QUALITY ASSURANCE
ECSA REPORT

Troy Dyer
26887193
Stellenbosch University

23/10/25

Abstract

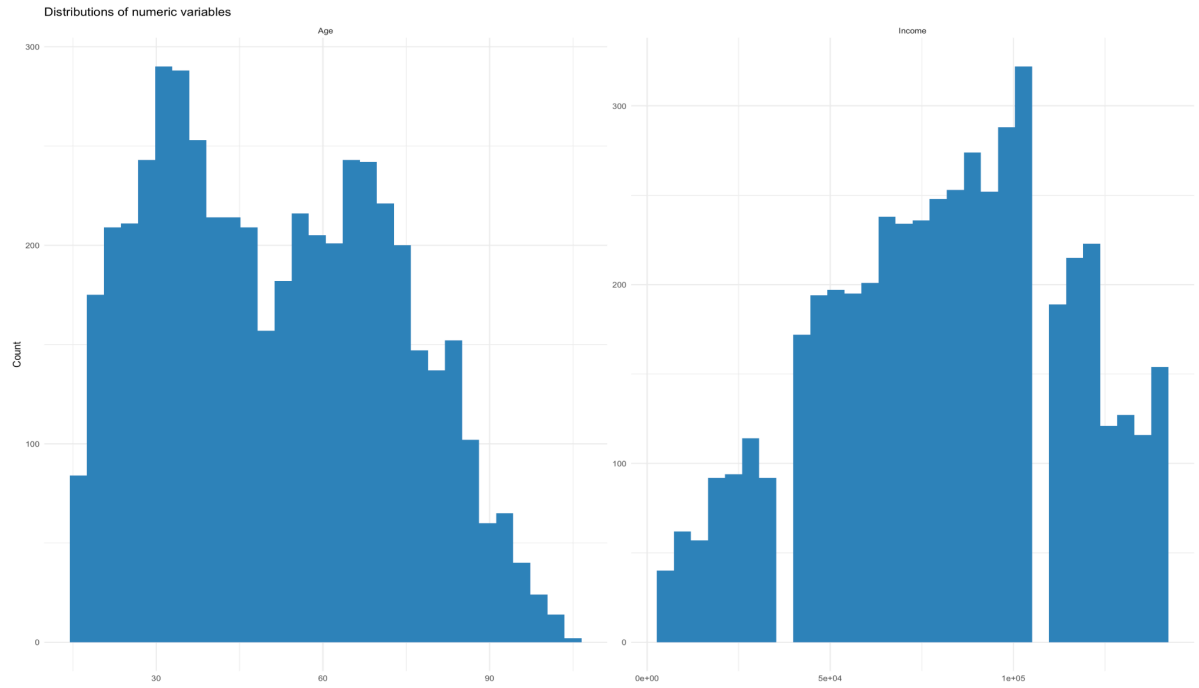
This is a data analysis report that follows the ECSA GA4 guide. The task involved applying a variety of different data analysis techniques and quality control measures. Statistical tools like SOC charts, process capability indices and ANOVA were applied to examine different measures in the data sets. Results highlight different areas for improvement and supported data to suggest some improvements.

TABLE OF CONTENTs

QUESTION 1
Data analysis project
==========

Customer data

==========
Figure 1



Age Distribution (left)

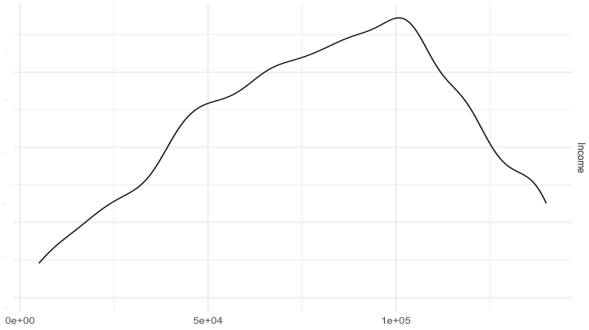
The age distribution shows multiple peaks. This shows that there is a larger spread over all the different ages and not concentrated only on a specific rage of ages. There is a large cluster of ages from 20s - 40s and another large cluster of ages from 55s - 70s. The frequency gradually decreases for customers older than 70. There is still a large number of ages above 70 which shows there is a diverse age of the customers.

Income Distribution (right)

The income data is skewed to the right. With most of the customers falling into the lower to mid income ranges. There is a very large cluster close to the 50000 mark which steadily decreases as the income range becomes higher. There are still a number of customers that have a large income but the numbers are a lot fewer than that of the lower to mid income range.

Plots

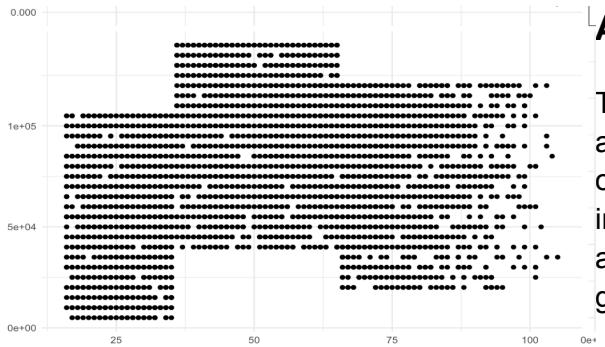
Figure 2



Income Distribution

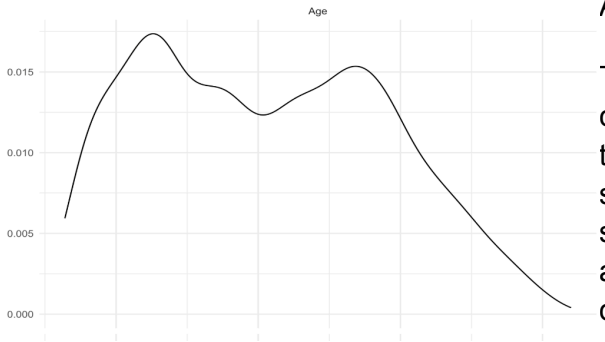
The income plot again shows the right-skewed trend, meaning most customers earn in the lower-to-middle income ranges, while only a smaller group earns very high incomes. The highest concentration is below the 100,000 range.

Figure 3



Age(y) vs. Income Scatter(x)

The scatter plot shows the relationship between age and income. There is a slight upward trend, meaning older customers tend to have somewhat higher incomes. However, there is a lot of spread at every age, so income levels vary widely within each age group.



Age Distribution

The age distribution shows that again there are two clusters that show a large number of customers in the ages of 20 - 40 and again at 55 - 70. And then slowly decrease from 70 on wards, while still showing a large number of customers at those older ages which shows the wide spread of ages in the data set

Figure 4

Correlation r = 0.158

The correlation between age and income is positive but weak. This means that while income tends to increase a little with age, age alone does not strongly predict how much a customer earns.

Figure 5

5 number summary

	skim_variable <chr>	n_missing <int>	complete_rate <dbl>	mean <dbl>	sd <dbl>	p0 <dbl>	p25 <dbl>	p50 <dbl>	p75 <dbl>
1	Age	0	1	51.5538	21.2161	16	33	51	68
2	Income	0	1	80797.0000	33150.1067	5000	55000	85000	105000

Figure 6

Interpretation age :

- The average is 51.5 and has a standard dev of 21.2 which shows a medium spread of the data
- The youngest customer is 16n and the oldest is 105 which is a very large range
- The middle 50% of the population (between Q1 and Q3) are aged 33–68, showing a wide age range and diverse demographic representation.
- The data appears to be slightly right-skewed, since the mean (51.6) is slightly higher than the median (51), suggesting some older participants pull the average up.
-

Interpretation income :

- The average income is R80 797, with a large standard deviation of R33 150 this shows a high income across individuals.
- The income distribution is positively skewed, as the mean is below the 3rd quartile but above the median, some participants likely earn significantly higher incomes.
- The central 50% earn between R55 000 and R105 000, which represents a relatively broad middle-income range.
- The minimum income of R5 000 shows that some individuals earn very little compared to the group average, possibly outliers or entry-level earners.

	Skew	kurtosis	explain
Age	0.2	-0.99	Skewness value shows it is slightly positive. The kurtosis value shows the data is platykurtic which shows the data is flatter and less peaked than a normal distribution
Income	-0.21	-0.75	The data for income is slightly negatively skewed. The income data is also playkurtic .

Table 1

==========

Products data

==========

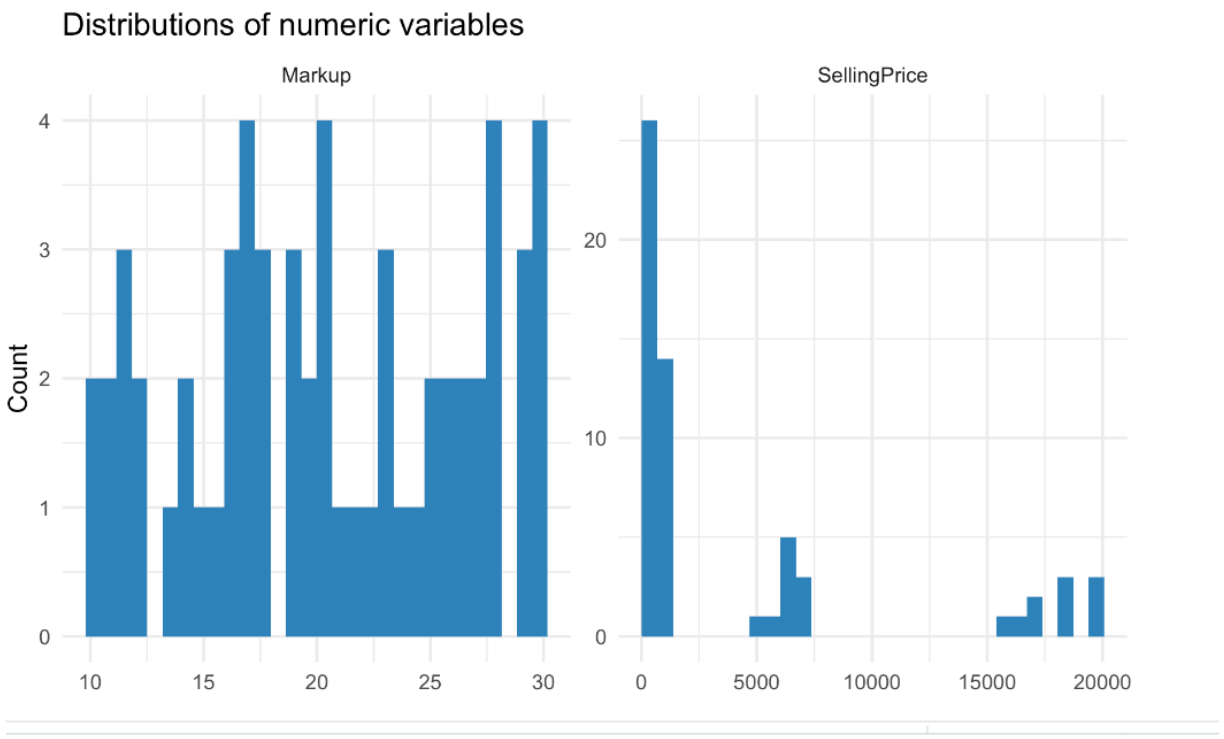


Figure 7

Markup Distribution (left)

The Markup values are mostly between 10 and 30. The data is fairly spread out with several small peaks which shows that products are priced with different markup strategies rather than being concentrated at a single level. Many items cluster around 15–20 and again near 28–30 which shows common pricing practices at these levels.

Selling Price Distribution (right)

The SellingPrice distribution is highly skewed to the right. A large number of products are priced in the low range, from zero to a few thousand. However, there are a few products with very high selling prices around 15,000–20,000 as can be seen in the histogram as there are a few spikes at these upper values. This indicates the dataset includes both low-cost items like everyday products and a small set of expensive products.

