UNIVERSITEIT·STELLENBOSCH·UNIVERSITY
jou kennisvennoot • your knowledge partner

Quality Assurance 344:

ECSA Project GR4 Data Analysis:

24/10/2025

Ralph Shpherd (27169014)

Table of Contents

...... i

Quality Assurance 344:...... i

Introduction:......1

Part 1 – Data Analysis: 2

1. **Inspection of Data:......2**
2. **Summary of Statistics:......3**
3. **Handling missing values: 6**
4. **Data filtering and subsetting:......6**
5. **Visual Analysis: 7**

Part 3 - SPC Analysis:......11

SPC (X and S-Charts) : 11

1. **Initialising X-Bar and S-Charts:......11**
2. **Continue Sampling:......12**
3. **Real-Life Interpretation: 13**
4. **Capability: Cp, Cpu, Cpl and Cpk (LSL = 0hrs, USL = 32hrs): 14**
5. **Control Issue Identification by Rules: 15**

PART 4 – Error Types:......16

Risk - Type I and Type II Error Analysis: 16

1. **Estimation of Type I Errors:......16**
2. **Estimation of Type II Errors:......17**
3. **Comparative Analysis: Identifying and Visualising Product Data Changes in the 2025 Update: 18**

Part 5 – Profit Optimisation:......20

1. Shop 1 Review: 20

2. Shop 2 Review: 21

Comparative Analysis: Shop 1 vs Shop 2:..22

Part 6 – DOA and MANOVA/ANOVA: 23

1. Data Analysis of DOA, MANOVA and ANOVA: 23

2. Results Compared: DOE and ANOVA / MANOVA 25

Part 7 - Reliability of Service:...27

1. Estimation of Reliable Service:...27

Profit Optimisation: 27

Conclusion: 29

References:...30

Introduction:

This report outlines a comprehensive analysis of several engineering datasets as part of the Quality Assurance 344 module under ECSA GA4 outcomes. The purpose of this investigation was to assess operational performance by employing a combination of statistical and process-control methods. Integrated descriptive and statistical measures were used to examine and assess data integrity, process stability, variation and capability across several operational and sales datasets.

The task was started by inspecting data with summary statistics to validate and present the key features for each of the customer, product and sales datasets with respect to dataset completion and summary measures. We employed visual and descriptive purposes such as histograms, boxplots, and scatter plots, to characterise distribution patterns and to identify possible process differences. We applied Statistical Process control charts, including X-bar and S-charts as part of our analysis to identify stability and special-cause variation.

The following sections explore the probabilities of Type I and Type II error to evaluate the credibility of control decisions and quantify the risk of false positives and undetected process shifts. Using comparative data from previous and new datasets facilitates an analysis of improvements around price and markup equivalency. Profit and efficiency models were created to estimate the most economical staff levels. ANOVA and MANOVA tests were used to test the statistical significance of differences observed in processes. The report concludes by examining service reliability by implementing a binomial model to quantify performance and support judgement regarding resource allocation.

The primary goal of the report is to demonstrate analytical skills as well as critical and evidence-based judgement in process performance and operational efficiency using quantitative approaches in accordance with engineering professionalism standards.

Part 1 – Data Analysis:

1. Inspection of Data:

Figure 1: Number of entries

```
$customers_rows_cols
[1] 5000    5

$products_rows_cols
[1] 60  5

$headoff_rows_cols
[1] 360   5

$sales_rows_cols
[1] 100000      9
```

In the early stages of inspection of data one can see that there are an immense number of entries in the various spreadsheets. The information is shown by row and column with only sales having 9 columns and the rest having 5.

Figure 2: Main headings for the data set

```
[1] "city"      "year"      "month"     "sales"     "volume"    "median"
[7] "listings"  "inventory" "date"
```

Figure 3: Individual headings in each file once processed:

Customers:

```
[1] "customer_id" "gender"      "age"         "income"      "city"
```

Products:

```
[1] "product_id"    "category"      "description"   "selling_price"
[5] "markup"
```

Products Head Office:

```
[1] "product_id"    "category"      "description"   "selling_price"
[5] "markup"
```

Sales:

```
[1] "customer_id"    "product_id"     "quantity"       "order_time"
[5] "order_day"      "order_month"    "order_year"     "picking_hours"
[9] "delivery_hours"
```

In each of the various spreadsheets there are multiple entries under different headings and these all provide vital information as to what the sheet contains. These will further be used in the coming data analysis.

2. Summary of Statistics:

Figure 4: Customers

```
 customer_id           gender               age             income      
 Length:5000        Length:5000        Min.   : 16.00   Min.   :  5000  
 Class :character   Class :character   1st Qu.: 33.00   1st Qu.: 55000  
 Mode  :character   Mode  :character   Median : 51.00   Median : 85000  
                                       Mean   : 51.55   Mean   : 80797  
                                       3rd Qu.: 68.00   3rd Qu.:105000  
                                       Max.   :105.00   Max.   :140000  
     city          
 Length:5000       
 Class :character  
 Mode  :character  
```

Figure 5: Products

```
  product_id          category         description        selling_price   
 Length:60          Length:60          Length:60          Min.   :  350.4  
 Class :character   Class :character   Class :character   1st Qu.:  512.2  
 Mode  :character   Mode  :character   Mode  :character   Median :  794.2  
                                                          Mean   : 4493.6  
                                                          3rd Qu.: 6416.7  
                                                          Max.   :19725.2  
     markup     
 Min.   :10.13  
 1st Qu.:16.14  
 Median :20.34  
 Mean   :20.46  
 3rd Qu.:25.71  
 Max.   :29.84  
```

Figure 6: Products Head Office

```
  product_id          category         description        selling_price    
 Length:360         Length:360         Length:360         Min.   :  290.5  
 Class :character   Class :character   Class :character   1st Qu.:  495.9  
 Mode  :character   Mode  :character   Mode  :character   Median :  797.2  
                                                          Mean   : 4411.0  
                                                          3rd Qu.: 5843.3  
                                                          Max.   :22420.1  
     markup     
 Min.   :10.06  
 1st Qu.:15.84  
 Median :20.58  
 Mean   :20.39  
 3rd Qu.:24.84  
 Max.   :30.00  
```

Figure 7: Sales

```
 customer_id         product_id           quantity       order_time
 Length:100000      Length:100000      Min.   : 1.0   Min.   : 1.00
 Class :character   Class :character   1st Qu.: 3.0   1st Qu.: 9.00
 Mode  :character   Mode  :character   Median : 6.0   Median :13.00
                                       Mean   :13.5   Mean   :12.93
                                       3rd Qu.:23.0   3rd Qu.:17.00
                                       Max.   :50.0   Max.   :23.00
   order_day     order_month       order_year   picking_hours
 Min.   : 1.0   Min.   : 1.000   Min.   :2022   Min.   : 0.4259
 1st Qu.: 8.0   1st Qu.: 4.000   1st Qu.:2022   1st Qu.: 9.3908
 Median :15.0   Median : 6.000   Median :2022   Median :14.0550
 Mean   :15.5   Mean   : 6.448   Mean   :2022   Mean   :14.6955
 3rd Qu.:23.0   3rd Qu.: 9.000   3rd Qu.:2023   3rd Qu.:18.7217
 Max.   :30.0   Max.   :12.000   Max.   :2023   Max.   :45.0575
 delivery_hours
 Min.   : 0.2772
 1st Qu.:11.5460
 Median :19.5460
 Mean   :17.4765
 3rd Qu.:25.0440
 Max.   :38.0460
```

Data Summary and Interpretation

Figures 4 – 7 provide an integrated view of customer, product, and sales information. Customer data of 5000 shows a broad age distribution between 16–105 years with an average of 51 years and mean income around R 80 000, indicating an older and economically spread-out customer base. Product data from both branch and head office are consistent. Average selling prices are around R 4400–R 4500 with markup values of around 20% which show minimal variation therefore suggesting stable pricing across locations. The larger head office dataset of 360 covers more items, reflecting centralised product management. Sales data of 100 000 observation reveal moderate order sizes with a mean of around 13 units and continuous activity across 2022–2023. Order and delivery times are well distributed with mean picking and delivery hours of 14.7 h and 17.5 h respectively, indicating efficient logistics.

Figure 8: Descriptive Statistics (Sales)

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
quantity	1	1e+05	13.50347	13.7601316	6.000	11.458100	5.9304	1.0000000	50.0000	49.00000	1.044341146	-0.2185180	0.043513357
order_time	2	1e+05	12.93230	5.4951268	13.000	13.117887	5.9304	1.0000000	23.0000	22.00000	-0.227168462	-0.7101693	0.017377117
order_day	3	1e+05	15.49683	8.6465055	15.000	15.495088	10.3782	1.0000000	30.0000	29.00000	0.002772591	-1.2007412	0.027342651
order_month	4	1e+05	6.44813	3.2834460	6.000	6.445538	4.4478	1.0000000	12.0000	11.00000	0.006928166	-1.1764404	0.010383168
order_year	5	1e+05	2022.46273	0.4986115	2022.000	2022.453413	0.0000	2022.0000000	2023.0000	1.00000	0.149493651	-1.9776714	0.001576748
picking_hours	6	1e+05	14.69547	10.3873345	14.055	13.543098	6.9188	0.4258889	45.0575	44.63161	0.735709308	0.4143469	0.032847636
delivery_hours	7	1e+05	17.47646	9.9999440	19.546	17.775077	8.8956	0.2772000	38.0460	37.76880	-0.470487992	-0.8716457	0.031622600

Figure 8 of around 100 000 records shows consistent operational activity across 2022–2023. Order quantities vary widely between 1–50 units and a mean of ≈ 13.5, indicating diverse customer demand. Order times are spread through the day with a mean of ≈ 13 h therefore suggesting steady workflow. Picking and delivery hours show high variation showing differing process durations. Slight positive skew in quantity and picking hours indicates occasional large orders or longer tasks, while near-zero kurtosis confirms roughly normal distributions.

Figure 9: Variable Character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique
customer_id	0	1	7	8	0	5000
product_id	0	1	6	6	0	60

If we observe the following figure 9 based on the descriptive stats we can see that although there are thousands of entries, there are on 5000 unique customers and 60 products that they can choose from.

