	Mouse	19725.18	11.70
MOU006	Mouse	6806.08	23.27
MOU007	Mouse	6777.62	11.05
MOU008	Mouse	516.41	22.83
MOU009	Mouse	752.75	29.11
MOU010	Mouse	366.70	20.64
MOU011	Mouse	511.53	25.05
MOU012	Mouse	527.56	16.79
MOU013	Mouse	1067.54	16.80
MOU014	Mouse	1092.07	23.14
MOU015	Mouse	19725.18	11.70
MOU016	Mouse	6806.08	23.27
MOU017	Mouse	6777.62	11.05
MOU018	Mouse	516.41	22.83
MOU019	Mouse	752.75	29.11
MOU020	Mouse	366.70	20.64
MOU021	Mouse	511.53	25.05
MOU022	Mouse	527.56	16.79
MOU023	Mouse	1067.54	16.80
MOU024	Mouse	1092.07	23.14
MOU025	Mouse	19725.18	11.70

As seen above these are the updated prices using the products_data.csv, the prices were repeated every 10 values for each respective product type, now below you'll find the newly calculated total sales value for 2023 per product type as well as the new week 1 analysis. Here is the total revenue for 2023 using the newly **updated** selling prices.

ProductType <chr>	Total_Revenue_2023 <dbl>
KEY	524926478
MON	492651283
MOU	480936410
SOF	479027793
CLO	441851801
LAP	344806731

Then here is the revenue for 2023 using the **old** product data selling prices.

ProductType <chr>	Total_Revenue_2023 <dbl>
LAP	1119099823
MON	543393489
CLO	92703320
KEY	74840190
SOF	61821700
MOU	53714561

From the above table we can see that the revenue for all product types increased from the old selling prices except for LAP and MON.

Part 5

Shop 1

The table below shows the reliability percentage (proportion of clients with a service time of less than 120s). The optimal number of baristas as well the optimum profit for the company regarding Shop 1.

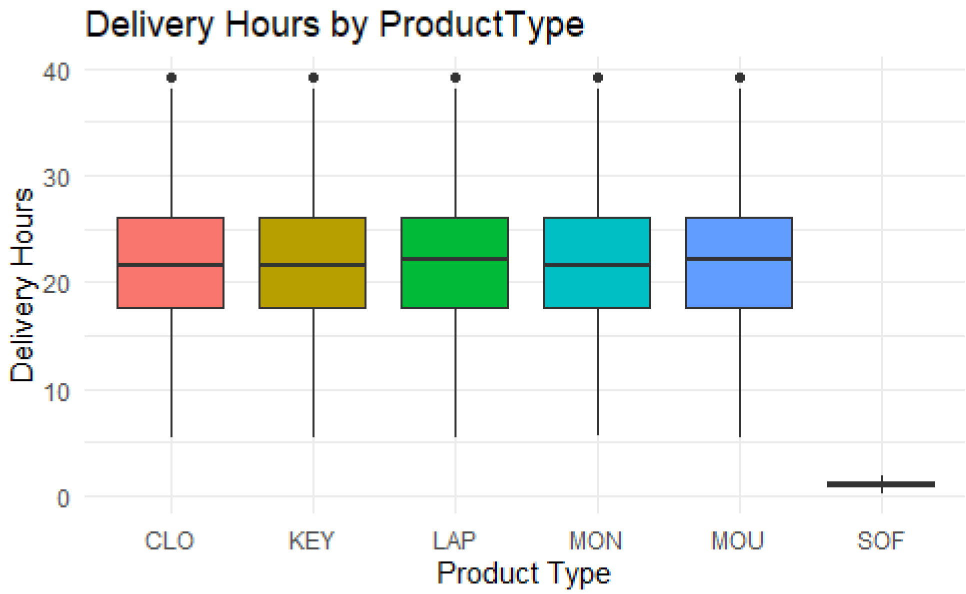
	c	mean_service_empirical	n_obs_for_c	rho	P_wait_le_t	revenue	personnel_cost	profit
	<int>	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
2	2	100.17098	3556	0.9529203	0.1692490	16438.36	2000	14438.36
3	3	66.61174	12126	0.4224489	0.9929183	16438.36	3000	13438.36
4	4	49.98038	29305	0.2377301	0.9999886	16438.36	4000	12438.36
5	5	39.96183	56701	0.1520618	1.0000000	16438.36	5000	11438.36
6	6	33.35565	97895	0.1057701	1.0000000	16438.36	6000	10438.36