Plots

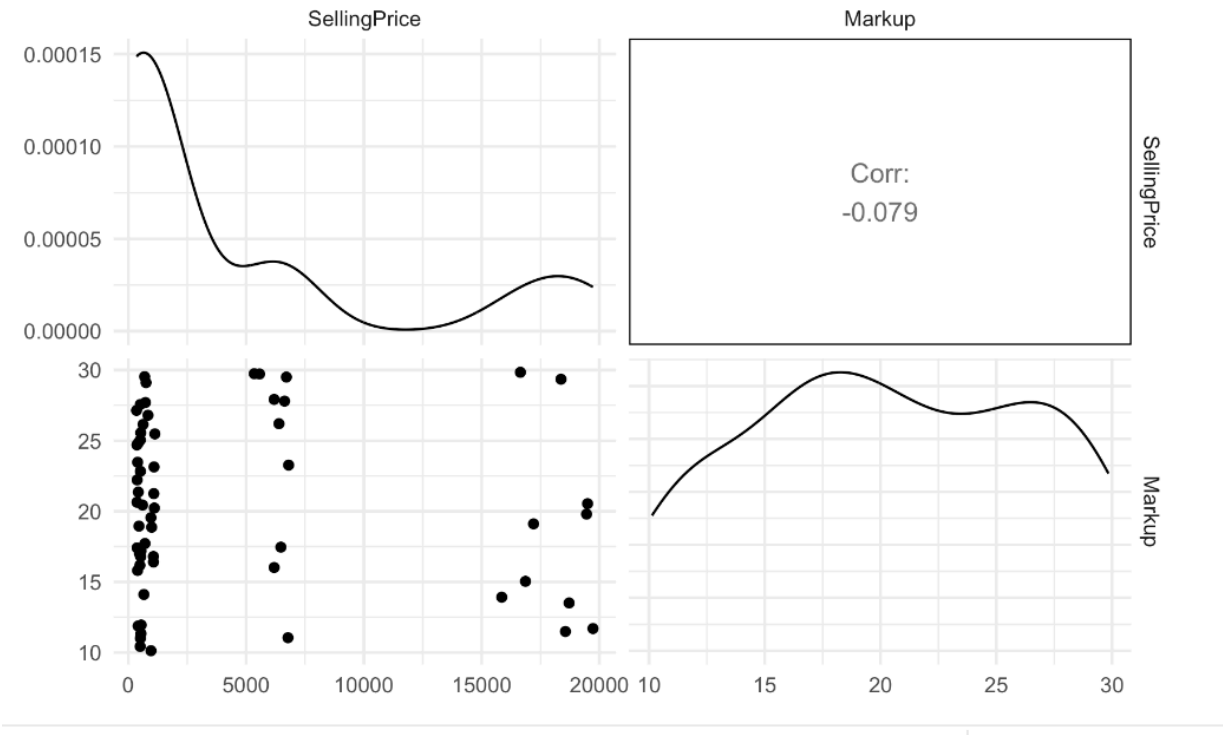


Figure 8

Selling price (top left)
The distribution shows how the data is skewed to the right as the data start very high and slowly decrease as the price increases. There are a few spikes as the price gets bigger which shows there are many everyday sold items and then a few more expensive items.

Markup (Bottom Right)

The markup distribution is fairly balanced it shows moderate variation with multiple small peaks. This shows that markups are not concentrated around a single value, but rather spread across a range. This shows that there is possible some different pricing strategies across the different products.

Correlation (top right)
The correlation is -0.079, which is very close to zero which indicates that there is almost no linear relationship between markup and selling price

Scatter plot (bottom left)
The scatter plot is positively skewed; this shows that there is a lot of items sold at lower prices, as there is a large cluster at lower selling prices. There are some smaller clusters at higher selling prices which shows there are some more expensive products being sold.

5 number summary

	skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
	<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	SellingPrice	0	1	4493.59283	6503.770150	350.45	512.1825	794.185	6416.6600	19725.18
2	Markup	0	1	20.46167	6.072598	10.13	16.1400	20.335	25.7075	29.84

Interpretation selling price:

- The selling price average is at R4493.56 and has a standard deviation of 6503.77 which shows quite a large spread.
- With a minimum selling price value of 350.45 and a maximum value of 19725.18 shows a wide range of item values
- Again the medium - p50 of R794 is much lower than the mean so this too tells us that it is positively skewed.
- With p25 and p75 having 50% of the prices between the values shows the different variation of prices

Interpretation Markup:

- The mean markup price being 20.46 and a standard deviation of 6.07, shows a moderate variation in markup between products.
- With the minimum being 10.13 and maximum being 29.84 shows that the values value into quite a consistency range
- With the medium being so close to the mean, shows that the data is close to being symmetrical.

	skewed	kutorsis	explain
Selling price	1.43	0.43	The skewed value shows that the selling price distribution is positively skewed. The kurtosis value suggests a slightly leptokurtic distribution, which means it is more peaked than normal distribution.
Mark up	-0.04	-1.24	The skewed value is so close to zero, that the data is basically symmetrical. The kutorsis value is platykurtic- it is flatter than a normal distribution.

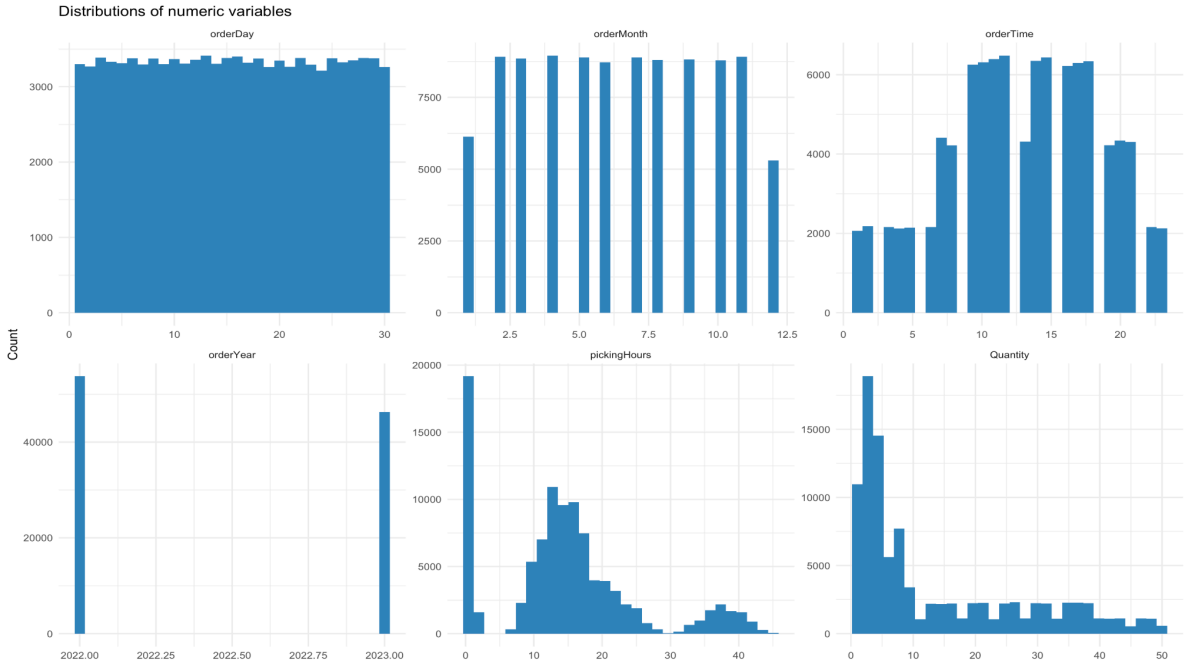
Table 2

==========

Sales 2022 and 2023

==========

Figure 9



order day (top left)

The Orderday spread is very uniform across all days of the month. There are no specific peaks in the distribution, this shows that customers make orders throughout the whole month.

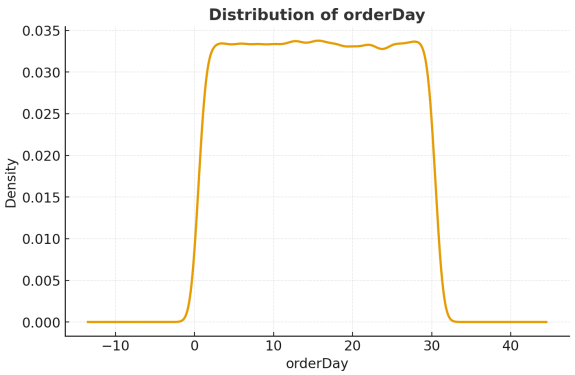


Figure 10

Ordermonths (top middle)

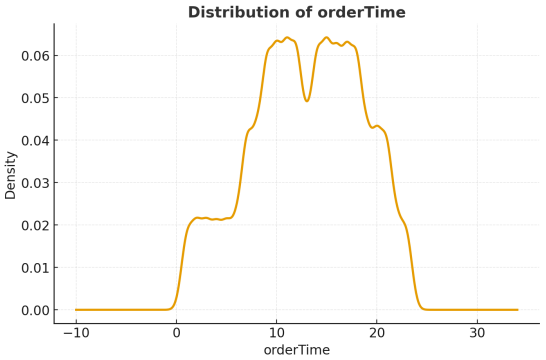
The months where customers order are pretty much consistent throughout the year with slight dips in January and December. This shows that sales stay stable throughout the year but during holiday seasons there may be less sales

Activity.

Figure 11

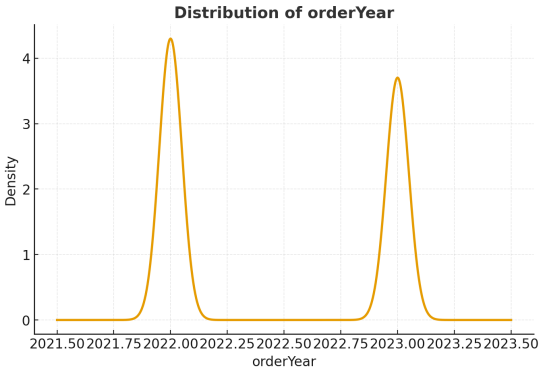
Figure 12

Ordertime (top right)
The ordertime histogram has a bimodal pattern, which shows two distinct peaks. Orders most frequent from 10:00 - 16:00, with small peaks around mid morning and late afternoon.
This shows that most orders are made during working hours.



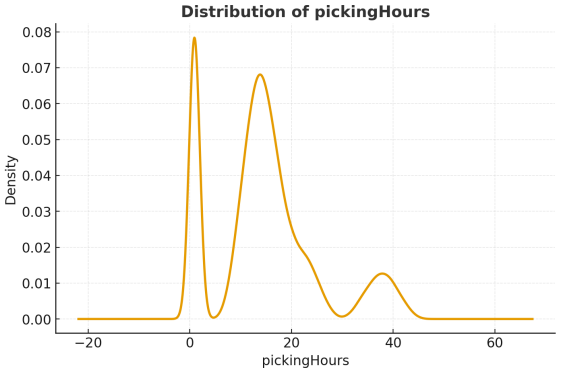
order year (bottom left)
There are only two years that the data covers and that is 2022 and 2023, with most orders being in 2022.

Figure 13



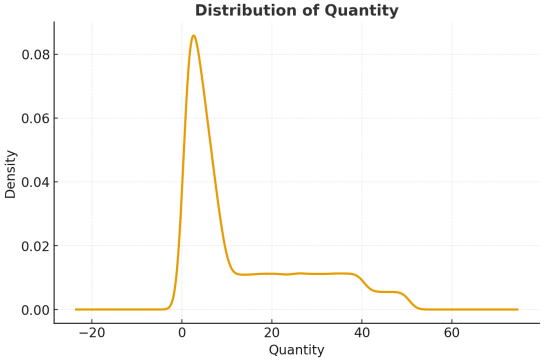
Picking hours (bottom middle)
The pickshours histogram is skewed to the right, with large clusters around the 5-15 hour mark and a smaller cluster around the 35-45 hour mark. This shows that the majority of the orders require short to moderate picking times, but a few orders take a lot longer. The small peak at the end may be due to bulk orders.

Figure 14



Quantity (bottom right)
The quantity histogram is positively skewed, with most orders containing smaller quantities, below 10 units. There are many small, frequent purchases with the vocational bulk order.

Figure 15



5 number summary

	skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
	<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Quantity	0	1	13.50347	13.7601316	1.0000000	3.000000	6.000	23.00000	50.0000
2	orderTime	0	1	12.93230	5.4951268	1.0000000	9.000000	13.000	17.00000	23.0000
3	orderDay	0	1	15.49683	8.6465055	1.0000000	8.000000	15.000	23.00000	30.0000
4	orderMonth	0	1	6.44813	3.2834460	1.0000000	4.000000	6.000	9.00000	12.0000
5	orderYear	0	1	2022.46273	0.4986115	2022.0000000	2022.000000	2022.000	2023.00000	2023.0000
6	pickingHours	0	1	14.69547	10.3873345	0.4258889	9.390833	14.055	18.72167	45.0575
7	deliveryHours	0	1	17.47646	9.9999440	0.2772000	11.546000	19.546	25.04400	38.0460

Figure 16

Interpretation Quantity:

- With the mean of 13.5 and a standard deviation of 13.76, shows that there is a large spread of values.
- The range being 1 - 50 shows a there is a big spread in the variation of quantity
- With most orders being below 10, and a few that are larger bring the mean up.