Figure 10: Variable Numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
quantity	0	1	13.50	13.76	1.00	3.00	6.00	23.00	50.00	
order_time	0	1	12.93	5.50	1.00	9.00	13.00	17.00	23.00	
order_day	0	1	15.50	8.65	1.00	8.00	15.00	23.00	30.00	
order_month	0	1	6.45	3.28	1.00	4.00	6.00	9.00	12.00	
order_year	0	1	2022.46	0.50	2022.00	2022.00	2022.00	2023.00	2023.00	
picking_hours	0	1	14.70	10.39	0.43	9.39	14.05	18.72	45.06	
delivery_hours	0	1	17.48	10.00	0.28	11.55	19.55	25.04	38.05	

This further confirms the statement made in the descriptive stats. It further provides histograms to visualise the distribution of the various data. It further shows that there are 0 missing entries in the 7 variables.

3. Handling missing values:

Figure 11: Quantity of missing values

```
$customers_total
[1] 0

$products_total
[1] 0

$headoff_total
[1] 0

$sales_total
[1] 0

$sales_by_col
   customer_id     product_id       quantity     order_time      order_day
             0              0              0              0              0
   order_month     order_year  picking_hours delivery_hours
             0              0              0              0
```

As seen in figure 11 there are no missing values to be found in the various data sets that were provided. Therefore, they are complete and can be properly analysed going forward.

4. Data filtering and subsetting:

Figure 12: Top 10 customers by total quantity ordered

CustomerID	TotalQuantity	Orders
CUST1193	14704	326
CUST1791	14626	322
CUST596	14212	319
CUST3721	13852	301
CUST2527	13773	303
CUST2277	13538	298
CUST1427	13335	296
CUST4729	12938	279
CUST3944	12855	286

Figure 13: Top 5 products by total quantity ordered

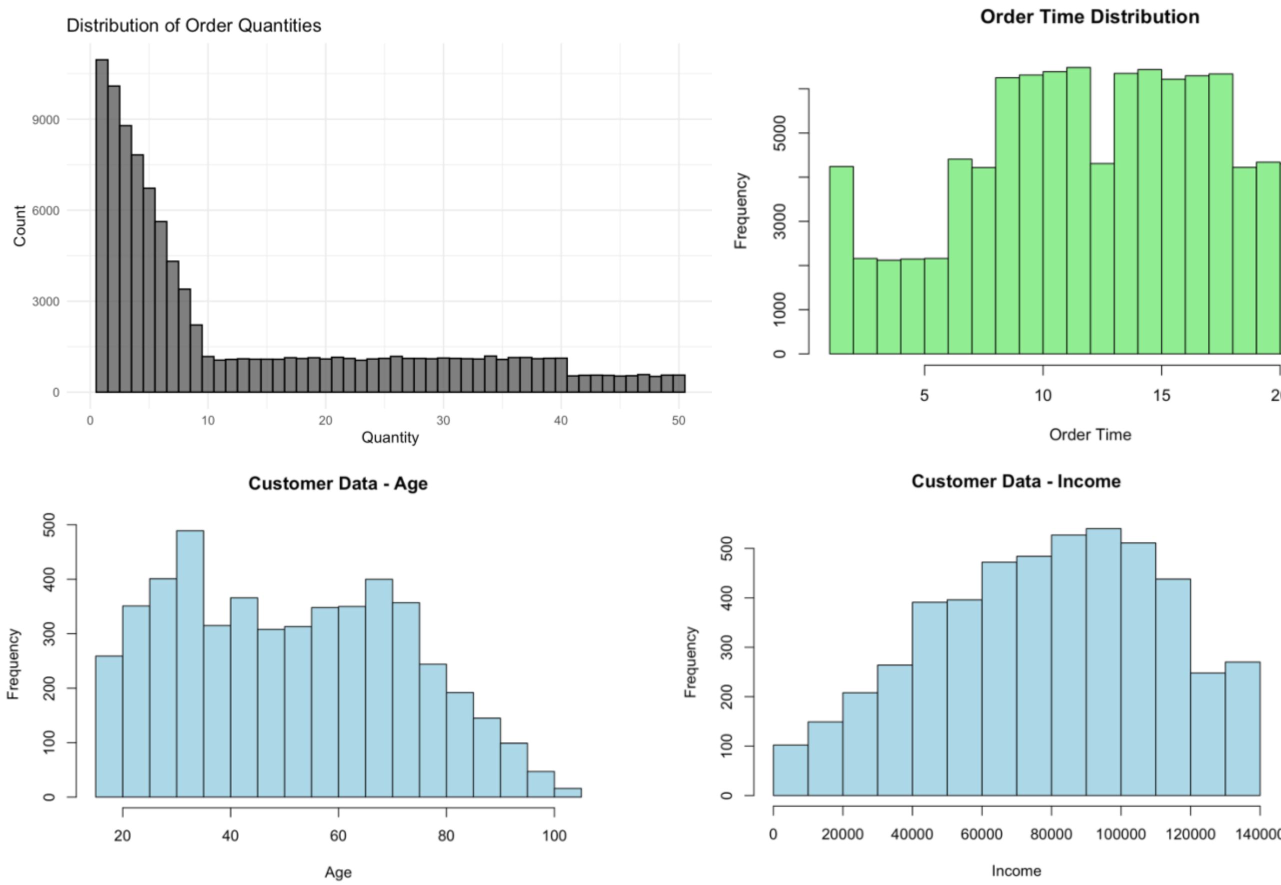
ProductID <chr>	TotalQuantity <int>
MOU059	29675
SOF001	29336
SOF004	29219
SOF010	29168
MOU058	28924

This analysis effectively highlights key customers and products driving sales performance. The top 10 customers show similar total quantities ordered showing a balanced customer base, while the top 5 products have noticeably high demand suggesting strong product performance and customer preference.

5. Visual Analysis:

Histograms:

Figure 14-17: Histograms



The Order Quantity histogram shows a strong right skew, with most orders consisting of fewer than 10 units. This suggests that smaller orders are more common reflecting potential retail-level transactions rather than bulk purchases.

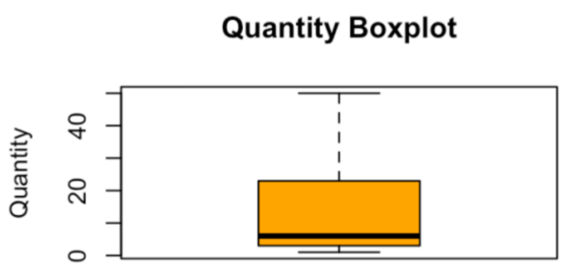
The Order Time distribution is relatively uniform across the 24-hour cycle, indicating consistent ordering throughout the day without a dominant time window. This may reflect an automated or continuous online ordering process. It can also be noted that there are no orders to be processed on Sundays.

The Customer Age histogram displays a wide distribution ranging from about 20 to 100 years, with a concentration between 30 and 70 years. This implies a broad but mature customer base. This is also validated in the previous data handling.

The Customer Income histogram is centered around R80 000–R100 000. This reflects a middle to upper income market segment with customers who can afford the company's products.

Boxplots:

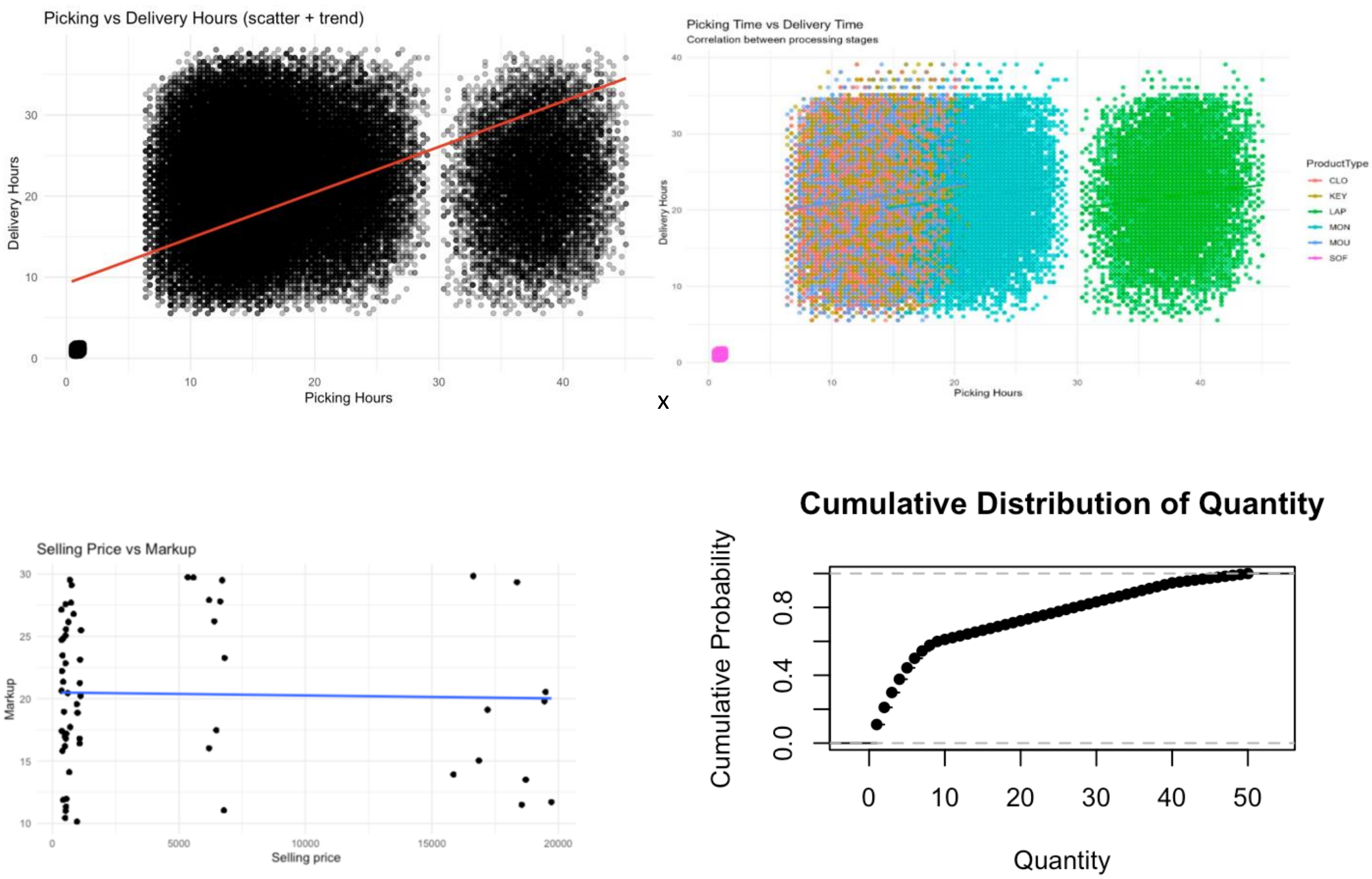
Figures 18-19: Boxplots



The Quantity boxplot shows that most orders are small with a median of around 6 units and a few larger outliers up to 50 unit showing that high-quantity orders are uncommon. The Picking vs Delivery Hours boxplot shows that delivery generally takes longer than picking, with median times of roughly 19 and 14 hours respectively. Both distributions show moderate variability and some extreme values, meaning potential occasional delays or operational faults in certain cases.

