****

**Faculty of Information Science and Technology (FIST)**

**PROJECT FRONT PAGE AND DECLARATION**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Subject Code** | | **TBS2251** | | **Subject Name** | | **BUSINESS STATISTICAL ANALYSIS** | | |
| **Lecturers** | Dr Lew Sook Ling & Dr Siti Fatimah | | | | | **Term/Session** | 2 / 2022/2023 | |
| **Title of Project** | | | Analysis of manufacturing data | | | | | |
| **Lab Section** | | | 1DL | | **Group Number** | | | 5 |
| **Student Details** | | | | | **Define which sections,**  **pages or components you contributed to this Project** | | | Contribution (estimate  in % of 100) |
| **1. Leader** | **Student ID: 1191101340**  **Seat No:**  **Name: Grayson Goh Jin Yi**  **Programme: ST** | | | | - Conclusion  - Hypothesis Checking  - Datasets | | | 20% |
| **2. Member** | **Student ID: 1191101497**  **Seat No:**  **Name: Foo Haw Liang**  **Programme: ST** | | | |  | | |  |
| **3. Member** | **Student ID: 1211301430**  **Seat No:**  **Name: Lim Yu Xian**  **Programme:** | | | | - | | |  |
| **4. Member** | **Student ID: 1191100577**  **Seat No:**  **Name: Siah Kah Chuan**  **Programme: ST** | | | | * Introduction * Regression analysis between total planned production and total manpower needed | | | 20% |
| **5. Member** | **Student ID: 1201101507**  **Seat No:**  **Name: Teow Yong Wei**  **Programme:** | | | |  | | |  |
| Note: \*\*AI / BIA / DCN / ITM / ST etc. | | | | | Total Percentage | | | **100%** |

**Marking Form**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Report (70)** | **MARKS** | | **Presentation (30)** | **MARKS** | |
| Introduction |  | 5 | Clarity |  | 10 |
| Methodology |  | 20 | Presentation Slides |  | 10 |
| Data set |  | 10 | Teamwork |  | 10 |
| Data Analysis  Result |  | 25 |  |  |  |
| Conclusion |  | 5 |  |  |  |
| Format |  | 5 |  |  |  |
| **Total** |  | **70** | **Total** |  | **30** |

A picture containing text, font, graphics, logo

Description automatically generated

**Faculty of Information Science and Technology (FIST)**

***The following declaration must accompany all written work submitted.***

**Declaration**

* 1. I understand what plagiarism is and I am aware of the University’s policy in this regard.
  2. I declare that the work hereby submitted is my own original work. Where other people’s work has been used (either from a printed source, Internet or any other source), this has been properly acknowledged and referenced in accordance with the Multimedia University’s requirements.
  3. I have not used work previously produced by another student or any other person to hand in as my own.
  4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

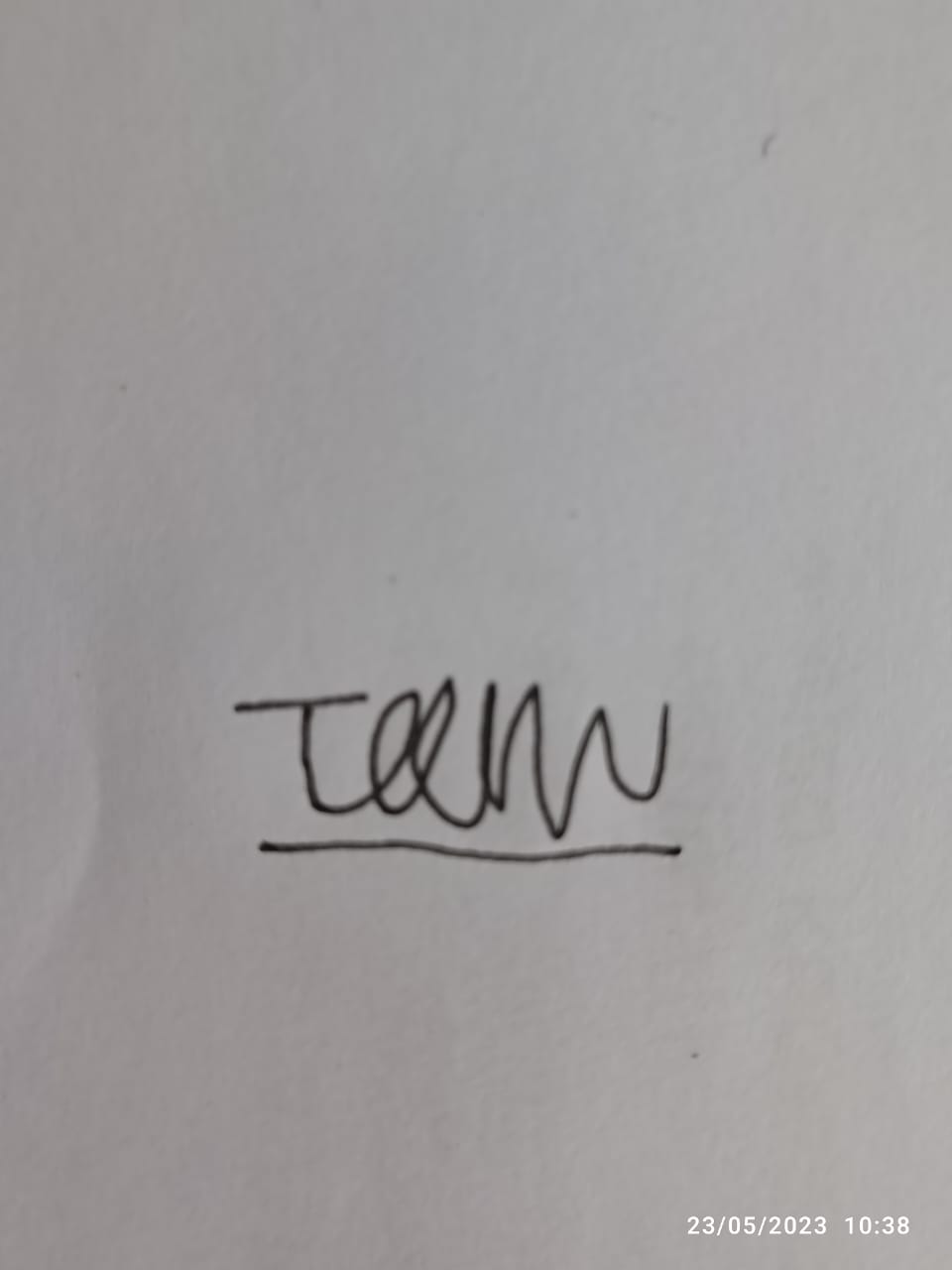
**A signature on a white background

Description automatically generated with medium confidence**

**SIGNATURE:**

**NAME: SIAH KAH CHUAN**

**DATE: 10/6/2023**

**SIGNATURE: **

**NAME: Teow Yong Wei**

**DATE: 13/6/2023**

**SIGNATURE: **

**NAME: Lim Yu Xian**

**DATE: 13/62023**

**SIGNATURE:**  

**NAME: Foo Haw Liang**

**DATE: 13/6/2023**

**SIGNATURE:** 

**NAME: Grayson Goh Jin Yi**

**DATE: 13/6/2023**

A picture containing text, font, graphics, logo

Description automatically generated

**Faculty of Information Science and Technology (FIST)**

**Marking Rubrics Report (25%)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Component** | **Sophisticated** | **Competent** | **Not Yet Complete** |
| **Introduction (5 Points)** | A good and well-written executive summary project. | An adequate introduction of the  project report. | Missing section to the written introduction project report. |
| **Methodology (20 Points)** | Systematic and clear explanation of the methodology is provided and related to the project. | Methodology of the project provided but not related to the project. | Missing section irrelevant methodology to provided. |
| **Data Set (10 Points)** | Comprehensive data set from a reliable source with explanation of the attributes and data file. | A data set from a reliable source but lacks depth and explanation of the attributes; and data file provided. file. | Data set from unknown source; no data and explanation provided. |
| **Data Analysis and Results (25 Points)** | Provide analysis and discuss the results using methods introduced in this course. | Provide analysis but lacks discussion on the findings/results. | Only descriptive statistics/charts provided with no analysis or discussion. |
| **Conclusions (5 Points)** | Able to conclude findings with supporting evidence. | Provide conclusions but lack supporting evidence. | Generally conclude the work done. |
| **Report Format (5 Points)** | Error-free formatting based on MMU guidelines. | Formatting based on MMU guidelines but has errors. | Report is not inappropriate format and have many errors. |