Interpretation orderTime:

- With mean being 12.93 and the standard dev of 5.5 mean that the spread of values is not too high.
- The values lie between 1-23
- The p50 is 13 which is very close to the mean so the distribution is roughly symmetric with two small peaks at mid morning and early afternoon which shows typical work day activity

Interpretation orderDay:

- The mean of order day is 15.5 and has a standard div of 15.5.
- This is a nearly uniform distribution and almost symmetric as the mean is very close to the p50 value of 15

Interpretation orderMonth

- The mean is 6.45 which is very close to the medium of 6 so this shows that the distribution is basically symmetrical.
- Most of the data falls in the IQR so this shows that there is a uniform distribution but there is a slight dip in Jan and Dec.

Interpretation pickingHours:

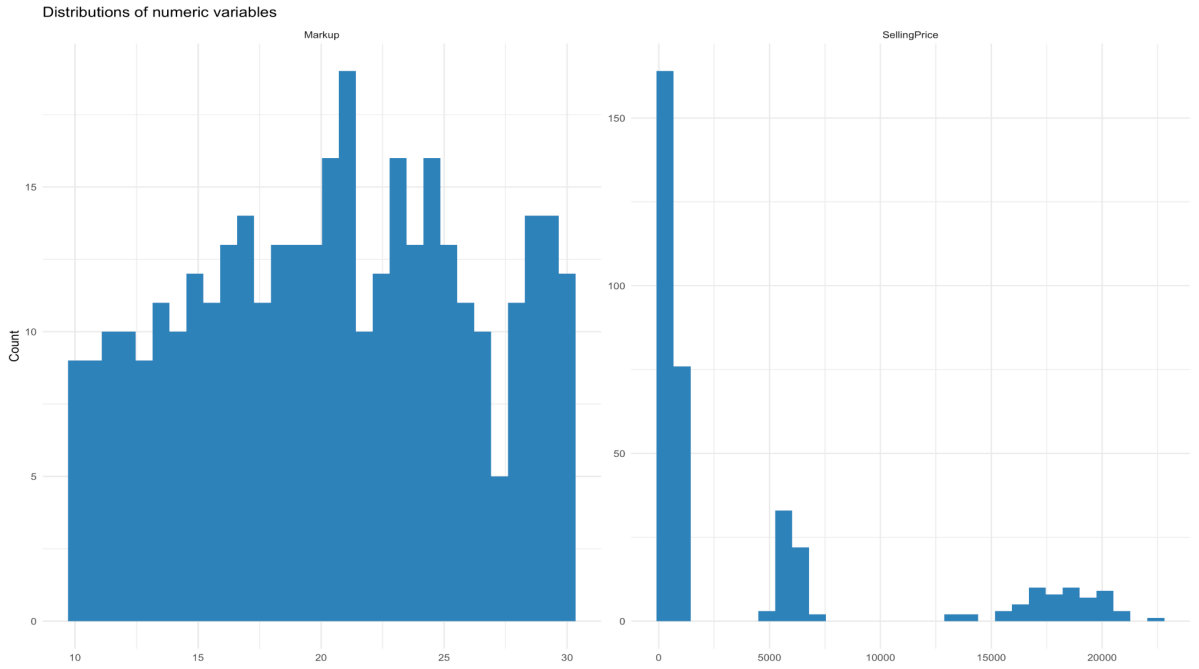
- With a mean of 14.69 and a standard dev of 10.38 there is large spread of variables.
- The data is skewed to the right, which means most orders require shorter picking times.

==========

Head Office products data

==========

Figure 17



Markup (left) - Selling price (right)

The mark up distribution almost has a bell shape curve, centered around 20-25 this shows a near normal distribution. This can show that there is a consistent profit across products. Where selling price on the other hand is highly skewed to the right with three distinct peaks, this indicates that there are multiple tired products. The majority of them fall under the lower price range while there are a few more expensive products. This confirms that the data has both high and low costing products.

Plots

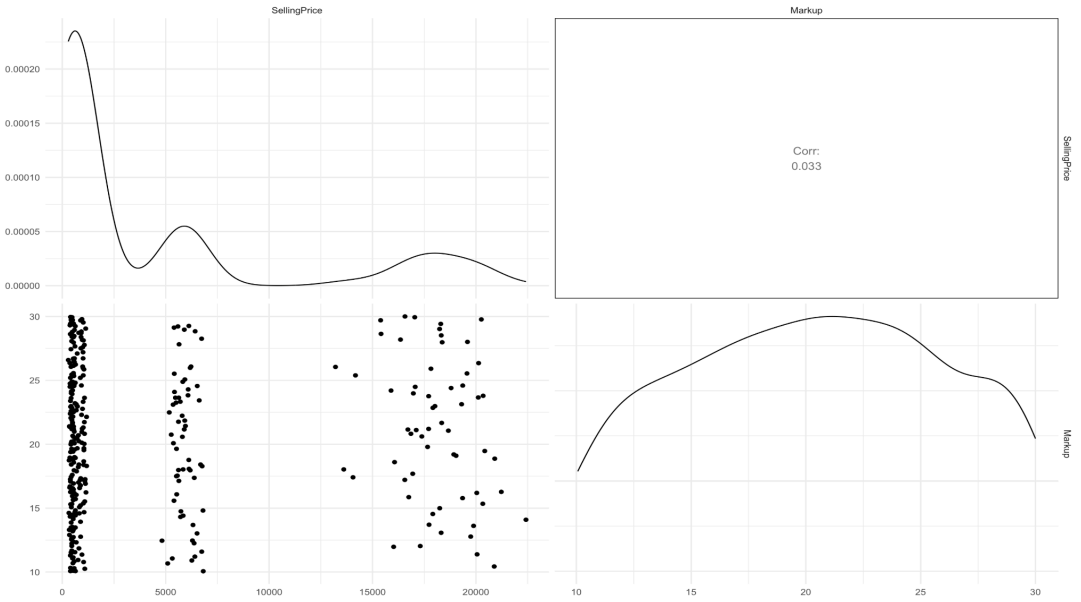


figure18

Selling price (top)left

- The curve shows a strong skewed right with most values concentrated at the lower end of the price range
- There are some smaller peaks at mid and high- price ranges, suggesting that the data contains three major products
- The long tail indicates that there are fewer expensive products

Markup distribution (bottom right)

- The density curve is roughly bell shaped centered around 20 - 25
- There is no strong skewness and the distribution is fairly symmetrical
- Mark ups are stable and consistent across the products, reflecting standardized profit margins rather than category specific profit margins

Scatter plot (bottom left)

- Each dot represents a product's selling price and its corresponding markup
- There's no visible clusters or pattern
- This shows that the selling price and markup are not strongly related
- The absence of a trend indicates that the markup percentage remains relatively independent of selling price. The company likely applies similar markup rates across both low- and high-priced items.

5 number summary

	skim_variable <chr>	n_missing <int>	complete_rate <dbl>	mean <dbl>	sd <dbl>	p0 <dbl>	p25 <dbl>	p50 <dbl>	p75 <dbl>	p100 <dbl>
1	SellingPrice	0	1	4410.9619	6463.822788	290.52	495.9375	797.215	5843.332	22420.14
2	Markup	0	1	20.3855	5.665949	10.06	15.8400	20.580	24.845	30.00

Figure 19

Interpretation selling price

- The mean of 4410 is significantly larger that the medium of 797.215 this shows that the distribution is heavily skewed to the right

- The large standard deviation confirms substantial price variability

- The results suggest three major price zones

Interpretation markup

- The mean of 20.39 and median of 20.58 are so close that this suggest abd symmetrical and close to normal distribution

- The range is 10-30 this is narrow compared to selling price, this shows the consistent profit margins