Scatterplots:

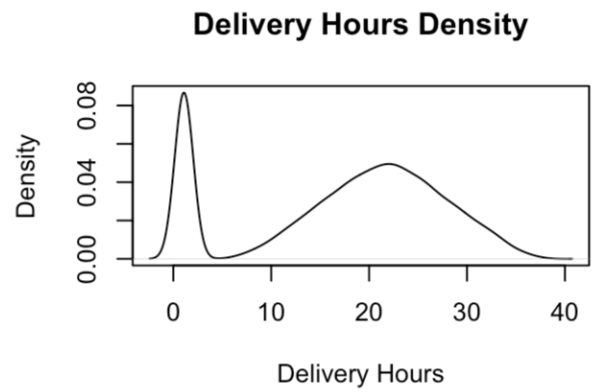
Figure 20-23: Scatterplots



The Picking vs Delivery Hours scatterplot shows a positive trend meaning longer picking times generally lead to longer deliveries The further provide colourful plot shows the various products and how they are distributed amongst the picking and delivery. The Cumulative Quantity plot shows most orders have fewer than 10 items, confirming that small orders dominate. The Selling Price vs Markup scatterplot shows almost no correlation indicating consistent profit margins across product prices. Overall, the results suggest stable operations and a uniform pricing approach.

Line Plots:

Figure24-25: Line Plots



The Monthly Order Quantities line plot shows steady sales activity with minor fluctuations across 2022 and 2023. Order volumes remain generally high suggesting consistent customer demand with only brief dips likely due to seasonal or operational factors. The dips also were repeating on specifically the 12th month being December/January suggesting closing over those periods. The Delivery Hours Density plot has two peaks indicating two common delivery time ranges with one very short and one around 20 hours therefore suggesting both quick deliveries and longer scheduled ones within the system.

Part 3 - SPC Analysis:

SPC (X and S-Charts) :

In the following section many graphs were produced as there are 6 main products and give the sales of 2026 and 2027 there are around 60 graphs that have been plotted as we are required to produce a X and S chart. I have selected various graphs and products specifically to answer each of the questions that form a part of part 3.

1. Initialising X-Bar and S-Charts:

Figure 26: SPC CLO011

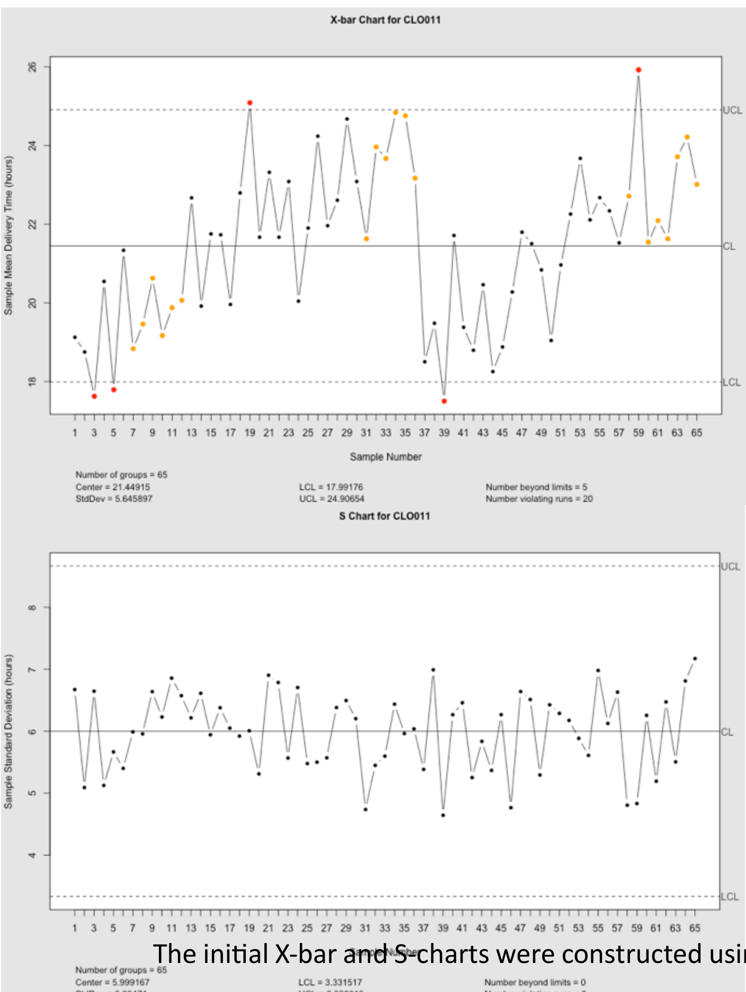
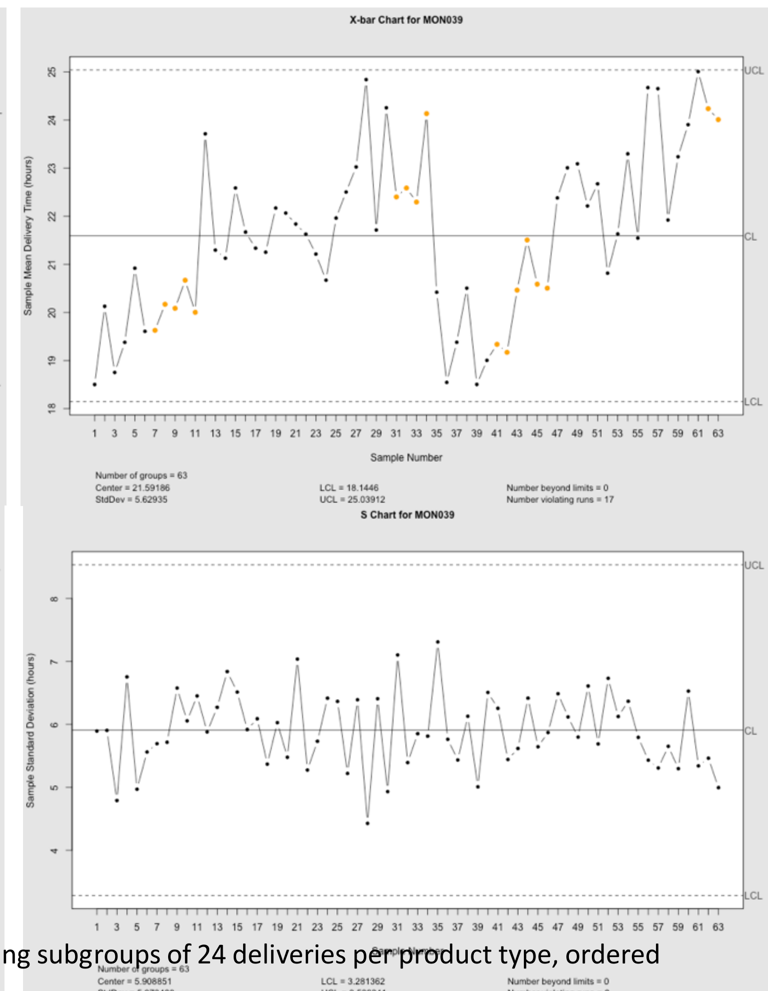


Figure 27: SPC MON039



The initial X-bar and S-charts were constructed using subgroups of 24 deliveries per product type, ordered by Year, Month, Day, and order time. The process centre lines and control limits at ±1 σ, ±2 σ, and ±3 σ were established using Phase I subgroups 1 through 30. Figure 26 presents the s-chart and X-bar chart for CLO011, which demonstrates very good statistical control. Its standard-deviation values remain entirely within the ±1 σ band for 26 consecutive samples. This therefore indicates consistent variation and a stable mean. The chart functions as a perfect example of how to start a chart properly while keeping the process under control. Figure 27 shows the corresponding charts for MON039, where a run of 23 consecutive s-values also lies within the ±1 σ region. The consistency observed in both CLO011 and MON039 confirms that the chosen subgroup size of 24 is appropriate, the sampling order is correct, and the calculated limits effectively capture only random variation. Together, these charts validate that the process-monitoring framework is correctly established for subsequent control analysis.