**Presentation (15%)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Component** | **Sophisticated (8-10)** | **Competent (5-7)** | **Not Yet Complete (1-4)** |
| **Content (10 Points)** | The presentation was interesting and with the required information on the work done. | The presentation includes all the important information on the project. | The presentation is not interesting and lacks information. |
| **Clarity Creativity (10 Points)** | The presentation is original, creative, clear and support the project. | The presentation is sufficient to present the main idea of the project. | The presentation lacks important information on the idea; further interrogations were required to understand the idea. |
| **Teamwork (10 Points)** | The presentation shows adequate planning and communication and teamwork among members. | The presentation was prepared according to plan with adequate communication and teamwork among members. | The presentation was dull and shows lack of planning and poor communication  and teamwork among members. |

**Table of Contents**

[1. Introduction 1](#_Toc137647826)

[2. Methodology 3](#_Toc137647827)

[2.1 Descriptive Analysis 3](#_Toc137647828)

[2.2 Hypothesis Testing 6](#_Toc137647829)

[2.3 Linear Regression 6](#_Toc137647830)

[2.4 Forecasting 7](#_Toc137647831)

[3. Data Sets 9](#_Toc137647832)

[3.1 Introduction to the datasets 9](#_Toc137647833)

[3.2 Raw Datasets 12](#_Toc137647834)

[4. Data Analysis and Result 14](#_Toc137647835)

[4.1 Regression analysis of total planned production and total manpower needed 14](#_Toc137647836)

[4.1.1 Dataset preparation 14](#_Toc137647837)

[4.1.2 Interpreting regression analysis 16](#_Toc137647838)

[4.1.3 Building model 17](#_Toc137647839)

[4.1.4 Residual plot 18](#_Toc137647840)

[4.1.5 Conclusion 19](#_Toc137647841)

[4.2 Performance Evaluation of Shipment Predictions Using Four-Week Moving Average Method 20](#_Toc137647842)

[4.2.1 Introduction 20](#_Toc137647843)

[4.2.2 105-02050 ORING, BUTTON 21](#_Toc137647844)

[4.2.3 105-04633 I-RING 23](#_Toc137647845)

[4.2.4 110-04674 Gasket 25](#_Toc137647846)

[4.2.5 105-04751 ORING, SMALL 27](#_Toc137647847)

[4.2.6 105-04752 ORING, LARGE 29](#_Toc137647848)

[4.2.7 105-07638 ORING, SEAL 31](#_Toc137647849)

[4.2.8 105-06116 ORING, MIC 33](#_Toc137647850)

[4.2.9 105-04802 SNOUT, COVER 35](#_Toc137647851)

[4.2.10 110-02888-N BUTTON AND 110-02898-T BUTTON, THIN 37](#_Toc137647852)

[4.3. Hypothesis checking about the average total planned production per week 39](#_Toc137647853)

[4.3.1 Dataset preparation 39](#_Toc137647854)

[4.3.2 Hypothesis Checking 40](#_Toc137647855)

[4.4 Regression analysis for Actual shipment 2021 vs 2022 vs 2023 43](#_Toc137647856)

[4.4.1 Original data 43](#_Toc137647857)

[4.4.2 Trendline analysis 44](#_Toc137647858)

[4.4.3 Result Analysis 45](#_Toc137647859)

[4.4.4 Result Conclusion 46](#_Toc137647860)

[5. Conclusion 47](#_Toc137647861)

[References 49](#_Toc137647862)

**List of tables**

[Table 3.1 Attribute table of Part Information & Process parameter 9](#_Toc137647872)

[Table 3.2 Attribute table of Customer forecast 10](#_Toc137647873)

[Table 3.3 Attribute table of Actual Shipment 2023 10](#_Toc137647874)

[Table 4.1 Total planned production Vs. Total manpower needed 14](#_Toc137647875)

[Table 4.2.1 105-02050 ORING, BUTTON: Demand vs. Shipment Error Analysis 21](#_Toc137647876)

[Table 4.2.2 105-04633 I-RING: Demand vs. Shipment Error Analysis 23](#_Toc137647877)

[Table 4.2.3 110-04674 Gasket: Demand vs. Shipment Error Analysis 25](#_Toc137647878)

[Table 4.2.4 105-04751 ORING, SMALL: Demand vs. Shipment Error Analysis 27](#_Toc137647879)

[Table 4.2.5 105-04752 ORING, LARGE: Demand vs. Shipment Error Analysis 29](#_Toc137647880)

[Table 4.2.6 105-07638 ORING, SEAL: Demand vs. Shipment Error Analysis 31](#_Toc137647881)

[Table 4.2.7 105-06116 ORING, MIC: Demand vs. Shipment Error Analysis 33](#_Toc137647882)

[Table 4.2.8 105-04802 SNOUT, COVER: Demand vs. Shipment Error Analysis 35](#_Toc137647883)

[Table 4.2.9 110-02888-N BUTTON AND 110-02898-T BUTTON, THIN: Demand vs. Shipment Error Analysis 37](#_Toc137647884)

[Table 4.3 Total planned production in each week 39](#_Toc137647885)

[Table 4.4 Shipment data for part number 105-02050 and 105-04633 over three years 43](#_Toc137647886)

**List of figures**

[Figure 2.1 Sample of Scatter Plot 4](#_Toc137647864)

[Figure 2.2 Sample of Line Chart 5](#_Toc137647865)

[Figure 4.1 Regression Summary Output 16](#_Toc137647866)

[Figure 4.2 Scatter plot between total planned production and total manpower needed 17](#_Toc137647867)

[Figure 4.3 Residual Output and plot 18](#_Toc137647868)

[Figure 4.4 Normal distribution of lower tailed rejection area 42](#_Toc137647869)

[Figure 4.5 Trendline for shipment data of part number 105-02050 over three years 44](#_Toc137647870)

[Figure 4.6 Trendline for shipment data of part number 105-04633 over three years 44](#_Toc137647871)

# 1. Introduction

Data analysis is an overall term that comprises of several phrases which includes, data collection, data pre-processing, data analysis, data visualization and finally result. Data analytics always start with a data collection process. In our case over here, we have gathered data sets that is related to the manufacturing process of a plastic & rubber factory through the year of 2023 and also the customer forecast data along with the actual shipping plan to predict the future shipping plan. These data sets comprise of several elements that are commonly seen in a factory, such as the production cycle time, production output rate, and other production & shipping details.

After that, we will perform simple data pre-processing to remove duplication and reorganize the data to make it looks more neatly. The purpose of doing so is to reduce the calculation complexity and overhead in the later data analysis stage. Unique part number is assigned to each parts to help us achieving the above purposes.

The next phrase would be to perform data analysis. This process serves as the core of the whole analysis process as this is where those unstructured data will be converted into useful information to help organizations drive innovative and improvement. There are a few methods fall under this stage, which are descriptive analysis, prescriptive analysis, predictive analysis regression and trendline, forecasting techniques and many more. However, due to the nature and constraint of collected data, the suitable method to be implemented in this project would be focusing on the regression analysis, hypothesis testing and forecasting techniques. At first, we will be applying descriptive analysis to compute the total number of tools, machine, and manpower needed in the manufacturing process. After that, we will use the total number of machines and manpower needed to generate a simple linear regression model to discover the relationship between machines and manpower needed. Besides, we also perform moving average onto the actual shipment data to predict near future shipment plan to ensure the shipping progress are on track. Apart from that, we also conduct hypothesis testing to discover whether the average total planned production is less than 30. Lastly, there is also a regression analysis conducted onto the actual shipment data over the years 2021, 2022, 2023 to find out variability between these three years. To get a clearer picture of the implementation of these methodologies, reader may refer to the data analysis section under this report.

The upcoming process would be the data visualization where we would convert the analyse data into varies illustration such as histogram, line chart, scatter chart instead of keeping them in decimal and text format which is difficult for viewers to extract the needed information from it.