- The std dev of 5.7 means most markups fall with in the $\pm$6% of the mean

Question 3

3.1

Figure 20

ProductID <chr>	n <dbl>	m_used <int>	xbarbar <dbl>	sbar <dbl>	UCL_x <dbl>	LCL_x <dbl>	CL_x <dbl>	UCL_s <dbl>	LCL_s <dbl>	CL_s <dbl>	U1_x <dbl>	U2_x <dbl>	L1_x <dbl>	L2_x <dbl>	U1_s <dbl>	U2_s <dbl>	L1_s <dbl>	L2_s <dbl>
CLO011	24	30	21.178967	6.1135893	24.963663	17.3942707	21.178967	8.8321194	3.3950591	6.1135893	22.440532	23.702097	19.9174013	18.6558360	7.0197660	7.9259427	5.2074125	4.3012358
CLO012	24	30	21.645000	5.9582356	25.333522	17.9564777	21.645000	8.6076846	3.3087866	5.9582356	22.874507	24.104015	20.4154926	19.1859851	6.8413853	7.7245350	5.0750859	4.1919363
CLO013	24	30	21.386789	5.7521782	24.947749	17.8258291	21.386789	8.3099996	3.1943567	5.7521782	22.573775	23.760762	20.1998023	19.0128157	6.6047853	7.4573925	4.8995710	4.0469639
CLO014	24	30	21.371389	5.8815105	25.012414	17.7303642	21.371389	8.4968422	3.2661788	5.8815105	22.585064	23.798739	20.1577140	18.9440391	6.7532877	7.6250649	5.0097333	4.1379561
CLO015	24	30	21.151067	5.7659893	24.720576	17.5815569	21.151067	8.3299521	3.2020264	5.7659893	22.340903	23.530740	19.9612301	18.7713935	6.6206436	7.4752979	4.9113350	4.0566807
CLO016	24	30	21.430928	5.8873666	25.075578	17.7862778	21.430928	8.5053024	3.2694309	5.8873666	22.645811	23.860694	20.2160445	19.0011611	6.7600119	7.6326571	5.0147214	4.1420762
CLO017	24	30	21.167856	5.8153978	24.767952	17.5677588	21.167856	8.4013311	3.2294645	5.8153978	22.367888	23.567920	19.9678233	18.7677911	6.6773756	7.5393533	4.9534200	4.0914422
CLO018	24	30	21.226128	5.9980509	24.939298	17.5129573	21.226128	8.6652046	3.3308972	5.9980509	22.463851	23.701575	19.9884043	18.7506808	6.8871021	7.7761534	5.1089997	4.2199484
CLO019	24	30	21.672328	5.9531444	25.357698	17.9869572	21.672328	8.6003296	3.3059593	5.9531444	22.900785	24.129241	20.4438709	19.2154141	6.8355395	7.7179345	5.0707494	4.1883543
CLO020	24	30	20.944183	5.8847739	24.587228	17.3011384	20.944183	8.5015567	3.2679911	5.8847739	22.158532	23.372880	19.7298350	18.5154867	6.7570348	7.6292958	5.0125130	4.1402520
KEY041	24	30	20.880356	5.9255464	24.548641	17.2120699	20.880356	8.5604595	3.2906333	5.9255464	22.103117	23.325879	19.6575937	18.4348318	6.8038508	7.6821551	5.0472420	4.1689377
KEY042	24	30	21.062933	5.8578286	24.689297	17.4365693	21.062933	8.4626296	3.2530276	5.8578286	22.271721	23.480509	19.8541453	18.6453573	6.7260956	7.5943626	4.9895616	4.1212946
KEY043	24	30	21.280478	5.7435019	24.836066	17.7248891	21.280478	8.2974653	3.1895385	5.7435019	22.465674	23.650870	20.0952816	18.9100853	6.5948230	7.4461441	4.8921808	4.0408596
KEY044	24	30	21.304478	5.7389422	24.857244	17.7517119	21.304478	8.2908780	3.1870064	5.7389422	22.488733	23.672988	20.1202225	18.9359672	6.5895874	7.4402327	4.8882969	4.0376516
KEY045	24	30	21.424983	5.9988249	25.138633	17.7113337	21.424983	8.6663228	3.3313270	5.9988249	22.662867	23.900750	20.1871001	18.9492169	6.8879908	7.7771568	5.1096589	4.2204930
KEY046	24	30	21.062117	5.8420653	24.678722	17.4455111	21.062117	8.4398569	3.2442737	5.8420653	22.267652	23.473187	19.8565815	18.6510463	6.7079958	7.5739264	4.9761348	4.1102043
KEY047	24	30	21.011667	5.9388710	24.688201	17.3351323	21.011667	8.5797092	3.2980328	5.9388710	22.237178	23.462690	19.7861552	18.5606437	6.8191504	7.6994298	5.0585916	4.1783122
KEY048	24	30	21.288567	5.7997175	24.878956	17.6981770	21.288567	8.3786783	3.2207567	5.7997175	22.485363	23.682160	20.0917701	18.8949736	6.6593711	7.5190247	4.9400639	4.0804103
KEY049	24	30	21.366917	5.9896462	25.074884	17.6589492	21.366917	8.6530625	3.3262298	5.9896462	22.602906	23.838895	20.1309275	18.8949384	6.8774516	7.7652571	5.1018407	4.2140353
KEY050	24	30	21.241528	5.8283023	24.849613	17.6334424	21.241528	8.4199738	3.2366307	5.8283023	22.444223	23.646918	20.0388326	18.8361375	6.6921928	7.5560833	4.9644117	4.1005212
LAP021	24	30	21.070208	5.9108480	24.729395	17.4110219	21.070208	8.5392252	3.2824708	5.9108480	22.289937	23.509666	19.8504795	18.6307507	6.7869737	7.6630995	5.0347223	4.1585966
LAP022	24	30	21.313636	5.6753759	24.827050	17.8002217	21.313636	8.1990457	3.1517061	5.6753759	22.484774	23.655912	20.1424980	18.9713599	6.5165992	7.3578224	4.8341527	3.9929294
LAP023	24	30	21.525386	5.7673297	25.095726	17.9550466	21.525386	8.3318886	3.2027708	5.7673297	22.715499	23.905612	20.3352729	19.1451597	6.6221827	7.4770356	4.9124767	4.0576238
LAP024	24	30	21.567794	5.9149088	25.229495	17.9060941	21.567794	8.5450917	3.2847259	5.9149088	22.788361	24.008928	20.3472277	19.1266609	6.7916365	7.6683641	5.0381812	4.1614536
LAP025	24	30	21.468253	5.8575609	25.094451	17.8420544	21.468253	8.4622430	3.2528789	5.8575609	22.676986	23.885718	20.2595200	19.0507872	6.7257883	7.5940156	4.9893336	4.1211063
LAP026	24	30	21.256561	5.9316637	24.928634	17.5844885	21.256561	8.5692970	3.2940304	5.9316637	22.480585	23.704610	20.0325369	18.8085127	6.8108748	7.6900859	5.0524526	4.1732415
LAP027	24	30	21.300378	5.8652953	24.931364	17.6693914	21.300378	8.4734165	3.2571740	5.8652953	22.510707	23.721035	20.0900490	18.8797202	6.7346690	7.6040427	4.9959215	4.1265478
LAP028	24	30	21.194350	6.0403437	24.933702	17.4549976	21.194350	8.7263037	3.3543836	6.0403437	22.440801	23.687252	19.9478992	18.7014484	6.9356637	7.8309837	5.1450237	4.2497037
LAP029	24	30	21.148981	5.8144563	24.748494	17.5494666	21.148981	8.3999710	3.2289417	5.8144563	22.348819	23.548656	19.9491426	18.7493046	6.6762946	7.5381328	4.9526181	4.0907799
LAP030	24	30	21.533692	5.9571780	25.221559	17.8458241	21.533692	8.6061567	3.3081993	5.9571780	22.762981	23.992270	20.3044025	19.0751133	6.8401709	7.7231638	5.0741851	4.1911922
MON031	24	30	21.812483	5.8646853	25.443092	18.1818745	21.812483	8.4725354	3.2568353	5.8646853	23.022686	24.232889	20.6022804	19.3920775	6.7339687	7.6032520	4.9954020	4.1261187
MON032	24	30	21.303783	5.7368316	24.855243	17.7523240	21.303783	8.2878289	3.1858343	5.7368316	22.487603	23.671423	20.1199636	18.9361438	6.5871640	7.4374964	4.8864992	4.0361667
MON033	24	30	21.481233	6.1232813	25.271929	17.6905373	21.481233	8.8461213	3.4004414	6.1232813	22.744799	24.008364	20.2176680	18.9541027	7.0308946	7.9385079	5.2156680	4.3080547
MON034	24	30	21.457439	5.8479445	25.077684	17.8371937	21.457439	8.4483504	3.2475387	5.8479445	22.664187	23.870936	20.2506905	19.0439421	6.7147465	7.5815485	4.9811426	4.1143406
MON035	24	30	21.433194	5.6088533	24.905427	17.9609617	21.433194	8.1029425	3.1147641	5.6088533	22.590605	23.748016	20.2757835	19.1183726	6.4402164	7.2715794	4.7774903	3.9461272
MON036	24	30	21.201883	5.7267971	24.747131	17.6566360	21.201883	8.2733323	3.1802618	5.7267971	22.383632	23.565382	20.0201342	18.8383851	6.5756421	7.4244872	4.8779520	4.0291069
MON037	24	30	21.941781	5.8477796	25.561924	18.3216375	21.941781	8.4481121	3.2474471	5.8477796	23.148495	24.355209	20.7350662	19.5283518	6.7145571	7.5813346	4.9810021	4.1142246
MON038	24	30	21.878517	5.8745618	25.515240	18.2417937	21.878517	8.4868035	3.2623200	5.8745618	23.090758	24.302999	20.6662757	19.4540347	6.7453090	7.6160563	5.0038145	4.1330673

Overview

X-charts and S-chart were created for the different delivery times of my data. To ensure chronological order the data was ordered by year first then month, day and order times. The first 30 samples have been used to determine centerlines, 1-σ, 2-σ and 3-σ control limits for the charts.

Key for table above

Variable	Meaning / Description
ProductID	Product type identifier (e.g., CLO, KEY, LAP, MON, MOU, SOF).
n	Sample size per subgroup (24 observations).
m_used	Number of subgroups used to calculate limits (30 samples).
xbarbar	Overall average of subgroup means – the process centre line for the X-bar chart.
sbar	Average of subgroup standard deviations – the process centre line for the s-chart.
UCL_x / LCL_x	Upper / Lower control limits for the X-bar chart (±3σ limits).
CL_x	Centre line for X-bar chart (= xbarbar).
UCL_s / LCL_s	Upper / Lower control limits for the s-chart (±3σ limits on sbar).
CL_s	Centre line for s-chart (= sbar).
U_1x, U_2x, L_1x, L_2x	±1σ and ±2σ zones for the X-bar chart (used to detect Western Electric rule violations).
U_{1s}, U_{2s}, L_{1s}, L_{2s}	±1σ and ±2σ zones for the s-chart (similar purpose as above).

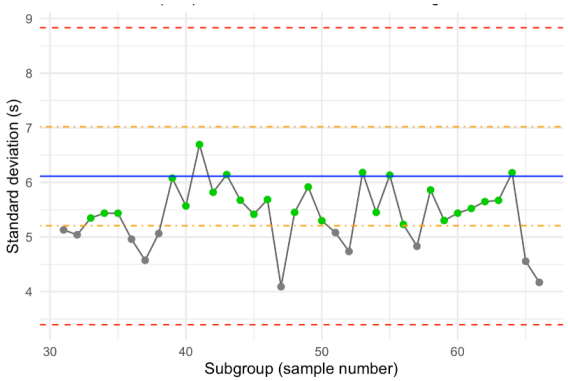
Interpretation of the S-Charts

Figure 21

The different S-charts show the speed of each product's delivery. Most of the product types remain in the ±3σ limits. This shows that there is some statistical control for the process variation.

figure 22

s-chart for CLO0011



There are some instances that move close to or above the UCL which can indicate a special cause variation. This is due to unusually high delivery times.

There are some products such as CLO0011and LAP028 that show periods of stable variation where most s values are in the +-σ limit. This shows process control and consistency in delivery times.

Interpretation of the X-bar Charts

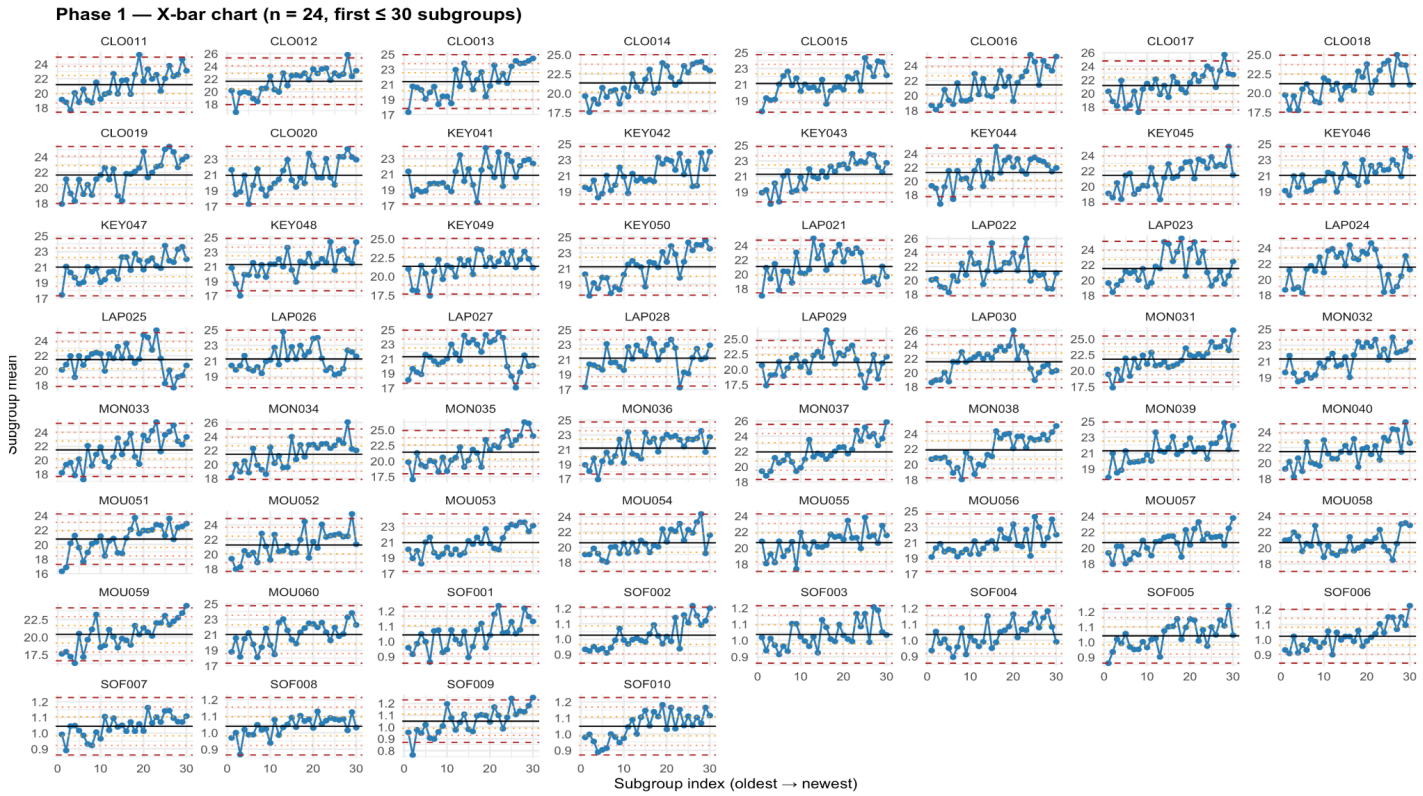


Figure 23

The X-bar chart is used to monitor the mean delivery time for each of the products over time. The goal is to determine whether the process is stable and in control. Each sub group represents a sample of 24 time deliveries.

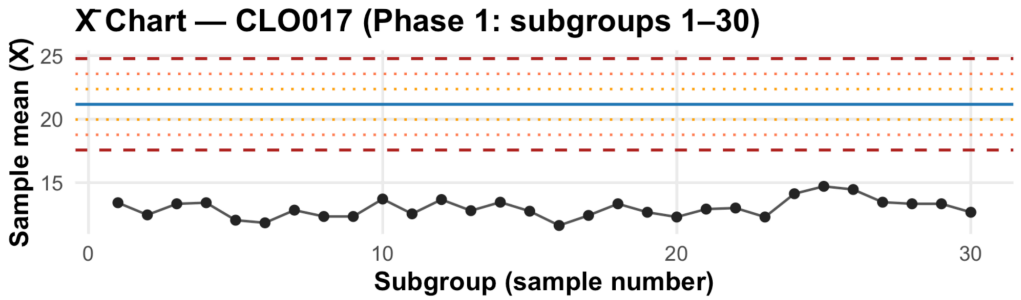


Figure 24

The horizontal axis is the subgroup index, this is time-ordered and it shows how the deliver performances evolve over time. The vertical axis is the sample mean delivery time for each batch of 24 sales. The center blue line indicates the average delivery time. The dashed red lines are the upper and lower control limits, and the orange dashed lines represent 1-sigma and 2-sigma zones.