2. Continue Sampling:

Figure 28: SPC CLO013

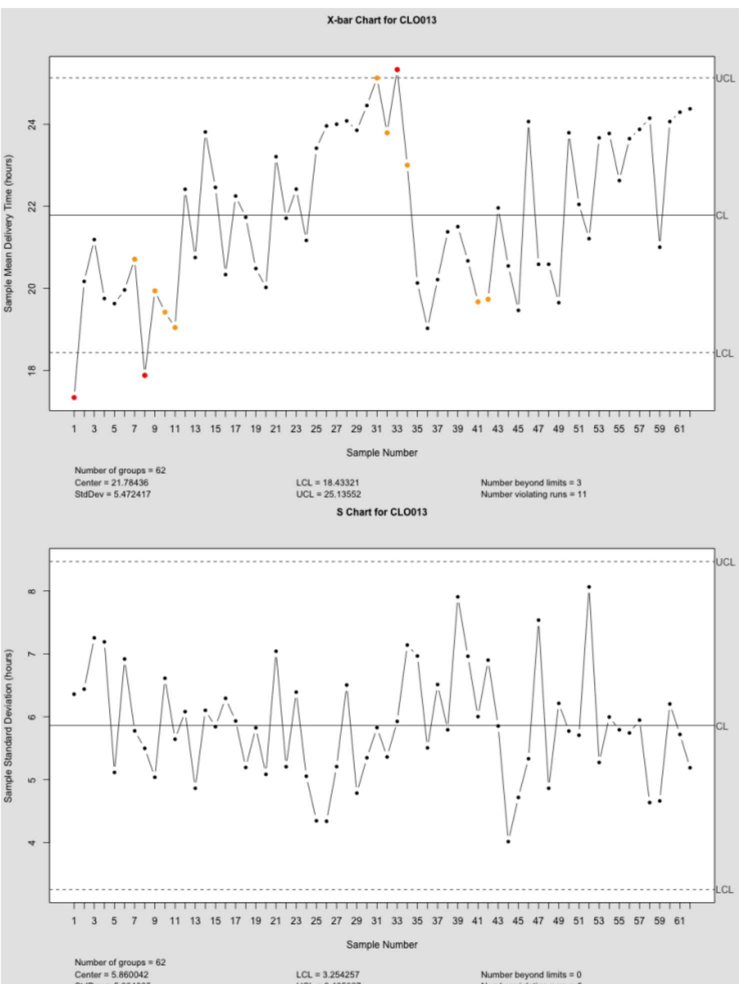
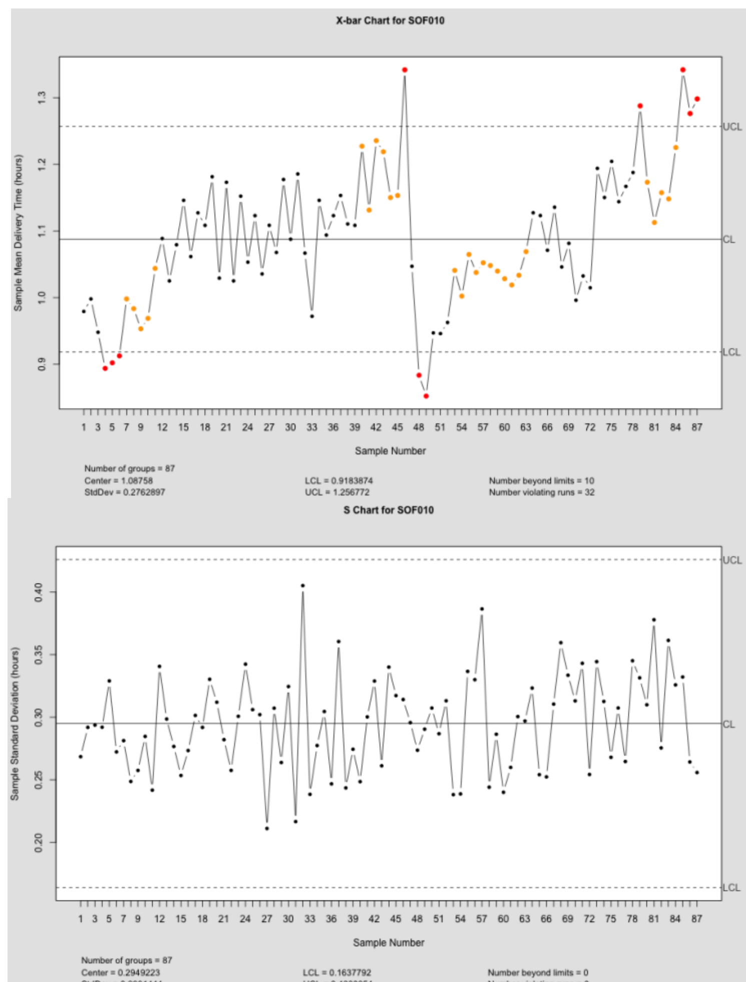


Figure 29: SPC SOF010



After the first thirty subgroup were used to develop initial control limits, further samples of twenty-four deliveries each were tracked in sequence against the Phase I limits to assess whether control was maintained. The Phase II assessment indicated that most product types stayed within expected variation, supporting the conclusion that delivery performance was stable. However, two product types showed patterns of non-randomness that warranted investigation. Specifically, as shown in Figure 28, product CLO013 showed six consecutive subgroup means (samples 26 to 31) above the +2 σ limit, demonstrating a sustained upward shift in the process mean and variation developing at the end of Phase I. Figure 29 shows product SOF010 exhibited runs of at least four means above +2 σ (samples 77-80 and samples 84 - 87) suggesting continued instability in Phase II. Therefore, the two product limits relate to a systematic change in the delivery process.

3. Real-Life Interpretation:

In real life usage the delivery data is made available based on the completion of sales in batches of twenty-four. Therefore, the s-chart must always be reviewed before the X-bar chart, because extreme spread variation compromises the reliability of the mean. The data show that for CLO011, the process variation was fully contained within control limits, with twenty-six consecutive samples 5 to 30 all falling between the ±1 σ limits. This indicates that the delivery process for this type of product is both stable and predictable, and the process is in control. CLO013 showed six consecutive subgroup means (samples 26 to 31) that were all above the +2 σ limit on the X-bar chart, which indicates a shift upwards in delivery times. Although the process s-chart limits were all within process control limits, the shift in the mean data may be an indication of variation associated with either workload imbalance or poor scheduling delays. CLO013 process requires management attention to investigate and eliminate the source of variation, while CLO011 can continue standard monitoring.

4. Capability: Cp, Cpu, Cpl and Cpk (LSL = 0hrs, USL = 32hrs):

Figure 30: Table of Complete Capabilities:

product_id	n	mean	sd	Cp	Cpu	Cpl	Cpk	capable_vs_VOC
CLO011	1000	21.272088000000007	6.2736471549554755	0.8501168780460661	0.5699986910870577	1.1302350650050745	0.5699986910870577	Not Capable
CLO012	1000	21.686243999999995	6.16817924509848	0.8646527802465284	0.5573636125115199	1.171941947981537	0.5573636125115199	Not Capable
CLO013	1000	21.467363999999996	6.210591149152055	0.8587480974434301	0.5653050703790115	1.1521911245078487	0.5653050703790115	Not Capable
CLO014	1000	21.33519399999999	6.077119832603587	0.8776087160105255	0.5849704187600847	1.1702470132609661	0.5849704187600847	Not Capable
CLO015	1000	21.53664799999999	6.018610433284723	0.8861403130261428	0.5794998760364204	1.1927807500158651	0.5794998760364204	Not Capable
CLO016	1000	21.53008	6.228785752419933	0.8562396501214206	0.5602975398499539	1.1521817603928872	0.5602975398499539	Not Capable
CLO017	1000	21.427162000000006	6.072971251172987	0.8782082299999702	0.58032208412852615	1.1760943758714142	0.58032208412852615	Not Capable
CLO018	1000	21.176997999999987	6.301021726259234	0.8464235746255085	0.57255275256369	1.120294396687327	0.57255275256369	Not Capable
CLO019	1000	21.538763999999997	6.133982922277726	0.8694731304131038	0.5684852258068912	1.1704610350193165	0.5684852258068912	Not Capable
CLO020	1000	20.895045999999997	5.957779298703167	0.8951881340240722	0.6213139406051974	1.169062327442947	0.6213139406051974	Not Capable
KEY041	1000	21.471012000000005	6.060009225708475	0.8800866689620946	0.5791513735288664	1.1810219643953228	0.5791513735288664	Not Capable
KEY042	1000	21.53875799999999	6.15817776349489	0.8660570607345637	0.566252031134561	1.1658620903345664	0.566252031134561	Not Capable
KEY043	1000	21.687798000000015	6.063591063700107	0.8795667909172692	0.5668919012769145	1.192241680557624	0.5668919012769145	Not Capable
KEY044	1000	21.542479999999998	6.058004870302481	0.8803778550060882	0.5754105641427043	1.1853451458694721	0.5754105641427043	Not Capable
KEY045	1000	21.837096000000017	6.297311838489117	0.8469222217543754	0.5379493271972758	1.1558951163114748	0.5379493271972758	Not Capable
KEY046	1000	21.776263999999994	5.953384233705913	0.8958490035193637	0.5724327317403156	1.2192652752984117	0.5724327317403156	Not Capable
KEY047	1000	21.498483999999994	6.094441410814761	0.8751143827339418	0.5743767307569138	1.17585203471097	0.5743767307569138	Not Capable
KEY048	1000	21.927504000000013	6.002367735853943	0.8885382515762462	0.5593623740530451	1.2177141290994473	0.5593623740530451	Not Capable
KEY049	1000	21.985003999999996	6.313649271784983	0.8447306943652103	0.5287484078215504	1.1607129809088703	0.5287484078215504	Not Capable
KEY050	1000	21.861354000000016	6.2734701592119215	0.8501408627092778	0.5387048285714972	1.1615768968470583	0.5387048285714972	Not Capable
LAP021	1000	21.355521999999993	6.146317712332895	0.867728220855185	0.5772821223045103	1.1581743194058598	0.5772821223045103	Not Capable
LAP022	1000	21.708243999999997	5.8136666795340695	0.9173785886467037	0.5900897871235156	1.2446673901698921	0.5900897871235156	Not Capable
LAP023	1000	21.76438799999999	5.844389525209488	0.9125561036491939	0.5837856378240589	1.241326569474329	0.5837856378240589	Not Capable
LAP024	1000	21.852922	6.071470903236282	0.8784252479066483	0.5570905942298812	1.1997599015834157	0.5570905942298812	Not Capable
LAP025	1000	21.756206	6.068480193048126	0.8788581594849801	0.5626776213114552	1.195038697658505	0.5626776213114552	Not Capable
LAP026	964	21.57464107883818	6.053739593443602	0.8809981418938977	0.5740451148825334	1.1879511689052622	0.5740451148825334	Not Capable
LAP027	1000	21.673698000000005	6.011877143420643	0.887132788328866	0.5725500678991213	1.2017155087586107	0.5725500678991213	Not Capable
LAP028	982	21.839415478615084	6.230184489497486	0.856047415983264	0.5436213827756942	1.168473449190834	0.5436213827756942	Not Capable
LAP029	1000	21.684191999999996	6.040962942147915	0.8828614551038815	0.5692143288407666	1.1965085813669964	0.5692143288407666	Not Capable
LAP030	1000	21.816322000000003	6.138973728788779	0.868766274128624	0.5529522495616022	1.1845802986956457	0.5529522495616022	Not Capable
MON031	1000	21.662808000000005	6.010447948670692	0.8873437352556869	0.57329016008345	1.2013973104279239	0.57329016008345	Not Capable
MON032	1000	21.143636000000004	5.987536703459887	0.8907391465761666	0.6043867752675134	1.1770915178848198	0.6043867752675134	Not Capable
MON033	1000	21.251855999999993	6.3248210716867765	0.8432386106870496	0.5664531258765221	1.120024095497577	0.5664531258765221	Not Capable
MON034	1000	21.27718200000001	6.137580810146367	0.8689634398811519	0.5823585509062202	1.1555683288560834	0.5823585509062202	Not Capable
MON035	1000	21.496357999999994	6.093399009692977	0.8752640890329714	0.574591290416154	1.1759368876497889	0.574591290416154	Not Capable
MON036	1000	21.514674	6.022891763174472	0.885510406469982	0.580304080514392	1.190716732425572	0.580304080514392	Not Capable
MON037	1000	21.465961999999998	5.920142964428748	0.9008791452130013	0.5931184468175797	1.2086398436084227	0.5931184468175797	Not Capable
MON038	1000	21.503337999999992	6.095439056133938	0.8749711520725179	0.5740172776909892	1.1759250264540466	0.5740172776909892	Not Capable
MON039	1000	21.078655999999995	6.074802419346509	0.8779435058411433	0.5992701899910712	1.1566168216912154	0.5992701899910712	Not Capable
MON040	1000	21.326304000000004	6.187527619140097	0.861949014471555	0.575011359248061	1.148886669695049	0.575011359248061	Not Capable
MOU051	1000	21.744439999999976	6.016143045468874	0.8865037438480113	0.5682245209536207	1.204782966742402	0.5682245209536207	Not Capable
MOU052	1000	21.75476400000002	5.935141283437014	0.8986025906774746	0.5753997257313819	1.2218054556235673	0.5753997257313819	Not Capable
MOU053	1000	21.875396000000013	6.168780427860204	0.864568514912004	0.5470883652720079	1.1820486645520003	0.5470883652720079	Not Capable
MOU054	1000	21.42171999999998	6.219815864717463	0.8574744734144314	0.5669128170394019	1.148036129789461	0.5669128170394019	Not Capable
MOU055	1000	21.401940000000014	6.008164308235482	0.8876810053318369	0.5879810347104447	1.1873809759532292	0.5879810347104447	Not Capable
MOU056	1000	21.72452799999999	6.114928390622284	0.8721824676659194	0.5601304078370044	1.1842345274948345	0.5601304078370044	Not Capable
MOU057	1000	21.404176000000017	6.032999094239538	0.8840268745317295	0.585437073363017	1.182616675700442	0.585437073363017	Not Capable
MOU058	1000	21.384764	6.030975889728502	0.8843234380055572	0.5867063746725224	1.181940501338592	0.5867063746725224	Not Capable
MOU059	1000	21.18894	6.319895793384507	0.8438957710214335	0.5702129883911863	1.117578553651681	0.5702129883911863	Not Capable
MOU060	1000	21.734307999999995	6.1126466334969844	0.8725080399882672	0.5598061753777025	1.185209904598832	0.5598061753777025	Not Capable
SOF001	1000	1.0694249999999998	0.310050488921854	17.201499510221904	33.253266919586366	1.149732100857441	1.149732100857441	Capable (Cpk≥1)
SOF002	1000	1.0646250000000002	0.3082288100979113	17.30316297051907	33.45498969869509	1.1513362423430542	1.1513362423430542	Capable (Cpk≥1)
SOF003	1000	1.0694249999999998	0.29547627756827727	18.049954389658005	34.89346674974351	1.2064420295725007	1.2064420295725007	Capable (Cpk≥1)
SOF004	1000	1.069825	0.3042943960286724	17.526886472239845	33.88185411196944	1.1719188325102496	1.1719188325102496	Capable (Cpk≥1)
SOF005	1000	1.078225	0.30837220658287745	17.29511680845971	33.42473190936933	1.1655017075500917	1.1655017075500917	Capable (Cpk≥1)
SOF006	1000	1.0594000000000001	0.3019032221682228	17.66570523835469	34.16171996861482	1.16969050809456	1.16969050809456	Capable (Cpk≥1)
SOF007	1000	1.0857499999999998	0.304488687574532	17.515702720574318	33.84280080184466	1.1886046393039726	1.1886046393039726	Capable (Cpk≥1)
SOF008	1000	1.0756750000000002	0.2924531213770907	18.236540982089583	35.24704501287234	1.2260369513068259	1.2260369513068259	Capable (Cpk≥1)
SOF009	1000	1.085725	0.30500570134483196	17.486011932949403	33.78546134677996	1.1865625191188431	1.1865625191188431	Capable (Cpk≥1)
SOF010	1000	1.069425	0.2964676499087927	17.989596284701268	34.77678481897962	1.2024077504229158	1.2024077504229158	Capable (Cpk≥1)

The process capability indices were determined using the first 1,000 delivery times for each product type with a lower specification limit of 0 hours and an upper specification limit of 32 hours. The indices C_p, C_{pu}, C_{pl} and C_{pk} were analysed to evaluate whether the processes are capable of consistently meeting customer delivery expectations. The results indicate that none of the product types achieved a C_{pk} value greater than 1.33, meaning that no process can currently be considered capable of meeting the Voice of the Customer (VOC). The top-performing and most consistent products were SOF008, SOF003 and SOF010 as seen in figure 30, which are relatively close to being capable and would require minimal improvement to meet specification limits. Conversely, the least capable product types were KEY049, KEY045 and KEY050 with Cpk of 0.529 – 0.539, all of which show substantial variation compared to their specification range. At the end of the day eight product types approached acceptable capability while the keyboard product line clearly requires significant process redesign to reduce variation and centre delivery times within specification limits.

5. Control Issue Identification by Rules:

The evaluation of the control charts was done according to established SPC rules to help identify any potential issues with the process. Rule A identifies any sample of s that is greater than the upper +3 σ limit. In the assessment, there were no violations for any product type, which confirms that process variation was still within the normal statistical range. Rule B indicates good control by looking for the longest number of samples of s that are in the ±1 σ range. The products that showed the most consecutive samples were CLO011 with 26 consecutive samples (samples 5 to 30), MON039 with 23 consecutive samples (samples 40 to 62) as seen in figure 26 and figure 27 respectively. These runs illustrate consistency and minimal variability in delivery. Rule C is used to identify four or more consecutive X-bar values that are greater than the +2 σ limit, and we found a total of 120 across all product types. The earliest value of the four or more consecutive values was for CLO013 during samples 26 to 31, showing a slow upward shift in the process mean, while SOF010 showed a similar outcome in later samples; specifically, for samples 77-80 and 84-87 as seen in figure 28 and 29 respectively. These values illustrate that while products such as CLO011 are in strong statistical control, products such as CLO013 and SOF010 are showing sustained mean shifts that require management investigation and possible correction.

PART 4 – Error Types:

Risk - Type I and Type II Error Analysis:

1. Estimation of Type I Errors:

The Type I error refers to the probability of incorrectly assuming that a process is out of control when it is actually stable and centred on its control limits. Under the null hypothesis (H_0), the process is assumed to be in control, with delivery data following a normal distribution centred on the calculated mean. The probability of observing a sample above the centreline is therefore $P(X > \text{CL}) = 0.5$, since the distribution is symmetric around the mean.

All Type I error probabilities were determined using standard theoretical values from the normal distribution. These probabilities hold for any stable process where the control limits are based on ±1σ, ±2σ, and ±3σ intervals.

Rule A which detects an s-sample exceeding the upper +3σ limit, the probability of such a false alarm under H_0 is $P(Z > 3) = 0.00135$. If 30 subgroups are examined in a week, the probability of at least one false signal is $1 - (1 - 0.00135)^{30} \approx 3.97\%$. Therefore even in a stable process around four percent of weeks may produce a false indication of excessive variation.

According to Rule B that considers process stability from extending runs of points within ±1σ, seven points in a row of a sample occurring purely by chance would happen with a probability of (0.6827) ^7 = 6.9%. In CLA011, there were twenty-six points in a row between samples 5–30 within ±1σ, which is only (0.6827)^26 ≈ 0.0049%. This value is so low that it indicates the process instability for CLO011 is genuine.

Referring to Rule C which is a signal for a shift from the mean where four in a row points are above the +2σ limit, the false-alarm rate per occurrence is (P(Z>2))^4 = (0.0228)^4 ≈ 2.7×10^(-7). Even if that was examined with 60 total potential starting points of the week, it would still be exceedingly low (<0.002%). Thus, violations observed such in CLO013 and SOF010 products would be even less likely to have occurred in isolation by chance which is another indication of a genuine shift.

These calculations provide further proof that SPC methodology (a sampling plan as evidenced) is also consistent with having a low false alarm (Type I error) rate, because there is evidence for process shifts in products such as CLO013 and SOF010, not just random variation.

Figure 31: Outcome of Error I

	Rule	Formula	Example.k.or.N.	Per.Test.Probability	Weekly.False.Alarm
1	A: s > +3σ	P(Z>3)	N=30	1.349898e-03	0.0397141697
2	B: k consecutive within ±1σ	(0.6827)^k	k=7,20,23,26	NA	NA
3	C: 4 Xbar > +2σ	(0.0228)^4	M=60	2.678772e-07	0.0000160725

2. Estimation of Type II Errors:

The Type II error shows the probability of failing to detect a real change in the process when it has moved away from the target mean. The bottle-filling process is assumed to be centred at 25.05 litres with control limits of 25.011 litres and 25.089 litres being the lower and upper respectively. The process mean shifts to 25.028 litres and the standard deviation of the sample means increases from 0.013 to 0.017 litres.