The last phrase is to summarize findings from the computed information. As a result of analyse, we have included several recommendations to help managers carried out a better informed decision.

Throughout this project, we aim to achieve the objectives listed below:

1. To uncover the association between planned production against number of manpower needed (linear regression)
2. To look into the variability among the predicted shipping data versus the actual shipping data (moving average)
3. To investigate the hypothesis about the average total planned production per week (hypothesis testing)
4. To find out the relationships for the actual shipment over the years 2021, 2022, and 2023 (linear regression)

# 2. Methodology

## 2.1 Descriptive Analysis

Descriptive analysis is a sort of data analysis that helps to explain, show, or summarise data points in a constructive way so that patterns can develop that fulfil every criterion of the data. It is one of the most crucial steps in statistical data analysis. It provides you with a conclusion about the distribution of your data, assists you in detecting typos and outliers, and allows you to spot commonalities among variables, preparing you for additional statistical analysis.

Basically, descriptive analysis is able to bring the advantages listed below:

i) **Data Summarization:** We can distil massive volumes of data into digestible summaries using descriptive analysis. Descriptive analysis gives a clear picture of the dataset by computing metrics like mean, median, mode, range, and standard deviation, making it simpler to comprehend and express the most important parts of the data.

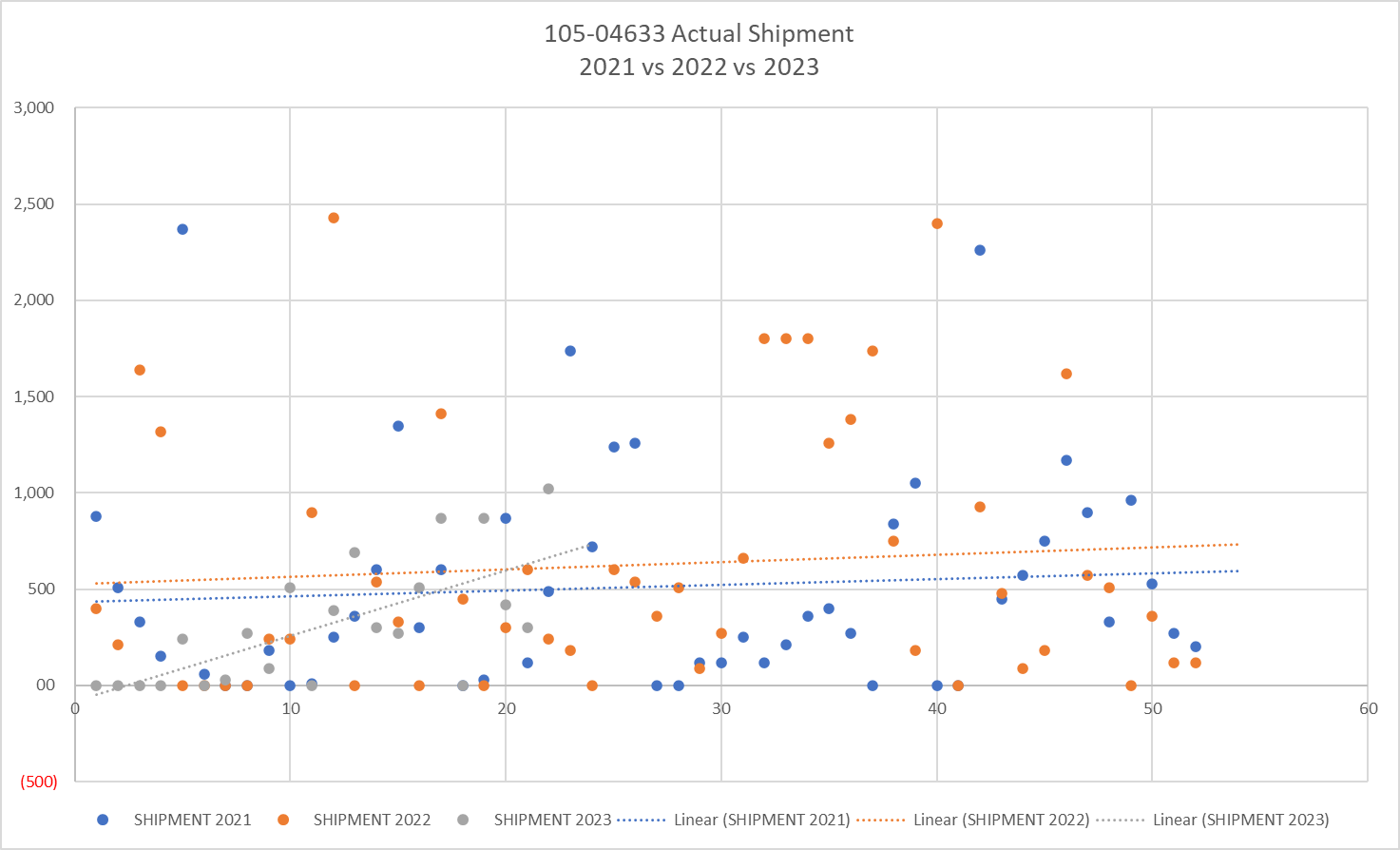
ii) **Data Exploration:** Exploring the dataset and gaining understanding of the underlying patterns and relationships is made easier by descriptive analysis. A preliminary knowledge of the dataset's structure and properties can be gained by descriptive analysis, which looks at the distribution of variables, spots outliers, and visualises the data using charts and graphs. The development of hypotheses can be guided by the results of this investigation.

iii) **Data Quality Assessment:** By using descriptive analysis, we can evaluate the accuracy and reliability of the data. Descriptive analysis assists in identifying data quality concerns that may affect the validity and reliability of later studies by looking for missing values, outliers, inconsistencies, or unexpected patterns. It enables data pretreatment and cleaning, making sure the data is ready for additional analysis.

iv) **Identifying Trends and Patterns:** Finding trends, patterns, and correlations within the dataset is made easier by descriptive analysis. Descriptive analysis offers insights into how variables are related to one another and how they vary over time or across different groups by looking at metrics like central tendency, variability, and correlation. This knowledge is useful for making choices and coming up with theories for additional investigation.

v) **Data Communication:** Effective data communication and presentation are made possible by descriptive analysis. Descriptive analysis facilitates the visual depiction of data by employing charts, graphs, and summary statistics, making it simpler to communicate information to stakeholders, clients, or coworkers. Data-driven storytelling and data-driven decision-making are both supported by descriptive analysis.

vi) **Benchmarking and Comparison:** Benchmarking and comparison of data across several time periods, groups, or factors are both possible with descriptive analysis. Descriptive analysis allows us to find variations, similarities, or contrasts and make accurate comparisons by comparing summary statistics, distributions, or graphical representations. For trend analysis, performance appraisal, and pinpointing areas for development, this information is helpful.



##### Figure 2.1 Sample of Scatter Plot

A picture containing line, plot

Description automatically generated

##### Figure 2.2 Sample of Line Chart

## 2.2 Hypothesis Testing

Hypothesis testing is an essential procedure in statistics. By using sample data, it is utilized to determine whether a hypothesis is tenable. The test offers proof that the hypothesis is plausible considering the available data. A hypothesis test determines which of two propositions about a population that are mutually exclusive is better supported by the sample data. A hypothesis test is what allows us to state that a result is statistically significant. Researchers can examine a hypothesis' plausibility using the technique known as hypothesis testing. It entails determining whether an assumption regarding a particular population parameter is accurate. Variance, standard deviation, and median are some examples of these population metrics. Testing hypotheses is important for assessing responses to queries involving data samples. A belief regarding a population parameter is referred to as a statistical hypothesis. This assumption might or might not be accurate. In other words, hypothesis testing is a legitimate method that scientists can use to confirm or disprove statistical theories. We carry out hypothesis testing to determine whether the total planned production will be more than the assumed value or not. At first, two types of hypotheses are defined, which is null hypothesis and alternate hypothesis. After that, we determine whether it is a two tailed test or one tailed test according to the symbol of alternate hypothesis. In the context of this report, we are testing for one tailed test in which standard deviation are not given. Thus, we will need to compute the mean, standard deviation, t-stat. and critical value with confidence interval of 95%. If t-stat fall under rejection area, we will reject null hypothesis, else failed to reject null hypothesis. For more information, readers may refer to the session 4.3 fall under this report.