3.2

X-Chart

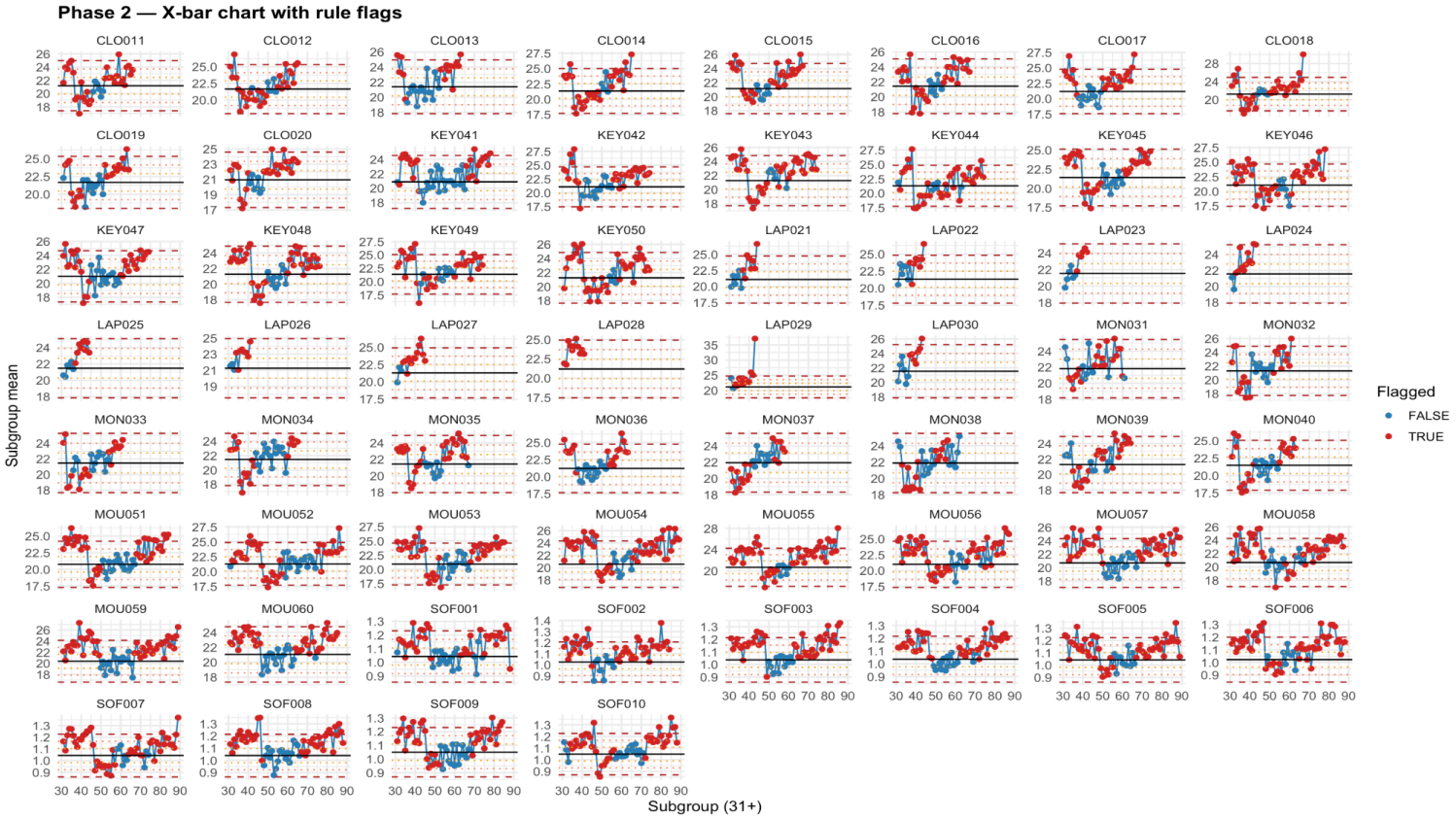


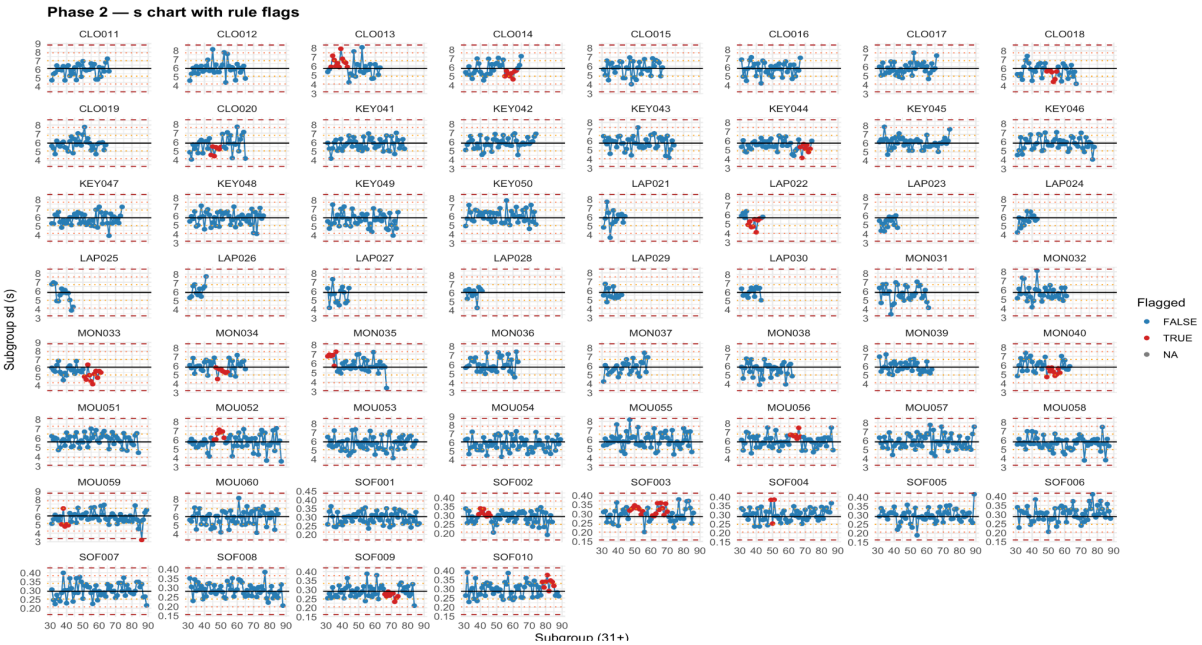
Figure 25

The phase 2 X- bar charts display subgroup means for each product type using control limits established from phase 1. There are now multiple red dots which all indicate points that violate one or more of western electric rules.

Most of the product types show a number of the red points which mean that their process averages drift beyond expected natural variation. A few charts show smaller values overall and tighter control, while others such as KEY041–KEY050 or MOU051–MOU060 exhibit significant variation and more red points.

Overall, the Phase 2 X-bar charts show that there were a number of products that went out of control. The high number of flagged points (red) demonstrates loss of control and the likely presence of special-cause variation.

S-Chart
Figure 26



When comparing the S-chart there are way fewer flagged points, which means there are fewer violations. This indicates that the spread has stayed relatively stable for most products.

The S-charts confirm that while the average delivery time for many products drifted, the variability within each sample stayed consistent. Therefore, special causes affecting process mean were more influential than causes affecting dispersion.

3.3

In the real world of manufacturing and distribution environments, process data is not collected once a year but rather continuously as sales and deliveries happen. So each new sub group of 24 deliveries represents an opportunity for monitoring an ongoing process. The S-Charts will be evaluated first, as there may be a slack of stability, making any conclusions from the X-chart would be unreliable. Once variability is under control monitoring for shifts in the process can be done through the X-charts.

From the phase 2 charts there were a number of product types that had repeated rule violations, like being outside the ±3σ limits. These are non random trends and they are either consistent under or over performers, this shows that they may have shifted due to special causes

In practice there would normally be a project manager that will intervene when a red point appears on the chart. The project manager would document it, then they would analyse what has caused this flagged point. If the S-Chart showed a sudden increase in standard deviation for a product, this would cause a review on delivery consistency or equipment performance. Same with the X-chart, if it showed an upward drift in delivery time averages, it would show signs that the process is no longer meeting the required service leave and this would need to be recalibrated straight away.

Regular reviews of the charts will alow managers to detect problems early and will maintain a stable and predictable delivery process that aligns with the customer expectations.

Process capabilities calculations

ProductID <chr>	n <int>	mean_x <dbl>	sd_x <dbl>	Cp <dbl>	Cpu <dbl>	Cpl <dbl>	Cpk <dbl>	Capable <lgl>
KEY045	1000	9.37600	3.489814	1.5282571	2.160956	0.8955587	0.8955587	FALSE
MON036	1000	10.21500	3.873730	1.3767955	1.874593	0.8789979	0.8789979	FALSE
CLO011	1000	9.93700	3.813341	1.3985984	1.928580	0.8686170	0.8686170	FALSE
KEY044	1000	8.99200	3.479661	1.5327161	2.204046	0.8613865	0.8613865	FALSE
CLO017	1000	9.82100	3.819889	1.3962012	1.935397	0.8570057	0.8570057	FALSE
MOU055	1000	8.31300	3.236279	1.6479830	2.439736	0.8562301	0.8562301	FALSE
MON040	1000	10.02500	3.910456	1.3638648	1.873183	0.8545465	0.8545465	FALSE
KEY050	1000	9.04600	3.546895	1.5036626	2.157192	0.8501332	0.8501332	FALSE
CLO013	1000	10.07600	3.960544	1.3466163	1.845201	0.8480316	0.8480316	FALSE
CLO014	1000	10.00800	3.970100	1.3433749	1.846469	0.8402810	0.8402810	FALSE
KEY043	1000	9.11900	3.617447	1.4743362	2.108393	0.8402795	0.8402795	FALSE
KEY041	1000	8.78400	3.487048	1.5294695	2.219260	0.8396787	0.8396787	FALSE
MON031	1000	10.26600	4.075764	1.3085480	1.777499	0.8395971	0.8395971	FALSE
CLO012	1000	9.81600	3.899404	1.3677305	1.896358	0.8391027	0.8391027	FALSE
MOU053	1000	8.32500	3.309128	1.6117038	2.384818	0.8385896	0.8385896	FALSE
KEY049	1000	9.05200	3.598645	1.4820393	2.125615	0.8384637	0.8384637	FALSE
SOF001	1000	8.24100	3.278974	1.6265250	2.415288	0.8377620	0.8377620	FALSE
MON039	1000	9.74900	3.889361	1.3712622	1.906997	0.8355272	0.8355272	FALSE
SOF004	1000	8.27100	3.308550	1.6119849	2.390674	0.8332955	0.8332955	FALSE
SOF005	1000	8.20500	3.283559	1.6242541	2.415570	0.8329378	0.8329378	FALSE
KEY048	1000	8.91400	3.567539	1.4949617	2.157043	0.8328805	0.8328805	FALSE
MOU051	1000	8.37200	3.350947	1.5915899	2.350380	0.8327994	0.8327994	FALSE
MON034	1000	9.95900	3.997912	1.3340296	1.837709	0.8303501	0.8303501	FALSE
MOU052	1000	8.20700	3.304099	1.6141568	2.400352	0.8279616	0.8279616	FALSE
MON038	1000	10.07200	4.057497	1.3144394	1.801439	0.8274396	0.8274396	FALSE
MOU058	1000	7.95800	3.206012	1.6635412	2.499679	0.8274038	0.8274038	FALSE
MON032	1000	9.90300	3.989675	1.3367838	1.846182	0.8273856	0.8273856	FALSE
SOF003	1000	8.10200	3.267763	1.6321054	2.437753	0.8264574	0.8264574	FALSE
KEY042	1000	8.97800	3.625280	1.4711505	2.116802	0.8254993	0.8254993	FALSE
SOF008	1000	8.15700	3.297457	1.6174079	2.410241	0.8245748	0.8245748	FALSE
SOF002	1000	8.26900	3.343025	1.5953616	2.366220	0.8245028	0.8245028	FALSE
KEY047	1000	8.76900	3.565158	1.4959600	2.172040	0.8198796	0.8198796	FALSE
LAP021	1000	12.42100	5.054830	1.0550964	1.291108	0.8190845	0.8190845	FALSE
KEY046	1000	8.52600	3.471221	1.5364432	2.254154	0.8187321	0.8187321	FALSE
CLO020	1000	9.68800	3.949209	1.3504814	1.883246	0.8177165	0.8177165	FALSE
LAP025	1000	12.52800	5.115602	1.0425622	1.268798	0.8163262	0.8163262	FALSE
SOF010	1000	8.11600	3.320477	1.6061948	2.397647	0.8147423	0.8147423	FALSE
MOU054	1000	7.88000	3.224903	1.6537965	2.493098	0.8144948	0.8144948	FALSE
SOF007	1000	7.98600	3.274833	1.6285820	2.444298	0.8128660	0.8128660	FALSE
CLO016	1000	9.46800	3.883305	1.3734007	1.934091	0.8127098	0.8127098	FALSE
SOF009	1000	8.30300	3.425628	1.5568922	2.305855	0.8079298	0.8079298	FALSE
CLO015	1000	9.45600	3.903755	1.3662059	1.924984	0.8074277	0.8074277	FALSE
MOU060	1000	8.23200	3.399667	1.5687812	2.330425	0.8071379	0.8071379	FALSE
MON037	1000	10.20600	4.227936	1.2614509	1.718254	0.8046480	0.8046480	FALSE
MOU057	1000	8.05900	3.341959	1.5958705	2.387921	0.8038200	0.8038200	FALSE
LAP027	1000	12.44700	5.166225	1.0323463	1.261592	0.8031009	0.8031009	FALSE
LAP024	1000	12.64100	5.249159	1.0160357	1.229340	0.8027317	0.8027317	FALSE
MON033	1000	10.02200	4.173840	1.2778003	1.755218	0.8003821	0.8003821	FALSE
MON035	1000	9.54000	3.978722	1.3404639	1.881676	0.7992516	0.7992516	FALSE
LAP023	1000	12.95600	5.416956	0.9845629	1.171876	0.7972498	0.7972498	FALSE
MOU056	1000	8.13700	3.407615	1.5651218	2.334281	0.7959623	0.7959623	FALSE
SOF006	1000	8.07700	3.410088	1.5639870	2.338454	0.7895202	0.7895202	FALSE
LAP022	1000	12.33800	5.227723	1.0202020	1.253701	0.7867033	0.7867033	FALSE
CLO018	1000	9.53800	4.063135	1.3126154	1.842748	0.7824828	0.7824828	FALSE
MOU059	1000	8.10900	3.460216	1.5413298	2.301494	0.7811652	0.7811652	FALSE
LAP029	1000	12.86300	5.511316	0.9677059	1.157437	0.7779751	0.7779751	FALSE
LAP030	1000	12.64300	5.430566	0.9820954	1.188151	0.7760395	0.7760395	FALSE
LAP028	982	12.70163	5.553886	0.9602887	1.158250	0.7623269	0.7623269	FALSE
LAP026	964	12.62759	5.655505	0.9430339	1.141802	0.7442655	0.7442655	FALSE