The Type II error probability is given by:

$$\beta = P(LCL \le \bar{X} \le UCL \mid \mu_1 = 25.028, \sigma_{\bar{X}} = 0.017)$$

Therefore:

$$z_U = \frac{25.089 - 25.028}{0.017} = 3.59, z_L = \frac{25.011 - 25.028}{0.017} = -1.00$$

$$\beta = \Phi(3.59) - \Phi(-1.00) = 0.841$$

The probability of **failing to detect the process shift is 84.1%**, while the power to detect it $(1-\beta)$ is only 15.9% this is also verified within the R-Code and seen in Figure X. This then means that the current control chart is relatively unchanged to this small mean shift and would likely continue to signal the process as in control. The bottle-filling process would consistently underfill bottles without triggering an alarm posing a risk to product quality.

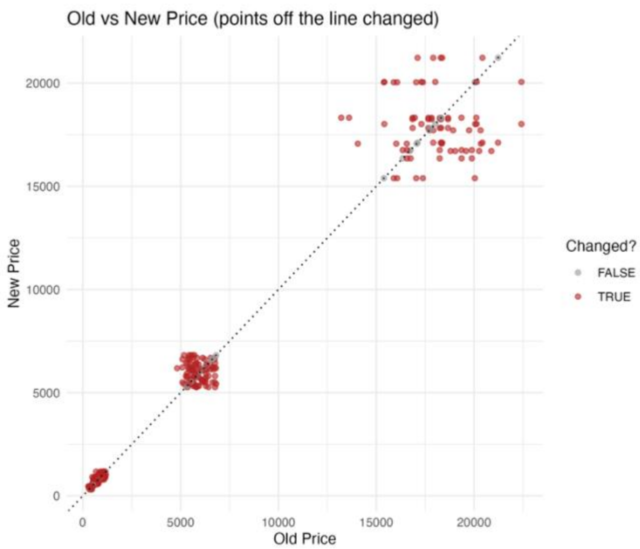
To reduce the likelihood of Type II errors the process could be improved by increasing the sample sizeand reducing natural variation.

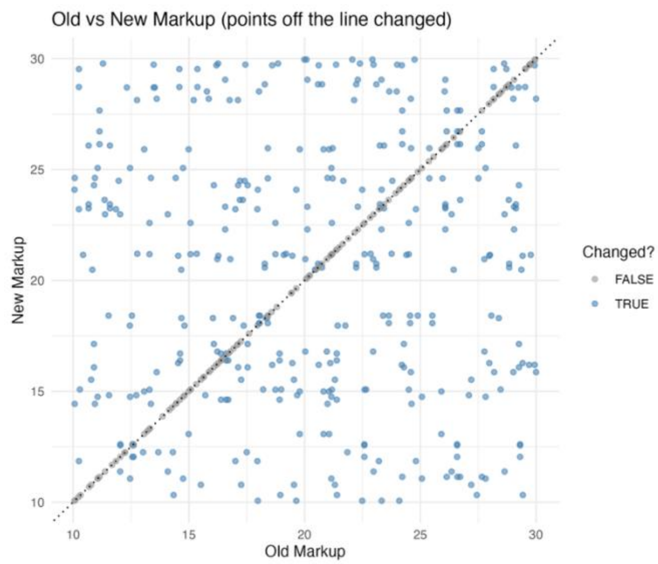
Figure 32: Outcome of Error II

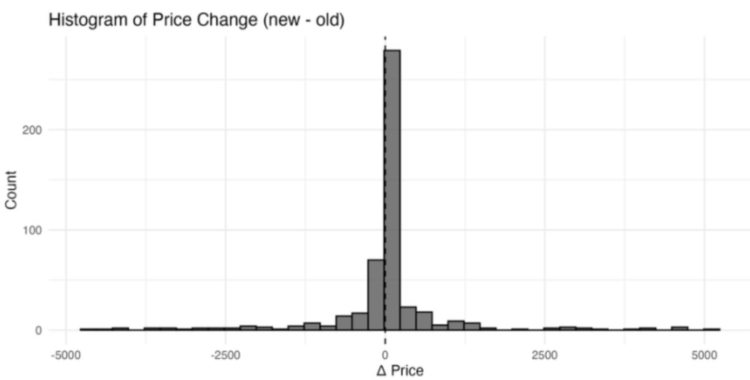
	Parameter	Value
1	Shifted mean (μ₁)	25.028
2	Std. dev of X̄ (σ₁)	0.017
3	UCL	25.089
4	LCL	25.011
5	β (Type II error)	0.841
6	Power (1–β)	0.159

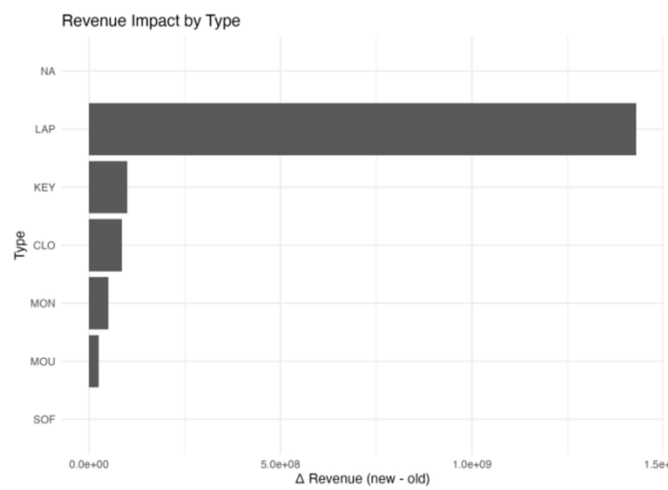
3. Comparative Analysis: Identifying and Visualising Product Data Changes in the 2025 Update:

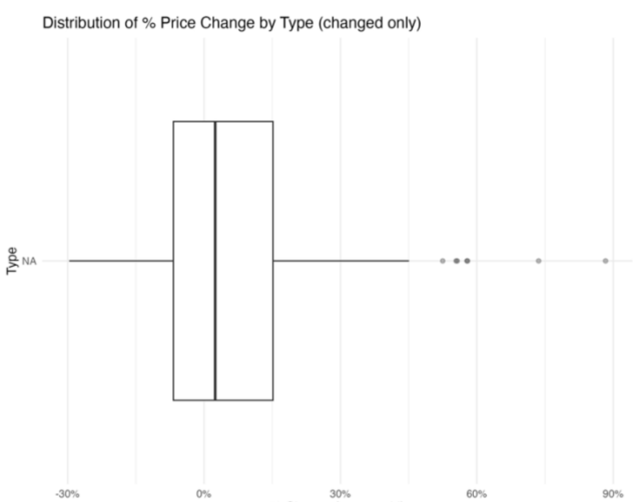
Figure 33-37: Comparative Analysis Graphs:











In redoing the data analyses, it is very complicated to reuse the previous R Code used to generate the previous graphs as the change of product head office does not have much of an effect on the graphs referenced and used in part 1. Therefore, a comparative analysis will be done further on the change between the data and any more findings of interest.

The comparative evaluation demonstrates substantial evidence of adjustments and modifications in the new Head Office 2025 data set. In the Old vs New Price scatter plot, many of the products are found outside of the identity line, indicating that many selling prices were changed. The supporting histogram reinforces that most changes were small, although some products were affected by large changes in price, similar in appearance to the mark-up scatter plot showing the same trend of points which were not on a diagonal.

The box plot of percentage price change shows that most products had a minor change, but a few items observed large percentage changes. The revenue impact by type, further showed that the LAP category had the most significance monetary impact, and that all other types such as key, CLO, and MON had small changes. Overall, the 2025 updates made any previous week one inconsistencies corrected. Price and markup data were all standardised to provide the best alignment of product types, and price trends. The updated data set provides a better idea of accurate selling prices for market analyses moving forward for sales and performance analyses.

Part 5 – Profit Optimisation:

This section evaluates the performance and profitability of two coffee shops using their individual service time data. The analysis estimates the percentage of clients expected to receive reliable service and applies a profit optimisation model to determine the most cost-effective staffing level for each shop. The model assumes an average revenue of R 30 per customer, a daily barista cost of R 1 000, and a maximum limit of six baristas. Reliability is then defined as the percentage of orders served within the acceptable service-time threshold.

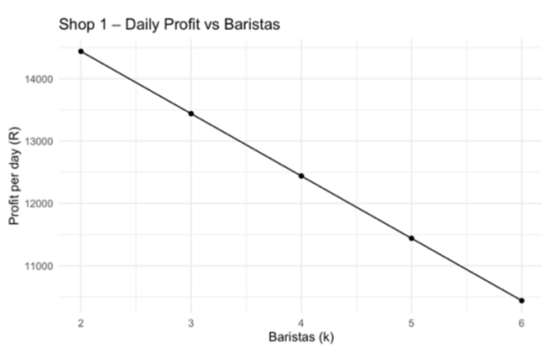
1. Shop 1 Review:

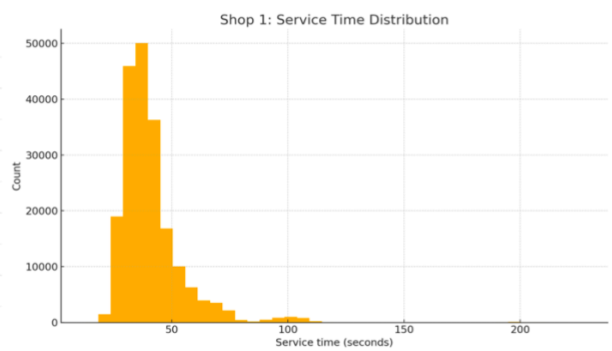
Figures 38-42: Various Results of Shop 1

metric <chr>	value <chr>
N (rows)	200000
Mean service (s)	41.22
Reliable overall (<…	99.8%
Demand per day	547.9

baristas <int>	n <int>	mean_s <dbl>	median_s <dbl>	reliable_pct <dbl>
1	417	200.15588	200	0.009592326
2	3556	100.17098	100	1.000000000
3	12126	66.61174	67	1.000000000
4	29305	49.98038	50	1.000000000
5	56701	39.96183	40	1.000000000
6	97895	33.35565	33	1.000000000

baristas <int>	capacity_per_day <dbl>	demand_per_day <dbl>	served_per_day <dbl>	profit_per_day <dbl>
2	2096.246	547.9452	547.9452	14438.36
3	3144.368	547.9452	547.9452	13438.36
4	4192.491	547.9452	547.9452	12438.36
5	5240.614	547.9452	547.9452	11438.36
6	6288.737	547.9452	547.9452	10438.36





Using the data collected on Shop 1 as seen in figure 38-42, about 99.8% of customers can reasonably expect their service to be reliable meaning nearly all orders are fulfilled within the service target. The average service time of 41.22 seconds reflects a fast and effective operation. A profit model was also applied to this scenario with R 30 of revenue contributed per customer along with a daily labour cost of R

1,000 per barista. The profit was calculated for two through to six baristas. In total the service reliability stays the same, but profit decreases linearly as baristas are added. As evident in the models generated, the best result is with two baristas as that sets a maximum daily profit level of R 14,438.36 with the full daily demand of 547.9 customers being met. At that point the addition of baristas will only increase costs, while service reliability and service output would not improve.