## 2.3 Linear Regression

A method for quantifying the relationship between one or more predictor variables and a numeric response variable is linear regression. When the result is a continuous value and the relationship between the variables is reasonably linear, it is appropriate. On the basis of the predictor variable(s), it can be used to forecast the response variable. It can also be used to determine the association between two quantitative variables by performing simple linear regression. To perform linear regression in part 4.1 and 4.4, we will need to define the independent variable (x-axis) and dependent variable (y-axis). After that, we can simply apply the trendline tools onto the scatter plot to compute the linear equation and R2. Alternatively, we can also apply the analysis toolpak onto the two variables and wait for excel to the analysis job. There are a few measures that would be useful in the computed results. One of it would be the R2. This shows that how fits the data points fall under the trendline. A higher R2 indicates that there more points fall on the trendline. Another measures that is relatively useful would be p-value. This is to check whether a model is significant or not. If the computed p-value is managed to be smaller than 0.05, it means that the model is very significant and it can be used to make reliable informed decisions.

## 2.4 Forecasting

Forecasting is the practise of estimating what will happen in the future based on past and present events. Essentially, it is a decision-making tool that assists firms in dealing with the impact of future uncertainty by reviewing historical data and trends. It is a planning tool that allows firms to chart their next steps and develop budgets that will ideally cover any uncertainties that may arise.

In general, forecasting is able to help business in terms of:

i) **Business Planning and Decision Making**: By anticipating future values or trends based on historical data, forecasting assists organisations in making future plans. It aids in making knowledgeable choices on the allocation of resources, budgeting, scheduling of production, keeping track of inventories, and predicting sales. Businesses may optimize operations, reduce risks, and increase profitability with the help of accurate projections.

ii) **Demand and Supply Management:** In many businesses, forecasting is crucial for managing supply and demand. Businesses can modify their production, inventory, and supply chain strategies as necessary by projecting future demand. This helps with stockout avoidance, inventory level optimization, and customer happiness. Forecasting also helps with resource allocation based on anticipated demand, labour management, and capacity planning.

iii) **Financial Planning and Budgeting**: For financial planning, budgeting, and decision-making, forecasting is essential. It enables organisations to create practical financial goals, allocate resources wisely, and plan investments by assisting them in estimating future revenues, costs, and cash flows. Informed financial decision-making, increased financial performance, and better risk management are all made possible by accurate financial predictions.

iv) **Market Analysis and Competitive Intelligence:** Forecasting offers information on market trends, consumer behaviour, and market competition. It assists companies in analysing consumer demand, spotting new trends, and modifying their tactics to stay competitive. Forecasting is also useful for estimating market share, gauging the performance of marketing initiatives, and analysing the potential of new goods and services.

v) **Resource Optimization and Efficiency:** Forecasting aids in resource allocation optimization and boosts operational effectiveness. Organisations can efficiently plan production schedules, manage inventory levels, enhance staffing, and allocate resources by forecasting future demand or resource needs. This lowers waste, lowers expenses, and boosts overall productivity.

vi) **Risk Assessment and Mitigation:** Forecasting helps with risk assessment and mitigation for a variety of issues, including supply chain interruptions, market volatility, and economic situations. Businesses can create backup plans, assess their risk exposure, and take proactive steps to reduce risks by foreseeing future scenarios. As a result, losses are reduced and company continuity is preserved.

vii) **Policy and Planning:** Long-term planning and government policy both heavily rely on forecasting. It assists in forecasting elements that affect policy choices, such as population increase, infrastructure needs, energy consumption, environmental effects, and other variables. Urban planning, transportation planning, environmental management, and other fields that depend on long-term estimates, can benefit from forecasting.

In order to perform forecasting and analysis, we will need to get ready the data to be forecasted in the table form. Later on, we will need to decide how many historical data should be used to forecast the future data. In the framework of this project, we are using past 4 historical shipment data to compute the upcoming shipment data. Thenceforth, we will need to check whether the forecasted result is accurate or not by determining the error rate which is the difference between actual data and forecasted data. MAD, MSE and MAPE is the typical measurement that is used to determine the accuracy of forecasting result easier. The lower the value of these measurement leads to a more accurate forecasted result.

# 3. Data Sets

## 3.1 Introduction to the datasets

The original dataset we use is consist of 3 parts, which is **the Part information & process parameter** and the **customer forecast** and **the Actual Shipment 2023 (up to date until May)**. The part information & process parameter store all the data of part information from the factory. Each attribute provides specific information about the part, including its characteristics, manufacturing details, and associated costs and prices. We also perform the pre-processing by adding the additional information of working hour and use it to calculate the **output per week** for the part and add a new column for output per week in the dataset.

In the part of customer forecast data, it refers to the anticipated demand or projection of future sales or requirements made by a customer or group of customers for a particular product or service. It is an estimate provided by the customer based on their analysis, market trends, historical data, and other factors that influence their purchasing patterns. The last part of the dataset is Actual shipment 2023 which provided a shipment record for different categories of products along with their prices, part numbers, and monthly quantities. The data only goes up to the month of May.

The below table briefly describe each attribute of 3 part of the datasets:

#### Table 3.1 Attribute table of Part Information & Process parameter

|  |  |
| --- | --- |
| Attribute Name | Description |
| Category | It indicates the material category of the part, whether it is made of rubber or plastic. |
| Part Number | This is a unique identifier assigned to each part for identification purposes |
| Description | It provides a brief description or name of the part. |
| Customer | It specifies the customer or client associated with the part. |
| Selling Price (USD) | This represents the price at which the part is sold to the customer, denoted in US dollars. |
| Cost (USD) | It denotes the cost associated with manufacturing the part, expressed in US dollars |
| Compression Machine | It refers to the type or capacity of the compression machine used for manufacturing the part. |
| Cycle Time (sec) | This indicates the time required for one complete cycle of manufacturing the part, measured in seconds. |
| Tool Cavity | It represents the number of cavities in the tool used for manufacturing the part. |
| Output per hour | This is the number of parts produced per hour. |
| Yield Rate | It specifies the percentage of parts that meet the desired quality standards during the manufacturing process. |
| Output per week: | This represents the total number of parts produced per week. |

#### Table 3.2 Attribute table of Customer forecast

|  |  |
| --- | --- |
| Attribute Name | Description |
| Category | It indicates the material category of the part, whether it is made of rubber or plastic. |
| Price | This represents the price at which the part is sold to the customer |
| Part Number | This is a unique identifier assigned to each part for identification purposes |
| Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sept, Oct, Nov, Dec | All these columns represent the months of the year. The values under each month column indicate the quantity of the corresponding part number that the customer has forecasted for that specific month. |

#### Table 3.3 Attribute table of Actual Shipment 2023

|  |  |
| --- | --- |
| Attribute Name | Description |
| Category | It indicates the material category of the part, whether it is made of rubber or plastic. |
| Price | This represents the price at which the part is sold to the customer |
| Part Number | This is a unique identifier assigned to each part for identification purposes |
| Jan, Feb, Mar, Apr, May | These columns represent the monthly shipments for each part number and category. The numbers in each cell indicate the quantity of products shipped in a particular month. A hyphen "-" denotes that no shipments were made for that specific month and part number. |

## 3.2 Raw Datasets

**Process Parameter & Part Information**



**Customer Forecast**









**Actual Shipment – 2023**





# 4. Data Analysis and Result

## 4.1 Regression analysis of total planned production and total manpower needed

### 4.1.1 Dataset preparation

By applying count function onto each of the individual parts in the “process parameter” worksheet, we are able to get the number of planned production and amount of manpower for each part in a particular week. After that we are going to using sum function to sum up the planned production for all the 10 parts to come out with a total planned production. Similar process is applied to manpower to get the total manpower needed as shown below.