Figure 27

Process Capability Analysis (Cp, Cpu, Cpl, Cpk)

The image above is all the process capability calculations for all product types, calculated from the first 1000 deliveries observed by each product. Each product was evaluated using specific limits of LSL= 0 and USL = 32 hours.

In theory many of the products have a potential capable spread if they were centered correctly as the Cp values for most of the products range from about 1.1 to 1.6. However due to the CPK values being below 1 consistently, this does confirm that the processes are not able to meet the VOC requirements. And this can further be seen in the capable column as all the products read false. So in practice this shows that significant portions of the delivery times have fallen outside of the acceptable limits, thus there is alot of process improvement required.

This is a table to show what each index means and defines them

Table 3 This was generated by AI

Index	Full Name	Meaning / Purpose
Cp	**Process Capability Index**	Measures the **potential capability** of a process assuming it is centered between the limits. It shows how well the process spread fits within the specification limits.
Cpu	**Upper Process Capability Index**	Measures process capability **relative to the upper specification limit**. Indicates how close the process mean is to the **upper limit**.
Cpl	**Lower Process Capability Index**	Measures process capability **relative to the lower specification limit**. Indicates how close the process mean is to the **lower limit**.
Cpk	**Process Capability (Adjusted for Centering)**	Represents the **actual process capability**, considering both **spread and centering**. It tells whether the process is both stable and centered within limits.

3.4

Figure 28

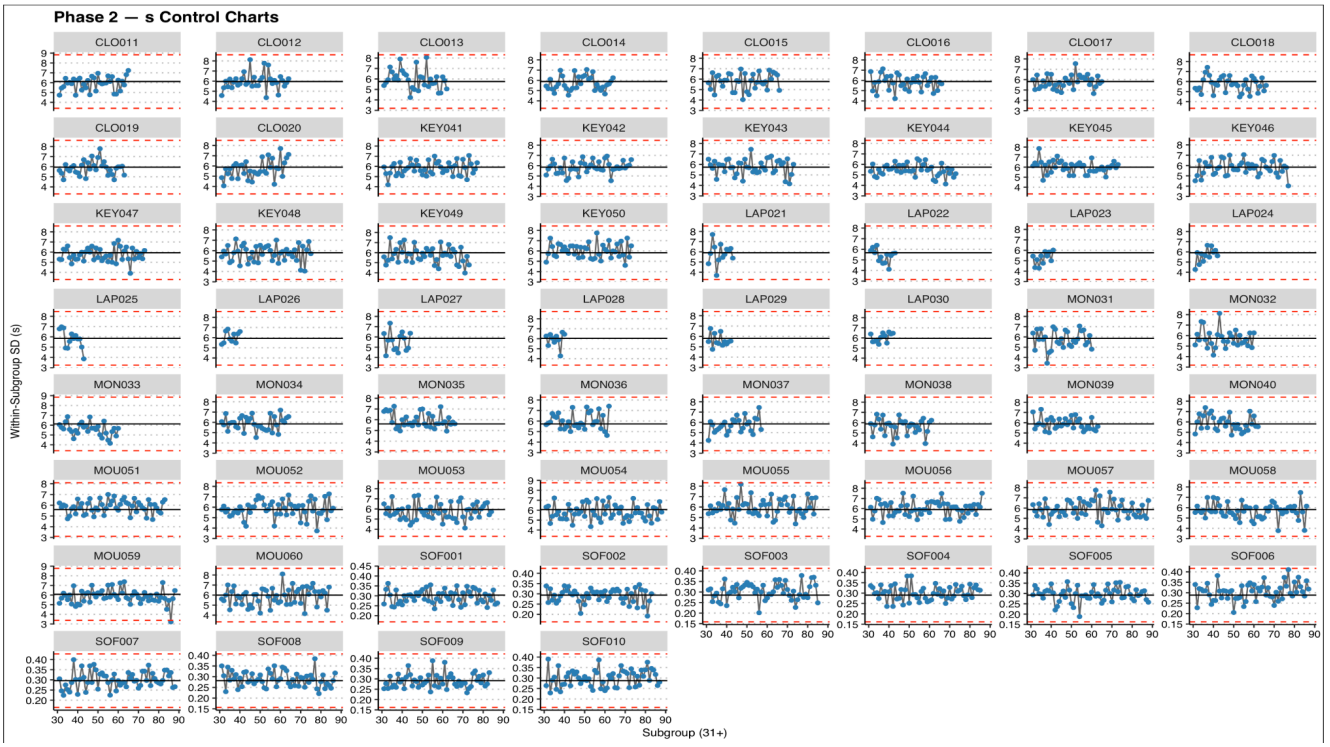
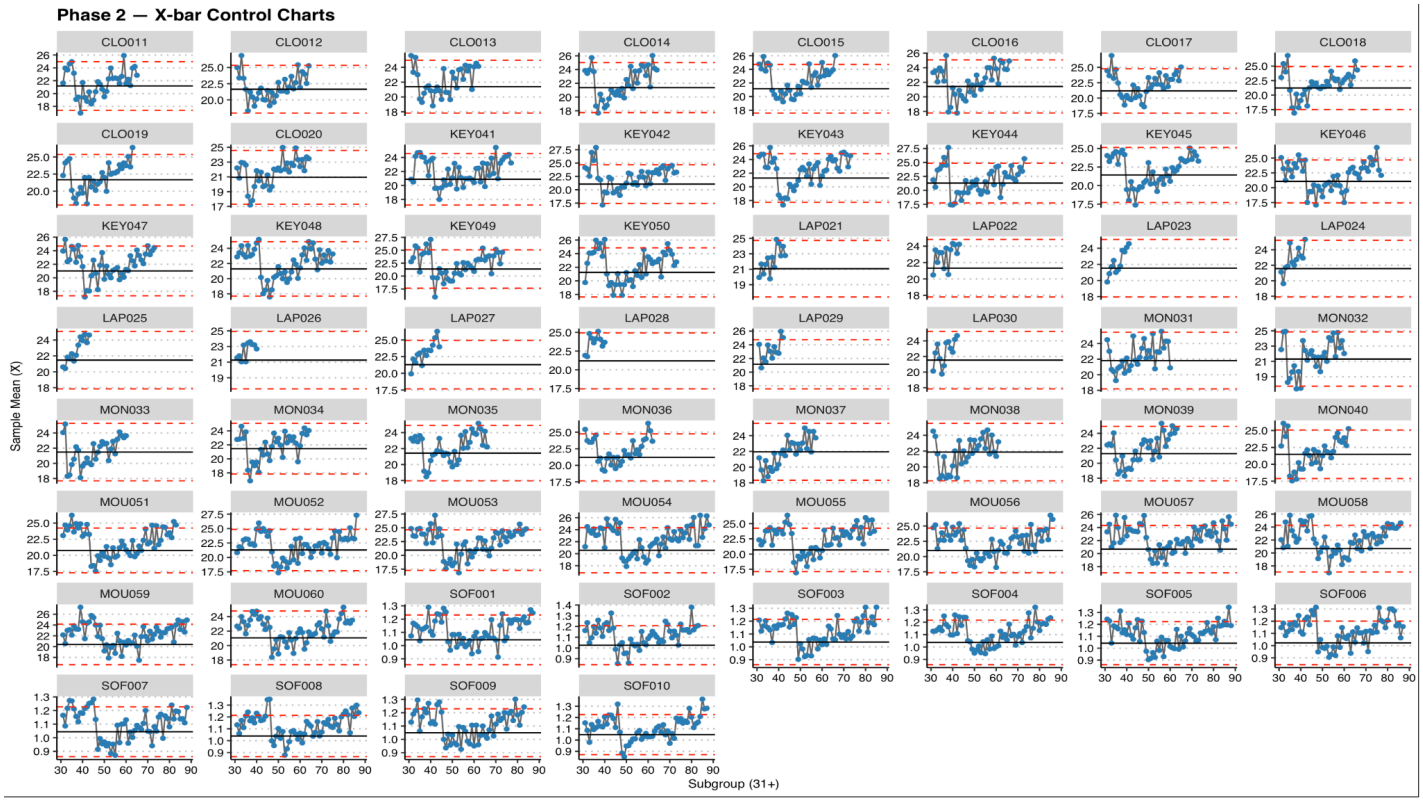


Figure 29



Rules

Rule	Description	Interpretation Goal
A	1 *s* sample outside of the upper +3σ control limit for all product types	Detect any points showing **unusually high variation** in delivery time
B	Find the most consecutive samples of *s* between the −1σ and +1σ control limits	Identify products showing **stable, consistent control**
C	4 consecutive $\bar{X}$ samples outside the upper +2σ limit	Detect products showing **a possible mean shift** or sustained deviation

Table 4

RULE A

- No products in the S-charts exceeded the +3σ control limit
- There for the variation is well controlled

figure 30

RULE B

- Each “long run” value show how many consecutive subgroups had there s values in the −1σ and +1σ control limits
- The high value 16-18 means excellent process consistency, the product's process times stayed within tight limits.
- KEY043 and KEY044 have runs of 18 this shows strong process stability.
- LAP027 has only 3 → indicates higher variability and less consistency.