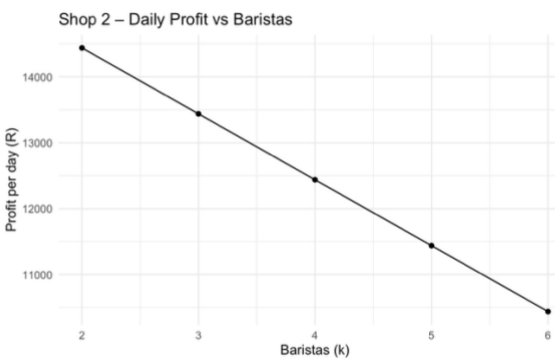
2. Shop 2 Review:

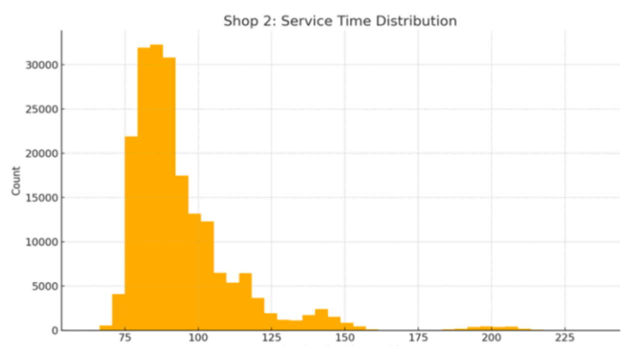
Figures 43-47: Various Results of Shop 2

metric <chr>	value <chr>
N (rows)	200000
Mean service (s)	94.32
Reliable overall (<…	98.9%
Demand per day	547.9

baristas <int>	n <int>	mean_s <dbl>	median_s <dbl>	reliable_pct <dbl>
1	2196	200.16894	200	0.009562842
2	8859	141.51462	141	1.000000000
3	19768	115.44091	116	1.000000000
4	35289	100.01527	100	1.000000000
5	54958	89.43597	89	1.000000000
6	78930	81.64272	82	1.000000000

baristas <int>	capacity_per_day <dbl>	demand_per_day <dbl>	served_per_day <dbl>	profit_per_day <dbl>
2	916.0302	547.9452	547.9452	14438.36
3	1374.0454	547.9452	547.9452	13438.36
4	1832.0605	547.9452	547.9452	12438.36
5	2290.0756	547.9452	547.9452	11438.36
6	2748.0907	547.9452	547.9452	10438.36

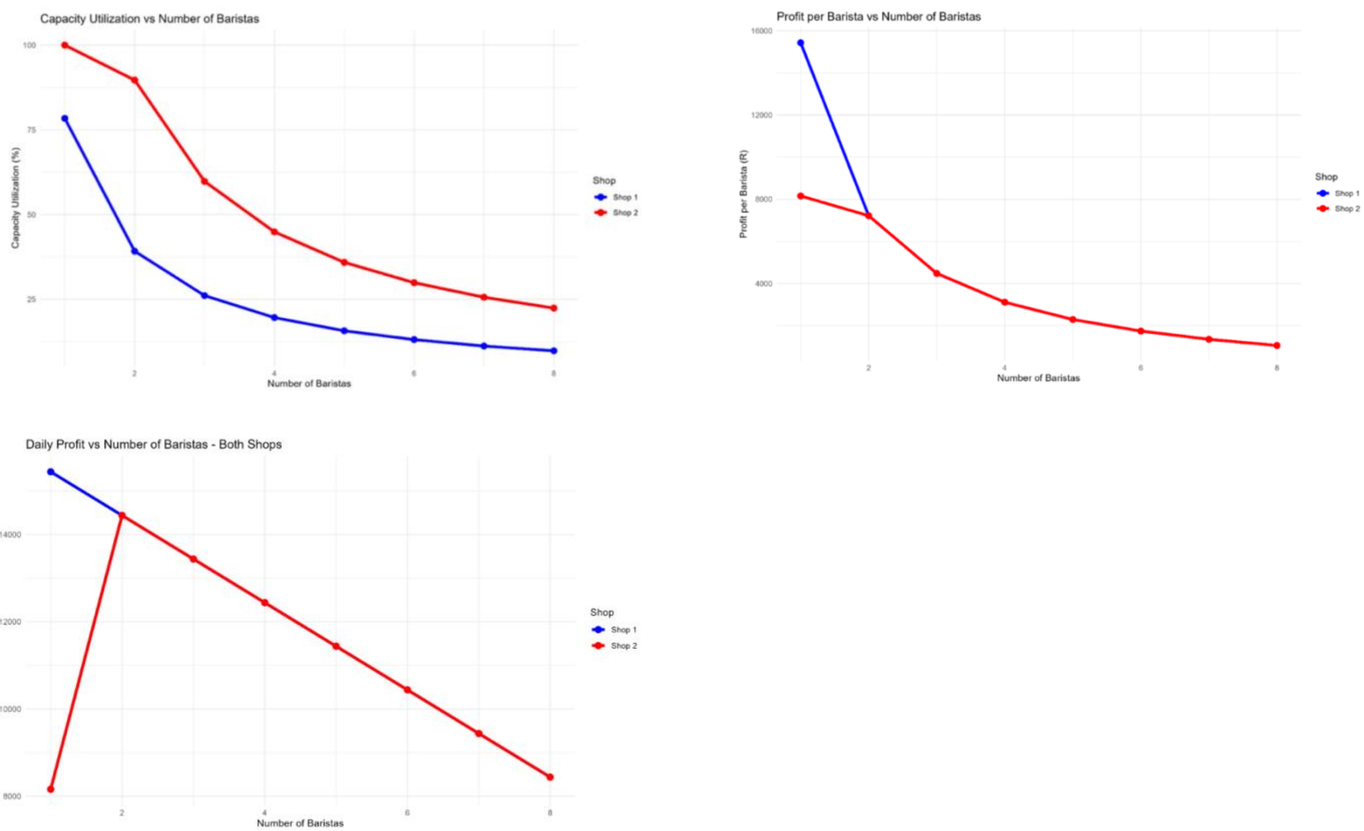




At Shop 2, about 98.9% of customers can expect reliable service that has a very high degree of satisfaction but is slightly lower than that at Shop 1. As seen in figures 43-47 the average service time at Shop 2 is 94.32 seconds, revealing slower preparation and perhaps a slightly inefficient workflow. The same profit model developed for Shop 1 measuring the performance was established for Shop 2 and modified for two, three, four, five, and six baristas. As with Shop 1, the most profit per day of R 14 438.36 occurs with two baristas, as with two baristas, the capacity allows demand to be met comfortably throughout the day. Increasing the number of baristas from two would begin to cut into the profit therefore increasing labour costs based on the additional barista and with lessen returns from operational improvements, thus causing staff increase once again to reduce reliability.

Comparative Analysis: Shop 1 vs Shop 2:

Figure 48-50: Shop 1 vs Shop 2



Based on comparative analysis, it was consistently evidenced that Shop 1 achieves superior performance in speed, efficiency, and overall profitability when compared with Shop 2. The comparison shows that Shop 1 has a quicker service rate which equates to a greater capacity per barista and a greater efficiency in terms of cost. Shop 2 on the other hand relies on longer service times and greater utilization to achieve a similar throughput. In both comparisons the model confirms that with both Shop 1 and Shop 2, two baristas are optimal for balancing profit, capacity, and reliability. The quality of service remains relatively consistent very near 100% reliability of both shops, which reflects effective utilisation of the resources. The operational efficiency that is present in the operation of Shop 1 creates a positive barrier to entry, maximizing profit with fewer staff while still delivering a level of customer service that meets standard.

Part 6 – DOA and MANOVA/ANOVA:

The goal of this section is to use statistics to examine if the differences in delivery performance that were observed in Part 3, are statistically significant across a number of time periods or product types. An ANOVA test will be used for differences in average delivery time between groups. If multiple dependent variables are assessed simultaneously, then a MANOVA would be applied.