#### Table 4.1 Total planned production Vs. Total manpower needed

|  |  |  |
| --- | --- | --- |
| **Week** | **Total Planned Production (x-axis)** | **Total Manpower Needed (y-axis)** |
| 1 | 39 | 154 |
| 2 | 29 | 110 |
| 3 | 24 | 89 |
| 4 | 22 | 83 |
| 5 | 22 | 83 |
| 6 | 21 | 75 |
| 7 | 18 | 64 |
| 8 | 16 | 58 |
| 9 | 18 | 74 |
| 10 | 17 | 64 |
| 11 | 16 | 59 |
| 12 | 21 | 72 |
| 13 | 24 | 83 |
| 14 | 22 | 77 |
| 15 | 22 | 76 |
| 16 | 23 | 77 |
| 17 | 23 | 80 |
| 18 | 22 | 75 |
| 19 | 22 | 75 |
| 20 | 20 | 67 |
| 21 | 19 | 70 |
| 22 | 21 | 79 |
| 23 | 22 | 82 |
| 24 | 23 | 84 |
| 25 | 23 | 83 |
| 26 | 34 | 113 |
| 27 | 38 | 141 |
| 28 | 30 | 98 |
| 29 | 31 | 99 |
| 30 | 37 | 120 |
| 31 | 36 | 121 |
| 32 | 35 | 120 |
| 33 | 36 | 122 |
| 34 | 36 | 122 |
| 35 | 31 | 97 |
| 36 | 30 | 86 |
| 37 | 29 | 84 |
| 38 | 38 | 130 |
| 39 | 47 | 161 |
| 40 | 54 | 204 |
| 41 | 55 | 206 |
| 42 | 51 | 194 |
| 43 | 50 | 186 |
| 44 | 46 | 180 |
| 45 | 52 | 187 |
| 46 | 57 | 212 |
| 47 | 51 | 187 |
| 48 | 49 | 184 |
| 49 | 40 | 147 |
| 50 | 31 | 120 |
| 51 | 39 | 141 |
| 52 | 37 | 143 |

After we managed to compile the raw data into a more organized data as shown in the table above, we are going to perform simple linear regression analysis onto the table above by using total planned production as independent variables (x-axis) and using total manpower needed as dependent variable (y-axis) to uncover relationship between total planned production and total manpower needed. In other words, we are going to determine how is the effect of total planned production onto the total manpower needed.

### 4.1.2 Interpreting regression analysis

A picture containing text, screenshot, line, font

Description automatically generated

##### Figure 4.1 Regression Summary Output

By using “analytics toolpak” tools in excel, we are able to get to the summary shown above. From the result, it provides lots of valuable information for us to make informed decisions. Multiple R with a value of 0.9851 with is very close to 1 indicate us that this model has a strong linear relationship. R square of 0.9704 indicates that 97.04% of variability in total manpower needed can be explained by the total planned production. As we are performing simple linear over here, adjusted R square would not play much role in our case. Standard error tells us that the average distance that the data points fall from regression line is 7.8478. As the manpower needed data is close to around 100 something, this small deviation would not affect much on the accuracy of data.

After that, we can conduct hypothesis testing to check whether the total planned production will affecting the total manpower needed. Hypothesis to be used in the experiment are shown below:

*H0: The total planned production does not affect the total manpower needed.*

*H1: The total planned production* *affects the total manpower needed*

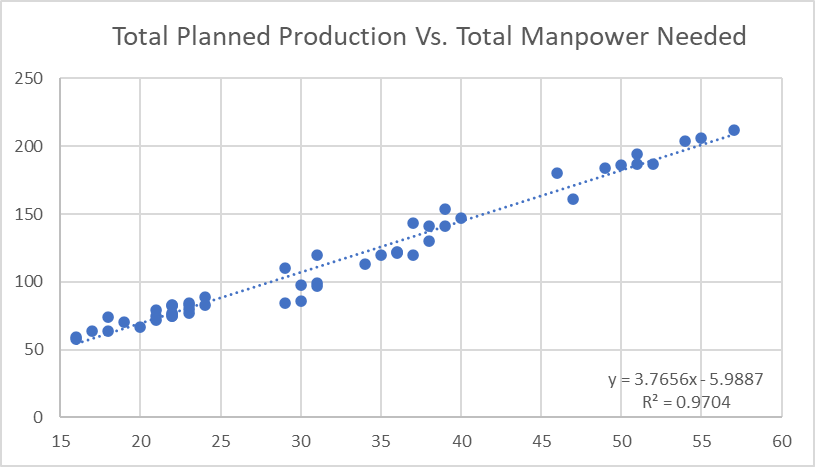
The p-value of value 6.88e-40 in the highlighted cell gives us strong evidence to reject the null hypothesis, which means total planned production will affects the total manpower needed. At the same time, it also tells us that this is a signification model as the p value is less than 0.05. The confidence interval of 95% in this model leads us to have a range of [3.58, 3.95]. Since the coefficient of total planned production which is 3.77 fall under this range, it further prove the significance of this model.

### 4.1.3 Building model

By referring to the coefficient column in the regression analysis, we could form a regression equation as shown below:

Y = 3.77(x) - 5.99, where x = total planned production.

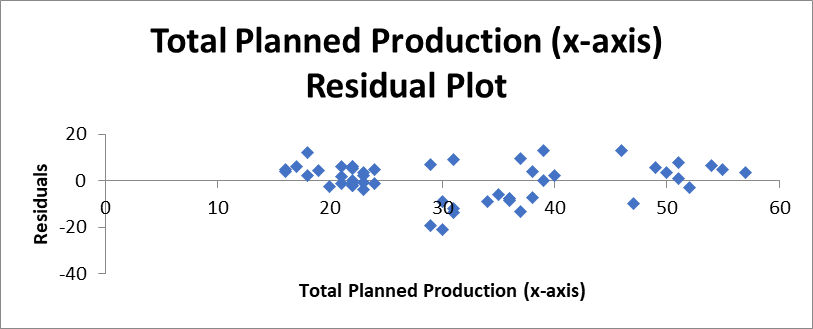
Alternatively, we can use scatter plot to get the similar equation.



##### Figure 4.2 Scatter plot between total planned production and total manpower needed

### 4.1.4 Residual plot





##### Figure 4.3 Residual Output and plot

By observing the standard residuals data above, we notice that all of the data above are fall within the range of [-3,3], thus we can conclude that there are no potential outliers exist in this model. The data in residual plot appears to be random and it shows no serious difference in the spread of the data for different X values. Therefore, it is managed to fulfil the assumption of linear regression.

### 4.1.5 Conclusion

This experiment shows that as total planned production increases, the total manpower needed increase as well. From the result, it shows that when total planned production increase by 1 unit, the total manpower will increase by approximately about 3.77, which round up to be 4 unit as manpower data cannot keep in continuous form. The significant of this model tells us that all the computed result are reliable, thus we can safely make informed decision based on this model.

## 4.2 Performance Evaluation of Shipment Predictions Using Four-Week Moving Average Method

### 4.2.1 Introduction

The tables below show the demand and shipment data for a product, as well as the forecast accuracy and error metrics. The forecast is based on a four-week moving average of the actual shipment data, which is a simple method that smooths out the fluctuations in the data. The forecast accuracy is measured by the standard error, which is the standard deviation of the errors between the forecast and the actual shipment. The error is the difference between the demand and the forecast, which can be positive or negative. The absolute error is the magnitude of the error, regardless of its sign. The squared error is the error raised to the power of two, which penalizes large errors more than small errors. The percentage error is the ratio of the absolute error to the demand, expressed as a percentage.

The dataset shows that there is no shipment data for the product in the first four weeks of the year. This could be due to various reasons, such as:

* Production or delivery delays: There might have been some technical issues, quality problems, supplier shortages, or other factors that prevented the product from being produced or delivered on time. This could have caused a backlog of orders or a loss of customers.
* Demand and supply mismatch: There might have been a discrepancy between the demand forecast and the production plan, which resulted in either excess inventory or stockouts. This could have affected the cash flow or customer satisfaction.
* Market changes: There might have been a change in the customer preferences or market conditions, such as new competitors, price fluctuations, or regulatory changes. This could have reduced the demand for the product or increased the costs of production or distribution.
* Data errors: There might have been an error in the data entry or reporting, which caused the shipment data to be missing or inaccurate. This could have distorted the performance evaluation or decision making.

However, since we obtained this dataset online, we do not have access to more information about the production and distribution process. Therefore, it is hard to determine the exact cause of the problem with certainty.