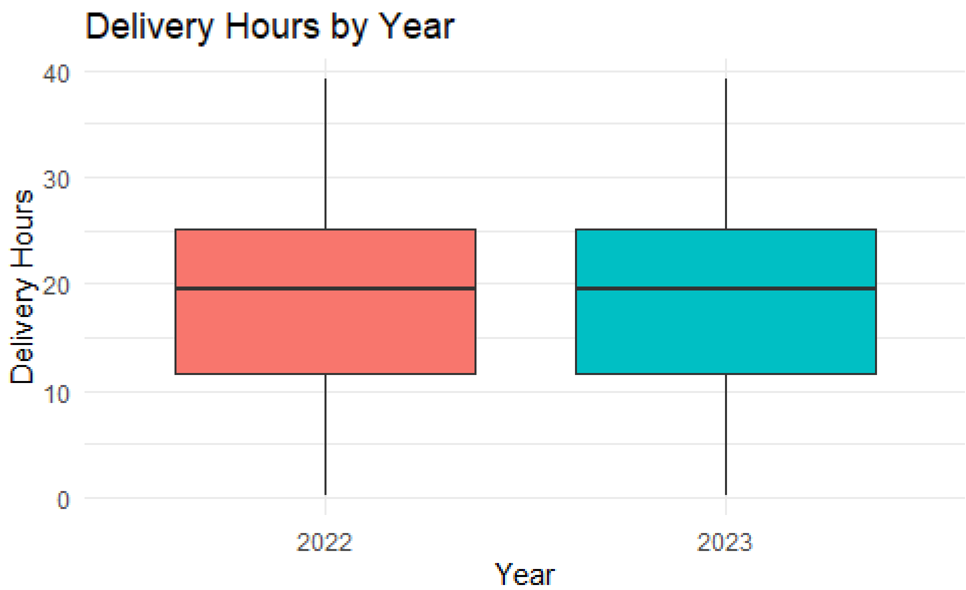
The reliability percentage is 99.29%, the optimal number of baristas is 3 baristas. Bringing in a profit of R13438.36.

Part 6

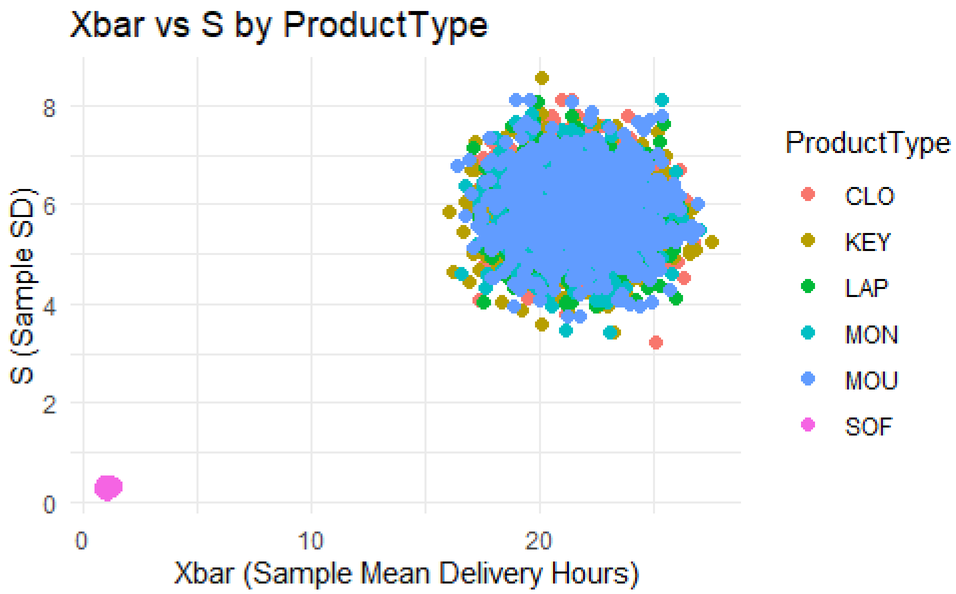
6.2

The dataset “sales2026and2027” contains delivery times (deliveryHours) for various product types. With how ANOVA & MANOVA work, because there was only one month of data available analysis was performed across products and across years (2022 and 2023). My dependent variables are as follows. ANOVA – deliverHours, MANOVA – Xbar (sample mean) and S (sample standard deviation) per product. Based on the ANOVA analysis, delivery hours differ significantly across product types.





Boxplots of delivery hours by product as shown above confirm this, showing that some products consistently have higher mean delivery times. This difference isn’t as clearly seen when the 2022 and 2023 are compared to each other. The MANOVA, using sample mean (Xbar) and sample standard deviation (S), also shows a difference across products however the difference is not as clearly seen as in the boxplots, however product type SOF is seen as an outlier. Since only month of data is available, month-based comparisons were not possible, but the results indicate that product type is a significant factor affecting delivery performance. These findings suggest that certain products may require closer process monitoring to improve consistency and reduce delays.



Part 7

7.1

The 397-day record indicates that days with at least 15 employees on duty are indicative of reliable service. The frequency distribution shows that 15 employees were on duty for 96 days. The frequency distribution showed that there were 366 trustworthy days out of 397, with 96 days having 15 workers and 270 days having 16 workers. This amounts to about 92% of all days. We can anticipate dependable service on roughly 0.92 x 365 = 337 days a year if this pattern holds. Consequently, clients can anticipate reliable service on approximately 337 days per year, while service may be less reliable on the remaining 28 days.