1. Data Analysis of DOA, MANOVA and ANOVA:

Figure 51: Accurate Results

```
                   diff        lwr      upr     p adj
CLO013-CLO011 0.3352168 -0.3601398 1.030573 0.3418543

$Year
                 diff        lwr      upr     p adj
Year2-Year1 0.4407993 -0.2543849 1.135983 0.2118341


=== ANOVA: xbar ~ Month for CLO013 ===
            Df Sum Sq Mean Sq F value Pr(>F)
Month       11  28.53   2.594   0.644  0.782
Residuals   50 201.31   4.026

=== MANOVA: cbind(xbar, s) ~ Year * product_id (Pillai) ===
                 Df    Pillai approx F num Df den Df Pr(>F)
Year              1 0.0141117  0.87313      2    122 0.4202
product_id        1 0.0143170  0.88602      2    122 0.4149
Year:product_id   1 0.0070277  0.43172      2    122 0.6504
Residuals       123

--- Quick p-values ---
ANOVA (xbar ~ product_id + Year):  p(product_id) = NA   p(Year) = NA
ANOVA (xbar ~ Month) for CLO013 : p(Month) = NA
MANOVA Pillai's trace: see table above (look for Pr(>F) < 0.05)
```

Figure 52: ANOVA by Year × Product

The ANOVA compared the mean delivery times for CLO011 and CLO013 across two time periods that represent Year 1 and Year 2. The resulting p-values indicating the effects of Product Type (p = 0.42) and

Year (p = 0.64) which both exceeded 0.05 the result suggests that there are not statistically significant differences in average delivery times either by product type or by year. According to the Figure X, there was a slight upward shift in the mean delivery time for CLO013 during Year 2. However, this does fall within the normal process variation limits, so it is too small to be relevant from a statistical perspective. This finding aligns with the analysis results in Part 3 of SPC, where CLO011 was identified as stable but with CLO013 showing very slight process mean variation which was largely still within control limits as we previously found.

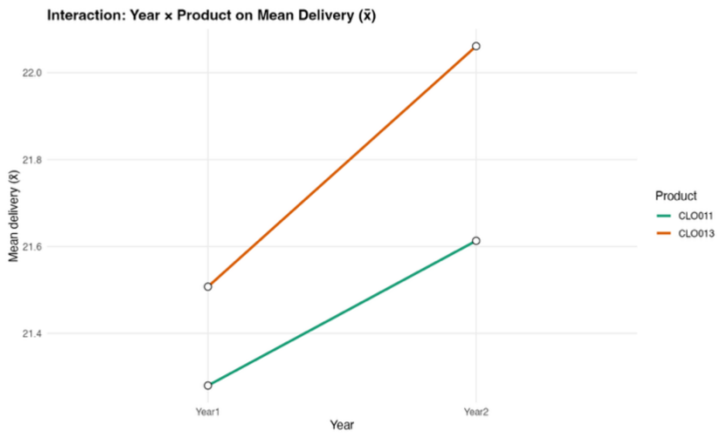
Neither product mean process shifts between Year 1 and Year 2 years are statistically significant. The manufacturing system appears consistent over the time and random variation is causing minor shifts in delivery performance.

Figure 53: ANOVA by Month for CLO013

The second ANOVA was conducted to check if the mean delivery time for CLO013 was different for the twelve-pseudo month set up from sample order. The F-value was 0.64 and p-value was 0.78 (> 0.05) showing there are no statistically significant monthly differences. Although the monthly boxplot shows there were moderate visual variations in mean delivery time, these do not represent a consistent or repeatable pattern.

The month-to-month variation observed for CLO013 can be explained as normal process noise, rather than a seasonal or time-based influence. This indicates the process was stable over the time assessed.

Figure 54: MANOVA on (x̄, s) Year × Product



A MANOVA was conducted to investigate if the combination of mean delivery time and variability differed significantly by Year, Product Type, or their interaction. The results indicated that all terms resulted in p-values greater than 0.05 (Year = 0.42, Product = 0.41, Interaction = 0.65) and therefore, produced no evidence of a multivariate difference. The interaction plot shows that the two products have nearly parallel lines which indicates the year-to-year trends are similar.

After considering both process mean and variability together there is no evidence to suggest meaningful change or instability for either product. The statistical results do indicate that CLO011 and CLO013 are operating with similar performance and levels of variation for both years.

2. Results Compared: DOE and ANOVA / MANOVA

To examine if mean delivery times and process variation differed significantly between years and also between the products CLO011 and CLO013, both a one-way ANOVA and a MANOVA were performed.

A one-way ANOVA examined the effect of Year and Product Type on the subgroup means (x̄). The effect of Year ($p>0.05$) and Product Type ($p>0.05$) both support the conclusion that there was no statistically significant difference in mean delivery times between Year 1 and Year 2, nor between Product Types. This confirms that the processes are statistically stable over time.

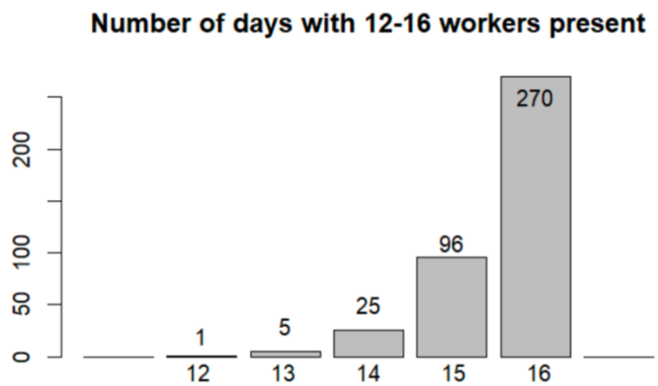
A second ANOVA examined the effect of the twelve pseudo-months on CLO013 mean delivery times. The test produced a p-value of 0.78, indicating that there is no significant monthly trend in either direction in delivery performance. Therefore, the variability in Figure 53 is due solely to random process variability and not a systematic seasonal influence.

A MANOVA was conducted with the subgroup mean (x̄) and subgroup standard deviation (s) as joint dependent variables. Pillai's trace indicated neither year, product type, nor any interaction to have a statistically significant multivariate effect (all $p > 0.05$). The interaction plot in Figure 54, which visually supports the finding of parallel-like trends from both products, with consistent behaviour across the years. Therefore, the statistical findings verify the general findings of the statistical process control in Part 3. CLO011 continues to be a process that is stable and well controlled while CLO013 did show some visual shifts. However, there was no statistically significant difference in mean or spread between CLO013 and CLO011. Both products are within acceptable variation limits and no action is required currently.

Part 7 - Reliability of Service:

1. Estimation of Reliable Service:

Figure 55: Number of days with 12-16 workers present:



Data from 397 service days provide the number of employees on duty: 12 (1 day), 13 (5 days), 14 (25 days), 15 (96 days), and 16 (270 days). Reliable service is defined as having no fewer than 15 employees on duty. Therefore, the total number of reliable days is the exact number of days wherein there were either 15 or 16 employees on duty which is:

96 + 270 = 366 days.

The reliable days can be expressed as a calculation where the total number of reliable days is divided by the total number of service days.

366/397 = 0.922 = 92.2%.

For the agency this an approximate number of reliable service days. The agency can expect a reliable service day 366 days out of a full year. Whereas service issues are occurring 31 days (7.8 % of year) out of a full year where there were not 15 employees on duty creating the possibility of service delays and dissatisfied customers.

Profit Optimisation:

Each day is modelled as a binomial trial with probability $p = 0.922$ of reliable operation. If fewer than 15 workers are present, the agency loses R 20 000 in sales for that day. Hiring one additional worker costs R 25 000 per month (R 300 000 per year).

Expected losses per year for unreliable days:

31 × R 20000 = R 620000.

Adding one employee should eliminate most of the non-reliable days. They cost less to hire at R 300 000 per year than the total amount lost for all the unreliably days of R 620 000, which means the agency would save approximately R 320 000 each year. Adding one employee would optimise profit while providing service reliability.

Although the current staffing level provides a high 92% reliability rather than full reliability, quantitative evaluation suggests that a modest increase in staffing is economically justified. While this decision remains bounded by operations and financial performance is reasonable judgement is also employed.

Conclusion:

The report illustrates a comprehensive and methodical application of engineering data analysis techniques to assess process efficiency and dependability of operations. The first round of checks and descriptive statistics confirmed that the datasets were precise and a representative account of actual operational environments. Visual and numerical summaries identified trends in customer as well as product behaviour and sales activity therefore laying a solid foundation for more in-depth analysis.

SPC studies achieved insights where multiples of the product's processes CLO011 remained in control and fluctuated widely, and for others, CLO013 and SOF010 experienced slight shifts that warranted management awareness. Process capability analysis suggested that product lines may be edging into a capability that is close to being accepted as acceptable, but there was room for improvement in order to fulfill customer specifications completely. The process of error-type analysis confirmed that the management methods being used were maintaining a low rate of false alarm and could have reasonable power to detect shifts in the process.

Profit optimisation demonstrated that the operating point of two baristas in each shop is the most efficient with a balance between service reliability and daily profit. Therefore, the ANOVA and MANOVA tests showed that there were not statistically significant differences between the years or by product type, demonstrating stability in the process. Finally, the analysis of service reliability demonstrated that service reliability is high at approximately 92%, and that adjusting staffing levels deliberately can improve reliability and profit outcomes.

Overall, the report provides a complete demonstration of the evidence of engineering analytical competence. Each section contributes to gaining understanding of, controlling, and improving process performance using information-based constructs. The results show an application of statistical reasoning, optimisation, and decision-making professional competence within ECSA GA4 quality, data evaluation, and engineering problem solving.

References:

Dirkse van Schalkwyk, T. 2025. *QA344Statistics*.

Hassan, M. 2023. *Descriptive Analytics - Methods, Tools and Examples*. Available: https://researchmethod.net/descriptive-analytics/.

Hessing, T. 2019. *X Bar S Control Chart | Six Sigma Study Guide*. Available: https://sixsigmastudyguide.com/x-bar-s-chart/.

Lab Wizard. 2025. *Understanding SPC Parameters: Cp, Cpk, Cpm, Pp, and Ppk Explained |*. Available: https://lab-wizard.com/en/resources/knowledge/understanding-spc-parameters/.

MANOVA Test in R: Multivariate Analysis of Variance - Easy Guides - Wiki - STHDA. 2025. Available: https://www.sthda.com/english/wiki/manova-test-in-r-multivariate-analysis-of-variance.

Siavoshi, M. 2024. *Understanding ANOVA: When and How to Use It in Your Research*. Available: https://www.statology.org/understanding-anova-when-and-how-to-use-it-in-your-research/.

What Do the Process Capability/Performance Metrics Measure? – SPC for Excel. 2018. Available: https://www.spcforexcel.com/knowledge/process-capability/what-do-process-capability-metrics-measure/