### 4.2.2 105-02050 ORING, BUTTON

#### Table 4.2.1 105-02050 ORING, BUTTON: Demand vs. Shipment Error Analysis

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | **MAD** | **MSE** | **MAPE** |
| **Week Number** | **DEMAND** | **SHIPMENT** | **4 MOVING AVERAGE** | **STANDARD ERROR** | **ERROR** | **|ERROR|** | **Error^2** | **|%Error|** |
| 1 | 870 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 2 | 671 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 3 | 627 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 4 | 604 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 5 | 498 | 240 | 00 | #N/A | 240 | 240 | 57,600,000 | 100% |
| 6 | 410 | 00 | 60 | #N/A | (60) | 60 | 3,600,000 | #DIV/0! |
| 7 | 387 | 00 | 60 | #N/A | (60) | 60 | 3,600,000 | #DIV/0! |
| 8 | 366 | 300 | 60 | 175 | 240 | 240 | 57,600,000 | 80% |
| 9 | 344 | 00 | 135 | 144 | (135) | 135 | 18,225,000 | #DIV/0! |
| 10 | 344 | 720 | 75 | 352 | 645 | 645 | 416,025,000 | 90% |
| 11 | 339 | 270 | 255 | 351 | 15 | 15 | 225,000 | 6% |
| 12 | 393 | 1,170 | 323 | 537 | 848 | 848 | 718,256,250 | 72% |
| 13 | 321 | 690 | 540 | 538 | 150 | 150 | 22,500,000 | 22% |
| 14 | 509 | 00 | 713 | 559 | (713) | 713 | 507656250000.00 | #DIV/0! |
| 15 | 509 | 210 | 533 | 581 | (323) | 323 | 104,006,250 | 154% |
| 16 | 507 | 600 | 518 | 400 | 83 | 83 | 6,806,250 | 14% |
| 17 | 507 | 210 | 375 | 402 | (165) | 165 | 27,225,000 | 79% |
| 18 | 498 | 00 | 255 | 225 | (255) | 255 | 65,025,000 | #DIV/0! |
| 19 | 498 | 1,320 | 255 | 555 | 1,065 | 1,065 | 1,134,225,000 | 81% |
| 20 | 399 | 570 | 533 | 554 | 38 | 38 | 1,406,250 | 7% |
| 21 | 422 | 185 | 525 | 574 | (340) | 340 | 115,600,000 | 184% |
| 22 | 502 | 788 | 519 | 575 | 269 | 269 | 72,495,563 | 34% |
|  |  |  |  |  |  |  |  | 51% |

The table reveals some insights about the performance of the forecast and the shipment. First, the forecast accuracy is low, as indicated by the large standard errors in most weeks. This means that the forecast deviates significantly from the actual shipment data. Second, the errors between the demand and the forecast are also large, especially in weeks 10, 12, 14, 15, 17, 19 and 21. This means that the forecast fails to capture the variations in the demand. Third, the percentage error shows that there are some weeks where the forecast overestimates or underestimates the demand by more than 50%, such as weeks 5, 8, 10, 12, 14, 15, 17 and 21. This means that the forecast is not reliable for planning purposes. The average percentage error is 51%, which is very high.

### 4.2.3 105-04633 I-RING

#### Table 4.2.2 105-04633 I-RING: Demand vs. Shipment Error Analysis

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | **MAD** | **MSE** | **MAPE** |
| **Week Number** | **DEMAND** | **SHIPMENT** | **4 MOVING AVERAGE** | **STANDARD ERROR** | **ERROR** | **|ERROR|** | **Error^2** | **|%Error|** |
| 1 | 914 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 2 | 705 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 3 | 662 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 4 | 645 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 5 | 539 | 240 | 00 | #N/A | 240 | 240 | 57,600,000 | 100% |
| 6 | 444 | 00 | 60 | #N/A | (60) | 60 | 3,600,000 | #DIV/0! |
| 7 | 418 | 30 | 60 | #N/A | (30) | 30 | 900,000 | 100% |
| 8 | 396 | 270 | 68 | 97 | 203 | 203 | 41,006,250 | 75% |
| 9 | 377 | 90 | 135 | 118 | (45) | 45 | 2,025,000 | 50% |
| 10 | 379 | 510 | 98 | 76 | 413 | 413 | 170,156,250 | 81% |
| 11 | 373 | 00 | 225 | 159 | (225) | 225 | 50,625,000 | #DIV/0! |
| 12 | 434 | 390 | 218 | 192 | 173 | 173 | 29,756,250 | 44% |
| 13 | 421 | 690 | 248 | 193 | 443 | 443 | 195,806,250 | 64% |
| 14 | 551 | 300 | 398 | 242 | (98) | 98 | 9,506,250 | 33% |
| 15 | 551 | 270 | 345 | 197 | (75) | 75 | 5,625,000 | 28% |
| 16 | 549 | 510 | 413 | 179 | 98 | 98 | 9,506,250 | 19% |
| 17 | 549 | 870 | 443 | 168 | 428 | 428 | 182,756,250 | 49% |
| 18 | 540 | 00 | 488 | 208 | (488) | 488 | 237,656,250 | #DIV/0! |
| 19 | 540 | 870 | 413 | 292 | 458 | 458 | 209,306,250 | 53% |
| 20 | 355 | 420 | 563 | 322 | (143) | 143 | 20,306,250 | 34% |
| 21 | 376 | 300 | 540 | 326 | (240) | 240 | 57,600,000 | 80% |
| 22 | 446 | 1,020 | 398 | 269 | 623 | 623 | 387,506,250 | 61% |
|  |  |  |  |  |  |  |  | 48% |

The table reveals some insights about the performance of the forecast and the shipment. First, the forecast accuracy is low, as indicated by the large standard errors in most weeks. This means that the forecast deviates significantly from the actual shipment data. Second, the errors between the demand and the forecast are also large, especially in weeks 8, 10, 13, 17, 19 and 22. This means that the forecast fails to capture the variations in the demand. Third, the percentage error shows that there are some weeks where the forecast overestimates or underestimates the demand by more than 50%, such as weeks 5, 7, 8, 10, 11, 13, 17, 18, 19 and 21. This means that the forecast is not reliable for planning purposes. The average percentage error is 48%, which is very high.

### 4.2.4 110-04674 Gasket

#### Table 4.2.3 110-04674 Gasket: Demand vs. Shipment Error Analysis

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | **MAD** | **MSE** | **MAPE** |
| **Week Number** | **DEMAND** | **SHIPMENT** | **4 MOVING AVERAGE** | **STANDARD ERROR** | **ERROR** | **|ERROR|** | **Error^2** | **|%Error|** |
| 1 | 153 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 2 | 121 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 3 | 124 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 4 | 145 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 5 | 147 | 252 | 00 | #N/A | 252 | 252 | 63,504,000 | 100% |
| 6 | 116 | 00 | 63 | #N/A | (63) | 63 | 3,969,000 | #DIV/0! |
| 7 | 108 | 00 | 63 | #N/A | (63) | 63 | 3,969,000 | #DIV/0! |
| 8 | 106 | 270 | 63 | 104 | 207 | 207 | 42,849,000 | 77% |
| 9 | 233 | 36 | 131 | 126 | (94) | 94 | 8,900,036 | 261% |
| 10 | 241 | 420 | 77 | 85 | 343 | 343 | 117,964,772 | 82% |
| 11 | 238 | 00 | 182 | 143 | (182) | 182 | 32,956,772 | #DIV/0! |
| 12 | 289 | 234 | 182 | 167 | 52 | 52 | 2,752,052 | 22% |
| 13 | 202 | 612 | 173 | 154 | 439 | 439 | 193,125,092 | 72% |
| 14 | 293 | 00 | 317 | 213 | (317) | 317 | 100,172,250 | #DIV/0! |
| 15 | 293 | 00 | 212 | 205 | (212) | 212 | 44,732,250 | #DIV/0! |
| 16 | 293 | 504 | 212 | 212 | 293 | 293 | 85,556,250 | 58% |
| 17 | 293 | 162 | 279 | 238 | (117) | 117 | 13,689,000 | 72% |
| 18 | 293 | 00 | 167 | 187 | (167) | 167 | 27,722,250 | #DIV/0! |
| 19 | 293 | 630 | 167 | 175 | 464 | 464 | 214,832,250 | 74% |
| 20 | 182 | 108 | 324 | 207 | (216) | 216 | 46,656,000 | 200% |
| 21 | 202 | 00 | 225 | 184 | (225) | 225 | 50,625,000 | #DIV/0! |
| 22 | 217 | 648 | 185 | 206 | 464 | 464 | 214,832,250 | 72% |
|  |  |  |  |  |  |  |  | 60% |

The table reveals some insights about the performance of the forecast and the shipment. First, the forecast accuracy is low, as indicated by the large standard errors in most weeks. This means that the forecast deviates significantly from the actual shipment data. Second, the errors between the demand and the forecast are also large, especially in weeks 5, 8, 9, 10, 13, 14, 15, 16, 17, 19 and 22. This means that the forecast fails to capture the variations in the demand. Third, the percentage error shows that there are some weeks where the forecast overestimates or underestimates the demand by more than 50%, such as weeks 5, 7, 8, 9, 10, 11, 14, 15, 18, 19 and 20. This means that the forecast is not reliable for planning purposes. The average percentage error is 60%, which is very high.