ProductID <chr>	Longest_Run <dbl>
CLO011	8
CLO012	10
CLO013	5
CLO014	5
CLO015	8
CLO016	6
CLO017	12
CLO018	6
CLO019	7
CLO020	4
KEY041	18
KEY042	10
KEY043	8
KEY044	18
KEY045	16
KEY046	7
KEY047	17
KEY048	9
KEY049	9
KEY050	14
LAP021	8
LAP022	6
LAP023	5
LAP024	9
LAP025	7
LAP026	6
LAP027	3
LAP028	7
LAP029	9
LAP030	12
MON031	7
MON032	12
MON033	10
MON034	8
MON035	12
MON036	6
MON037	8
MON038	8

1–38 of 60 rows

RULE C

- Each count shows how many times a product had 4 or more consecutive subgroups above +2σ in the $\bar{X}$ chart. Meaning that the delivery time was consistently higher than what was expected.
- Bottom 3
    - MON040 - 0
    - SOF009 - 0
    - MON039 - 1
- These all have pretty stable means and have no systemic drift
- Top 3
    - SOF008 - 4
    - SOF002 - 3
    - MOU059 -3

ProductID <chr>	Runs_Above2Sig <int>
MON039	1
MON040	0
MOU051	2
MOU052	1
MOU053	2
MOU054	2
MOU055	2
MOU056	1
MOU057	1
MOU058	2
MOU059	3
MOU060	1
SOF001	3
SOF002	3
SOF003	2
SOF004	2
SOF005	2
SOF006	2
SOF007	2
SOF008	4
SOF009	0
SOF010	1

figure 31

This is an example of a X-Chart that has several points above the +2σ limit.

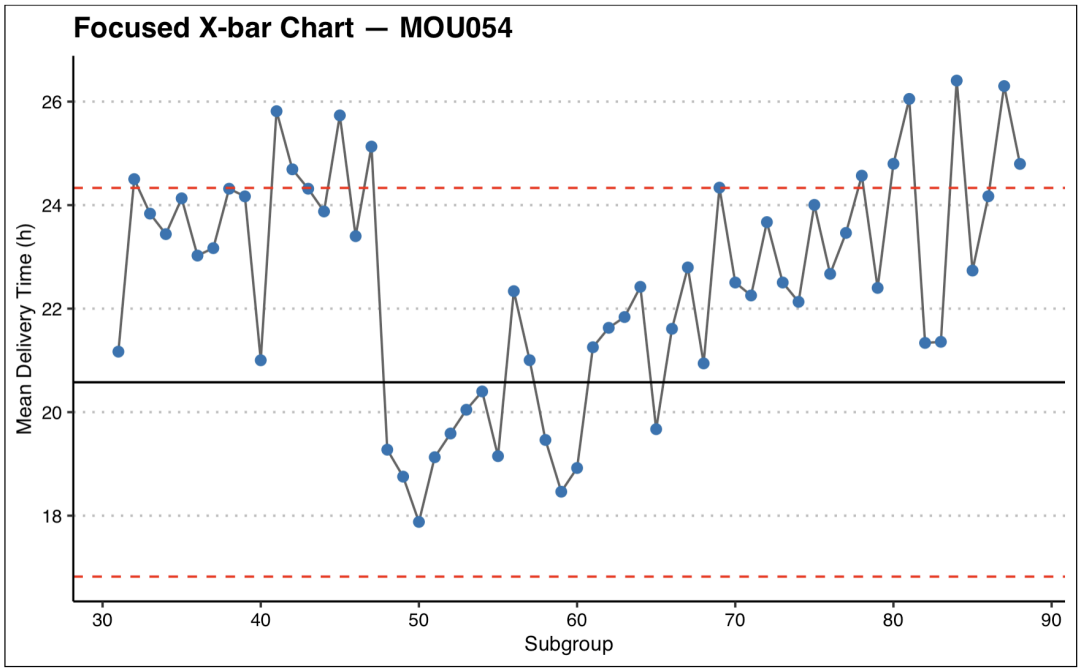


Figure 32

QUESTION 4
4.1

RULE A

- Under a stable process +3 is a one sided mormal tail
- Under H0 a +3σ one sided normal tail is the
  P(Z>3) = 0.00135
  Type 1 error
  pA = P(s > UCLs) ≈ 0.00135.
  Per week example
  $P(\geq\ 1\ false\ in\ a\ week)\ =\ 1\ -\ (1\ -\ 0.00135)^{N}.$

  If N = 60 == 7.8%
  If N= 100 == 12.6%

So even with a health process you should expect a red dot occasionally during a busy week

RULE B

- Rule B is looking for the longest consecutive sequence of s points inside the $\pm\ 1\sigma$. This is not an out of control alarm but rather an indicator of very tight and consistent variation.
- Rule B does not reject Ho. Therefore it does not create Type-1 risk. Therefore there is no type 1 probability to report
- Rule B is good news, not a false alarm

RULE C

- A run of 4 consecutive sub group mean above the +2σ line.
- Under Ho:
  $$P(Z > 2) = 1 - \Phi(2) \approx 0.0228.$$
- Assuming independence
  $$P(4\ \text{in a row above}\ +2\sigma)\ =\ (0.0228)^4\ \approx\ \mathbf{2.70 \times 10^{-7}}.$$
  $$P(\geq 1\ \text{false run in week})\ \approx\ 1 - \left(1 - 0.0228^4\right)^{N-3}.$$
- If N = 60 == 0.0015 %
- If N = 100 ==0.0026%

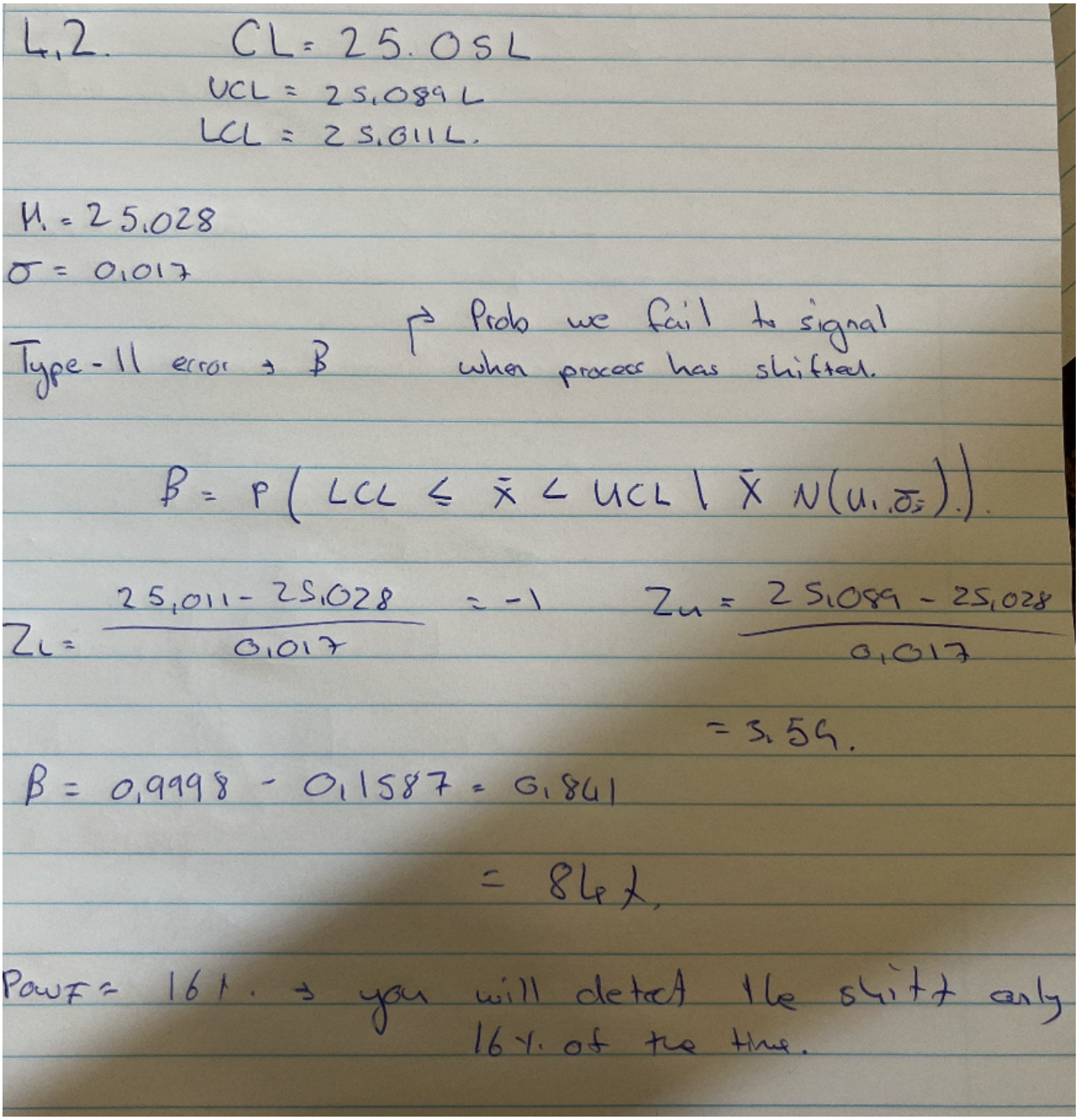
So therefore a 4 point push above +2σ is extremely unlikely by chance. If it does happen it is a strong signal the mean has drifted upwards

#The probability(1 sample > centreline) = 0.5 (You should know this, and explain why?)

- If the process is centered then each point is equally as likely to fall above or below the center line

Question 4.2

Working out for the question

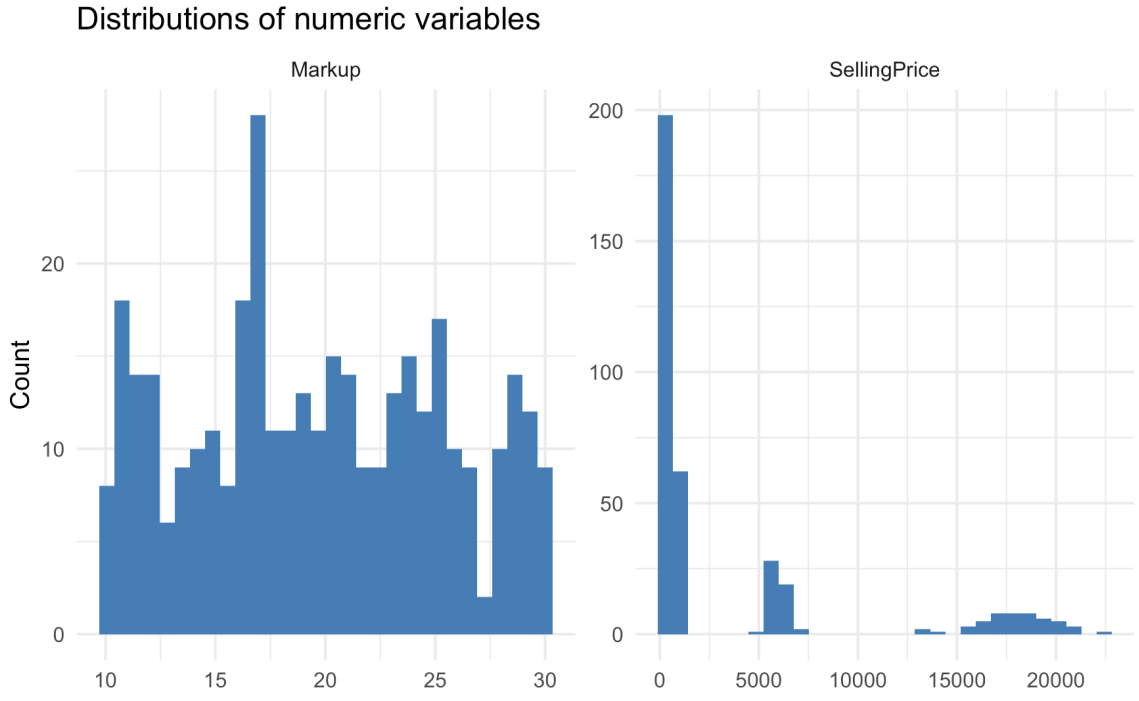


Overview

When using the frozen -3σ X-limits and using the values for UCL = 25.089 and LCL = 25.011. They type-II error for the X-bar chart is $\beta = 0.841$ or 84.1%. The true over all β will be lower once the S-charts are added because the increased variability makes the S-chart signal more likely.

4.3

Figure 33



Markup (left)
This is the new and updated data for the head office. The markup on the left now appears to be more uniform than before, though it still centers around 20-25. The bars on the histogram are more evenly distributed across the range of 10-30, with multiple smaller peaks throughout. There is no skewness observed, the data remains fairly symmetrical.there are no sharp gaps or isolated bars, this shows that it is a balanced distribution of mark up percentages. The updated version confirms a consistent markup policy, with no extreme outliers or heavy tails. This indicates the most products achieve markups within a similar range.