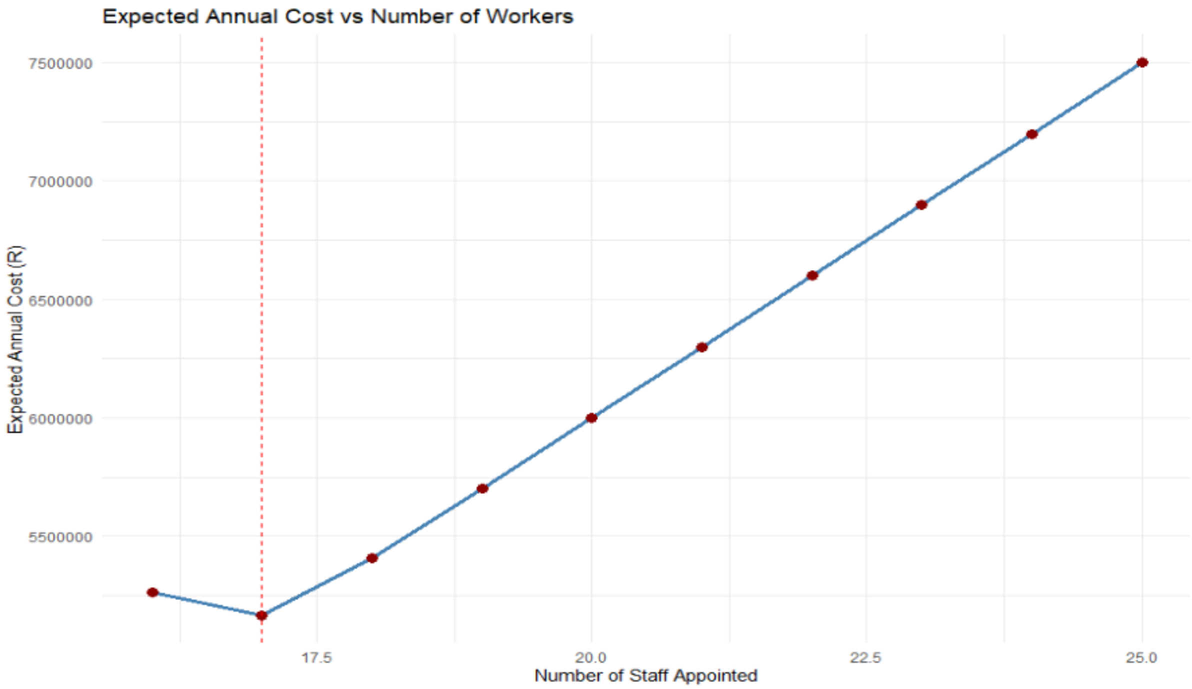
7.2

A binomial distribution was used to simulate the number of employees present each day, and the projected daily attendance probability was p = 16.95 16 ≈ 0.934.
Less than 15 employees cause issues, which leads to a daily loss of R20,000 on average. The cost of hiring each extra employee is R25,000 per month, or R300,000 annually.

```
[1] 15.58438
[1] 0.9740239

Optimal number of staff = 17 with expected yearly cost of R 5166220
```

Workers <int>	Expected_Annual_Cost <dbl>
16	5264506
17	5166220
18	5407593
19	5700740
20	6000063
21	6300005
22	6600000
23	6900000
24	7200000
25	7500000



The methodology used the projected lost-sales cost and annual pay to get the expected annual cost for different workforce levels. About 17 workers was the lowest predicted total cost.

Therefore, the company should hire roughly 17–18 employees in order to maximize profit and provide dependable service. This level strikes a compromise between the risk of service interruption from understaffing and staff costs.

Conclusion

In conclusion, this report shows an in-depth analysis of manufacturing and service process, using statistical process control and process capability analysis. This was done using descriptive statistics, SPC tools, and capability indices, we have further investigated process performance and reliability, identifying specific areas for improvement. The results provide important insights into process instability, particularly through the identification of out-of-bound samples and errors that indicate potential inefficiencies.

The process capability analysis, the interpretation of Upper Specification Limits in particular, have led to the classification of products as capable or not, offering targeted recommendations for optimizing production. Additionally, our analysis of the loss function highlights how variability impacts both operational costs and product quality.

By using statistical control tools, we have been able to identify as well as address key process issues, leading to more stable operations. Evaluating delivery times in regards to profitability provides valuable insight to improving to improving service delivery and boosting customer satisfaction. Conclusively, it can be said that the data-driven approach highlights current challenges and presents a clear path for continuous improvement, optimizing processes for better efficiency, cost-effectiveness, and customer satisfaction. Future research should refine these strategies and explore additional process variables to further enhance performance and reliability.

References

MANOVA test in R: Multivariate analysis of variance (no date) *STHDA*. Available at: https://www.sthda.com/english/wiki/manova-test-in-r-multivariate-analysis-of-variance (Accessed: 23 October 2025).

What is Statistical Process Control? SPC Quality Tools | ASQ (2025) *Asq.org*. Available at: https://asq.org/quality-resources/statistical-process-control?srsltid=AfmBOoq8lYDw-sq8R1CmPYzdR7L9rNkBSb_HZt02qPG0Rt9yVGSQjhlP (Accessed: 22 October 2025).

SERVICES – Premium Plus Health. Available at: https://premiumplushealth.com/services/