### 4.2.5 105-04751 ORING, SMALL

#### Table 4.2.4 105-04751 ORING, SMALL: Demand vs. Shipment Error Analysis

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | **MAD** | **MSE** | **MAPE** |
| **Week Number** | **DEMAND** | **SHIPMENT** | **4 MOVING AVERAGE** | **STANDARD ERROR** | **ERROR** | **|ERROR|** | **Error^2** | **|%Error|** |
| 1 | 162 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 2 | 138 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 3 | 118 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 4 | 173 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 5 | 186 | 195 | 00 | #N/A | 195 | 195 | 38,025,000 | 100% |
| 6 | 157 | 15 | 49 | #N/A | (34) | 34 | 1,139,063 | 225% |
| 7 | 138 | 00 | 53 | #N/A | (53) | 53 | 2,756,250 | #DIV/0! |
| 8 | 145 | 180 | 53 | 80 | 128 | 128 | 16,256,250 | 71% |
| 9 | 150 | 30 | 98 | 90 | (68) | 68 | 4,556,250 | 225% |
| 10 | 156 | 150 | 56 | 54 | 94 | 94 | 8,789,063 | 63% |
| 11 | 147 | 00 | 90 | 59 | (90) | 90 | 8,100,000 | #DIV/0! |
| 12 | 205 | 210 | 90 | 69 | 120 | 120 | 14,400,000 | 57% |
| 13 | 167 | 495 | 98 | 79 | 398 | 398 | 158,006,250 | 80% |
| 14 | 157 | 00 | 214 | 161 | (214) | 214 | 45,689,063 | #DIV/0! |
| 15 | 157 | 00 | 176 | 181 | (176) | 176 | 31,064,063 | #DIV/0! |
| 16 | 157 | 330 | 176 | 196 | 154 | 154 | 23,639,063 | 47% |
| 17 | 157 | 150 | 206 | 198 | (56) | 56 | 3,164,063 | 38% |
| 18 | 157 | 00 | 120 | 140 | (120) | 120 | 14,400,000 | #DIV/0! |
| 19 | 157 | 390 | 120 | 124 | 270 | 270 | 72,900,000 | 69% |
| 20 | 118 | 165 | 218 | 123 | (53) | 53 | 2,756,250 | 32% |
| 21 | 157 | 15 | 176 | 106 | (161) | 161 | 26,001,563 | 1075% |
| 22 | 159 | 420 | 143 | 123 | 278 | 278 | 77,006,250 | 66% |
|  |  |  |  |  |  |  |  | 119% |

The table reveals some insights about the performance of the forecast and the shipment. First, the forecast accuracy is low, as indicated by the large standard errors in most weeks. This means that the forecast deviates significantly from the actual shipment data. Second, the errors between the demand and the forecast are also large, especially in weeks 5, 8, 9, 10, 12, 13, 14, 15, 16, 17, 19 and 22. This means that the forecast fails to capture the variations in the demand. Third, the percentage error shows that there are some weeks where the forecast overestimates or underestimates the demand by more than 50%, such as weeks 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 18 and 21. This means that the forecast is not reliable for planning purposes. The average percentage error is 119%, which is very high.

### 4.2.6 105-04752 ORING, LARGE

#### Table 4.2.5 105-04752 ORING, LARGE: Demand vs. Shipment Error Analysis

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | **MAD** | **MSE** | **MAPE** |
| **Week Number** | **DEMAND** | **SHIPMENT** | **4 MOVING AVERAGE** | **STANDARD ERROR** | **ERROR** | **|ERROR|** | **Error^2** | **|%Error|** |
| 1 | 124 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 2 | 89 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 3 | 112 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 4 | 100 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 5 | 92 | 150 | 00 | #N/A | 150 | 150 | 22,500,000 | 100% |
| 6 | 65 | 60 | 38 | #N/A | 23 | 23 | 506,250 | 38% |
| 7 | 66 | 30 | 53 | #N/A | (23) | 23 | 506,250 | 75% |
| 8 | 57 | 75 | 60 | 58 | 15 | 15 | 225,000 | 20% |
| 9 | 71 | 15 | 79 | 58 | (64) | 64 | 4,064,063 | 425% |
| 10 | 73 | 90 | 45 | 22 | 45 | 45 | 2,025,000 | 50% |
| 11 | 78 | 00 | 53 | 28 | (53) | 53 | 2,756,250 | #DIV/0! |
| 12 | 72 | 45 | 45 | 33 | 00 | 00 | 00 | 0% |
| 13 | 116 | 165 | 38 | 33 | 128 | 128 | 16,256,250 | 77% |
| 14 | 117 | 00 | 75 | 54 | (75) | 75 | 5,625,000 | #DIV/0! |
| 15 | 117 | 00 | 53 | 57 | (53) | 53 | 2,756,250 | #DIV/0! |
| 16 | 117 | 255 | 53 | 58 | 203 | 203 | 41,006,250 | 79% |
| 17 | 117 | 150 | 105 | 95 | 45 | 45 | 2,025,000 | 30% |
| 18 | 117 | 00 | 101 | 87 | (101) | 101 | 10,251,563 | #DIV/0! |
| 19 | 117 | 210 | 101 | 97 | 109 | 109 | 11,826,563 | 52% |
| 20 | 87 | 45 | 154 | 98 | (109) | 109 | 11,826,563 | 242% |
| 21 | 71 | 00 | 101 | 69 | (101) | 101 | 10,251,563 | #DIV/0! |
| 22 | 86 | 240 | 64 | 72 | 176 | 176 | 31,064,063 | 73% |
|  |  |  |  |  |  |  |  | 70% |

The table reveals some insights about the performance of the forecast and the shipment. First, the forecast accuracy is low, as indicated by the large standard errors in most weeks. This means that the forecast deviates significantly from the actual shipment data. Second, the errors between the demand and the forecast are also large, especially in weeks 5, 9, 13, 14, 15, 16, 18, 19 and 20. This means that the forecast fails to capture the variations in the demand. Third, the percentage error shows that there are some weeks where the forecast overestimates or underestimates the demand by more than 50%, such as weeks 5, 7, 9, 10, 11, 13, 14, 15, 18 and 20. This means that the forecast is not reliable for planning purposes. The average percentage error is 70%, which is very high.