Selling price (right)
The selling price distribution retains its strong skewness to the right, but it does show some clearer grouping of prices. The first large spike remains near 0- 1000. Which still shows the majority of products sold are at lower costs. The selling price still follows a multi-modal pattern, showing that the products are segmented into distinct price tiers.

Figure 34

SellingPrice Distribution (Top Left)

- The density plot ids similar to the one before and remains highly skewed to the right.
- There are a number of smaller peaks between 5000- 2000. Suggesting three price clusters at low- , mid- and high-range items
- There is also a reduced high in the main peak which shows that there is now a more even spread.
- The updated distribution continues to show that most products are priced on the lower end, while a smaller subset of items drives up the average.

Markup Distribution (Bottom Right)

- **The mark up curve is roughly symmetrical. Which is stil the same as before.**
- The plot is slightly more flattened and has a broader peak.
- There are no extreme outliers or sharp peaks, suggesting controlled variation.
- The stable, near-normal distribution suggests that the company continues to apply consistent markup policies across product categories. The even spread demonstrates standardized profit margins with only moderate variation.

Scatterplot (Bottom Left)

- Each dot represents an individual product's SellingPrice–Markup pair.
- The scatter remains diffuse and unpatterned, with no visible linear or nonlinear trend.
- The data points are clustered vertically due to the broad ranges in the prices, confirming that the markup percentages are independent of absolute price levels.
- There's still no clear relationship between selling price and markups.

Correlation (Top Right)

- The updated correlation coefficient increased slightly to r = 0.070, but this value remains very close to zero.

- This confirms an extremely weak, statistically insignificant relationship between the two variables.

QUESTION 5

My code for this question gave me this.

Shop A

baristas	served/day	reliable %	lost	profit
2	547.48	**100%**	0.52	14 424
3	547.46	100%	0.53	13 424
4	547.42	100%	0.58	12 423
5	547.47	100%	0.53	11 424
6	547.60	100%	0.40	10 428

- The best number of baristas is 2 - this will give us the highest profit
- And 100% of the customers will receive reliable service, this mean it will be less than 5 min of service time

Shop B

baristas	served/ day	reliable %	lost	profit R/day
2	546.45	98.8 %	1.55	14 394
3	546.77	100 %	1.23	13 403
4	546.78	100 %	1.23	12 403
5	546.97	100 %	1.03	11 409
6	546.93	100 %	1.07	10 408

- For shop B the best number of baristas is also 2 - this will also give the highest profit of R14 394
- But its will give a 98.8% service level

Overview
Both of the shops will work well with only 2 baristas. So an increase in staff will add cost without meaningful improvements. The service is already excellent so it does not really play much of a factor, so cost control only dominates. The model aligned with the analysts notes: monitor service times, ensure staffing, but avoid overstaffing during quiet periods.

QUESTION 6
6.2

DATA USED

- Dependent variables
    - Quantity
    - Picking hours
    - Delivery hours
- Independent variables
    - Year
    - Month
    - Interaction between year:month

HYPOTHESIS

	Null hypothesis (H_0)	**Alternative (H_1)**
Year	There is no difference in the mean response between 2022 and 2023.	The mean response differs between 2022 and 2023.
Month	Mean response is the same across all 12 months.	At least one month differs significantly.
Year × Month	The month pattern is the same for both years (no interaction).	The pattern of months differs by year (interaction).

Table 5

Code output

```
 Response quantity :
             Df Sum Sq Mean Sq F value Pr(>F)
Year          1      0   0.068  0.0004 0.9848
Month        11    902  81.966  0.4351 0.9408
Year:Month   11   1652 150.192  0.7973 0.6431
Residuals  1557 293314 188.384

 Response pickinghours :
             Df  Sum Sq Mean Sq F value               Pr(>F)
Year          1     2.9   2.918  0.4047              0.52478
Month        11  2215.9 201.448 27.9337 < 0.0000000000000002 ***
Year:Month   11   159.2  14.470  2.0065              0.02453 *
Residuals  1557 11228.5   7.212
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response deliveryhours :
             Df Sum Sq Mean Sq F value              Pr(>F)
Year          1     27   27.49  0.7528              0.3857
Month        11   4263  387.54 10.6128 <0.0000000000000002 ***
Year:Month   11    320   29.11  0.7971              0.6433
Residuals  1557  56856   36.52
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 35

Factor	F-value	p-value	Interpretation
Year	**0.0004**	**0.9848**	**No significant difference between 2022 and 2023 (p ≫ 0.05).**
Month	**0.435**	**0.941**	**No significant difference in mean quantity across months (p ≫ 0.05).**
Year × Month	**0.797**	**0.643**	**No significant interaction between year and month.**

Table 6

Interpret graphs

Box plot - QUANTITY BY YEAR

- The median quantity in 2022 is slightly higher than 2023
- Both years show similar spread

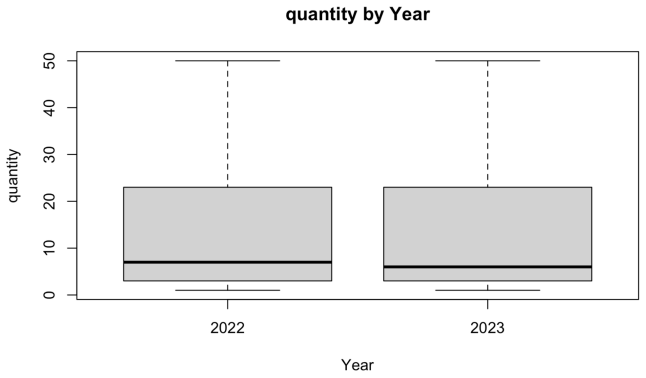


Figure 36

Box plots - QUANTITY BY MONTH

- Median quantity rises slightly mid-year, dips later
- The boxplot shows small variations in quantity; however, the ANOVA shows a value of p = 0.941, which shows that statistically it is not significant.
- Therefore, the null hypothesis is supported, indicating that production quantity does not vary significantly across months.

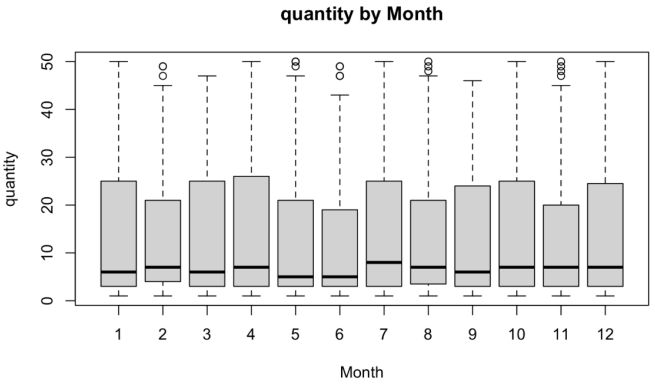


Figure 37

Line graphs - YEAR X MONTH ON QUANTITY

- The 2 lines for 2022 and 2023 roughly move parallel to one another.
- They have similar month to month fluctuations
- The ANOVA results show that the interaction is not statistically significant
- Therefor the null hypothesis is supported

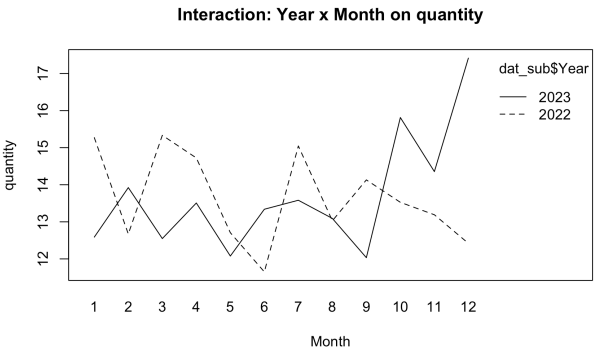


Figure 38

Conclusion

Although there are minor fluctuations that are visible in the plots, the ANOVA results confirm that both year and month effects are insignificant, and there is no meaningful year x month interaction. Thus over all the findings do support the Null hypotheses.

QUESTION 7

7.1

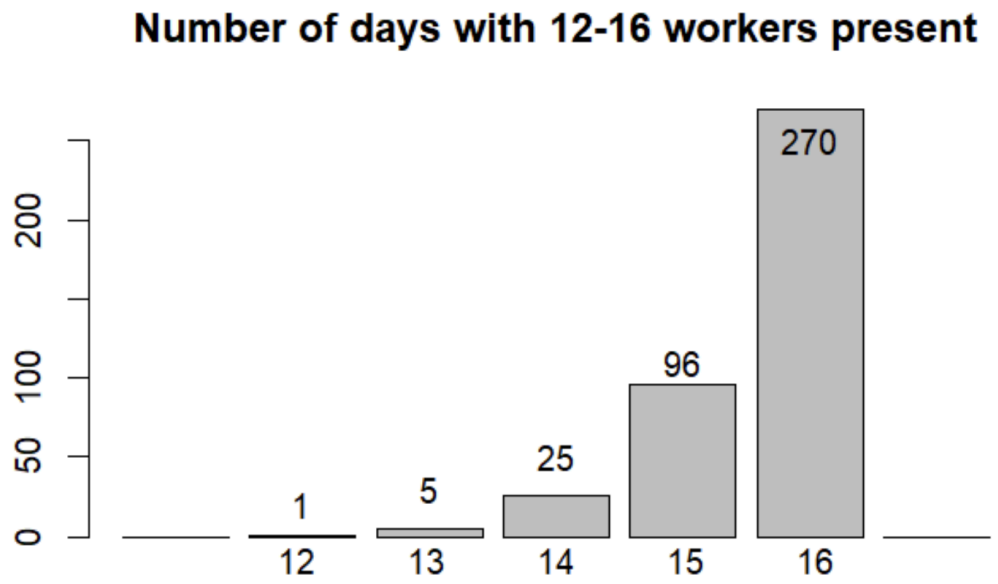


Figure 39

No. of Workers	Days
12	1
13	5
14	25
15	96
16	270

Reliable service mean having 15 or more workers

$$P(\geq 15\ workers) = \frac{96 + 270}{365} \approx 0.922$$

Therefore we have 365 x 0.922 = 336 days per year were reliable service is expected with a reliability of 92.2%

7.2

The weighted average p from my code is p = 0.974

- This means that on average 97.4% of the time works show up

$$P(X \geq 15) = 1 - P(X \leq 14)$$

This is equal to = 0.94
This means that we expect reliable service on 94% of the days which is: 365 x 0.94 = 343 days of the year

n <dbl>	reliability <dbl>	expected_loss <dbl>	extra_cost <dbl>	expected_profit <dbl>
14	0.0000000	7300000.0000000	0	-7300000.0
15	0.6738208	2381108.3181319	0	-2381108.3
16	0.9363690	464506.1593435	0	-464506.2
17	0.9909288	66219.8165223	300000	-366219.8
18	0.9989599	7592.9697751	600000	-607593.0
19	0.9998986	739.9417982	900000	-900739.9
20	0.9999913	63.4857935	1200000	-1200063.5
21	0.9999993	4.9135639	1500000	-1500004.9
22	1.0000000	0.3491348	1800000	-1800000.3

Figure 40

This table shows the expected annual profit evaluated for staffing levels from 14 to 22 employees. As can be seen in the table the optimal level is 17 workers which also archives a 99%reliability. Hiring more employees slightly increases reliability but your profit is reduced at higher fixed costs.

Figure 41

References

Asq.org. (2025). *Control Chart - Statistical Process Control Charts | ASQ*. [online] Available at: https://asq.org/quality-resources/control-chart?srsltid=AfmBOoomtZpWyxkQh9wkgz9cNwdk0Wq1hMmL5xFFKp4TaGwPVLrNU10A.

Sthda.com. (2025). *MANOVA Test in R: Multivariate Analysis of Variance - Easy Guides - Wiki - STHDA*. [online] Available at: https://www.sthda.com/english/wiki/manova-test-in-r-multivariate-analysis-of-variance.

Rproject.org. (2025). *rproject.org*. [online] Available at: https://cran.rproject.org/web/packages/MANOVA.RM/vignettes/Introduction_to_MANOVA.RM.html.

Bhandari, P. (2021). *Type I & Type II Errors | Differences, Examples, Visualizations*. [online] Scribbr. Available at: https://www.scribbr.com/statistics/type-i-and-type-ii-errors/.