### 4.2.7 105-07638 ORING, SEAL

#### Table 4.2.6 105-07638 ORING, SEAL: Demand vs. Shipment Error Analysis

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | **MAD** | **MSE** | **MAPE** |
| **Week Number** | **DEMAND** | **SHIPMENT** | **4 MOVING AVERAGE** | **STANDARD ERROR** | **ERROR** | **|ERROR|** | **Error^2** | **|%Error|** |
| 1 | 1,560 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 2 | 1,309 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 3 | 1,194 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 4 | 1,072 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 5 | 1,027 | 420 | 00 | #N/A | 420 | 420 | 176,400,000 | 100% |
| 6 | 874 | 90 | 105 | #N/A | (15) | 15 | 225,000 | 17% |
| 7 | 802 | 180 | 128 | #N/A | 53 | 53 | 2,756,250 | 29% |
| 8 | 781 | 540 | 173 | 159 | 368 | 368 | 135,056,250 | 68% |
| 9 | 755 | 120 | 308 | 197 | (188) | 188 | 35,156,250 | 156% |
| 10 | 757 | 1,140 | 233 | 131 | 908 | 908 | 823,556,250 | 80% |
| 11 | 746 | 00 | 495 | 347 | (495) | 495 | 245,025,000 | #DIV/0! |
| 12 | 868 | 690 | 450 | 414 | 240 | 240 | 57,600,000 | 35% |
| 13 | 942 | 1,230 | 488 | 410 | 743 | 743 | 551,306,250 | 60% |
| 14 | 1,102 | 690 | 765 | 468 | (75) | 75 | 5,625,000 | 11% |
| 15 | 1,102 | 570 | 653 | 340 | (83) | 83 | 6,806,250 | 14% |
| 16 | 1,098 | 1,170 | 795 | 278 | 375 | 375 | 140,625,000 | 32% |
| 17 | 1,098 | 1,050 | 915 | 289 | 135 | 135 | 18,225,000 | 13% |
| 18 | 1,079 | 00 | 870 | 193 | (870) | 870 | 756,900,000 | #DIV/0! |
| 19 | 1,079 | 1,890 | 698 | 398 | 1,193 | 1,193 | 1,422,056,250 | 63% |
| 20 | 711 | 840 | 1,028 | 576 | (188) | 188 | 35,156,250 | 22% |
| 21 | 751 | 60 | 945 | 564 | (885) | 885 | 783,225,000 | 1475% |
| 22 | 893 | 1,650 | 698 | 642 | 953 | 953 | 907,256,250 | 58% |
|  |  |  |  |  |  |  |  | 124% |

The table reveals some insights about the performance of the forecast and the shipment. First, the forecast accuracy is low, as indicated by the large standard errors in most weeks. This means that the forecast deviates significantly from the actual shipment data. Second, the errors between the demand and the forecast are also large, especially in weeks 5, 8, 9, 10, 11, 13, 14, 15, 16, 18, 19 and 22. This means that the forecast fails to capture the variations in the demand. Third, the percentage error shows that there are some weeks where the forecast overestimates or underestimates the demand by more than 50%, such as weeks 5, 8, 9, 10, 11, 13, 18 and 21. This means that the forecast is not reliable for planning purposes. The average percentage error is 124%, which is very high.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | **MAD** | **MSE** | **MAPE** |
| **Week Number** | **DEMAND** | **SHIPMENT** | **4 MOVING AVERAGE** | **STANDARD ERROR** | **ERROR** | **|ERROR|** | **Error^2** | **|%Error|** |
| 1 | 120 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 2 | 144 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 3 | 144 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 4 | 120 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 5 | 108 | 00 | 00 | #N/A | 00 | 00 | 00 | #DIV/0! |
| 6 | 90 | 30 | 00 | #N/A | 30 | 30 | 900,000 | 100% |
| 7 | 90 | 00 | 08 | #N/A | (08) | 08 | 56,250 | #DIV/0! |
| 8 | 90 | 00 | 08 | 12 | (08) | 08 | 56,250 | #DIV/0! |
| 9 | 120 | 00 | 08 | 12 | (08) | 08 | 56,250 | #DIV/0! |
| 10 | 120 | 00 | 08 | 13 | (08) | 08 | 56,250 | #DIV/0! |
| 11 | 110 | 00 | 00 | 06 | 00 | 00 | 00 | #DIV/0! |
| 12 | 110 | 00 | 00 | 05 | 00 | 00 | 00 | #DIV/0! |
| 13 | 90 | 100 | 00 | 04 | 100 | 100 | 10,000,000 | 100% |
| 14 | 90 | 00 | 25 | 38 | (25) | 25 | 625,000 | #DIV/0! |
| 15 | 45 | 160 | 25 | 40 | 135 | 135 | 18,225,000 | 84% |
| 16 | 00 | 00 | 65 | 62 | (65) | 65 | 4,225,000 | #DIV/0! |
| 17 | 00 | 00 | 65 | 70 | (65) | 65 | 4,225,000 | #DIV/0! |
| 18 | 00 | 00 | 40 | 62 | (40) | 40 | 1,600,000 | #DIV/0! |
| 19 | 00 | 200 | 40 | 64 | 160 | 160 | 25,600,000 | 80% |
| 20 | 05 | 300 | 50 | 86 | 250 | 250 | 62,500,000 | 83% |
| 21 | 12 | 00 | 125 | 119 | (125) | 125 | 15,625,000 | #DIV/0! |
| 22 | 29 | 100 | 125 | 133 | (25) | 25 | 625,000 | 25% |
|  |  |  |  |  |  |  |  | 26% |

### 4.2.8 105-06116 ORING, MIC

#### Table 4.2.7 105-06116 ORING, MIC: Demand vs. Shipment Error Analysis

The table reveals some insights about the performance of the forecast and the shipment. First, the forecast accuracy is low, as indicated by the large standard errors in most weeks. This means that the forecast deviates significantly from the actual shipment data. Second, the errors between the demand and the forecast are also large, especially in weeks 6, 13, 15, 19 and 20. This means that the forecast fails to capture the variations in the demand. Third, the percentage error shows that there are some weeks where the forecast overestimates or underestimates the demand by more than 50%, such as weeks 6, 7, 8, 9, 10, 13 and 15. This means that the forecast is not reliable for planning purposes. The average percentage error is 26%, which is relatively low compared to other products.

### 4.2.9 105-04802 SNOUT, COVER

#### Table 4.2.8 105-04802 SNOUT, COVER: Demand vs. Shipment Error Analysis

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | **MAD** | **MSE** | **MAPE** |
| **Week Number** | **DEMAND** | **SHIPMENT** | **4 MOVING AVERAGE** | **STANDARD ERROR** | **ERROR** | **|ERROR|** | **Error^2** | **|%Error|** |
| 1 | 802 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 2 | 579 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 3 | 549 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 4 | 468 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 5 | 519 | 1,110 | 00 | #N/A | 1,110 | 1,110 | 1,231,212,160 | 100% |
| 6 | 513 | 392 | 277 | #N/A | 115 | 115 | 13,133,160 | 29% |
| 7 | 510 | 00 | 375 | #N/A | (375) | 375 | 140,925,160 | #DIV/0! |
| 8 | 494 | 294 | 375 | 457 | (81) | 81 | 6,625,960 | 28% |
| 9 | 611 | 00 | 449 | 463 | (449) | 449 | 201,511,210 | #DIV/0! |
| 10 | 622 | 157 | 172 | 221 | (15) | 15 | 216,090 | 9% |
| 11 | 622 | 196 | 113 | 222 | 83 | 83 | 6,938,890 | 43% |
| 12 | 622 | 255 | 162 | 119 | 93 | 93 | 8,667,610 | 37% |
| 13 | 447 | 78 | 152 | 104 | (74) | 74 | 5,402,250 | 94% |
| 14 | 343 | 412 | 172 | 75 | 240 | 240 | 57,648,010 | 58% |
| 15 | 347 | 98 | 235 | 114 | (137) | 137 | 18,823,840 | 140% |
| 16 | 274 | 372 | 211 | 126 | 162 | 162 | 26,146,890 | 43% |
| 17 | 274 | 00 | 240 | 132 | (240) | 240 | 57,648,010 | #DIV/0! |
| 18 | 163 | 00 | 221 | 166 | (221) | 221 | 48,620,250 | #DIV/0! |
| 19 | 366 | 627 | 118 | 152 | 510 | 510 | 259,692,160 | 81% |
| 20 | 220 | 529 | 250 | 236 | 279 | 279 | 78,008,490 | 53% |
| 21 | 568 | 745 | 289 | 256 | 456 | 456 | 207,662,490 | 61% |
| 22 | 905 | 843 | 475 | 268 | 368 | 368 | 135,056,250 | 44% |
|  |  |  |  |  |  |  |  | 46% |

The table reveals some insights about the performance of the forecast and the shipment. First, the forecast accuracy is low, as indicated by the large standard errors in most weeks. This means that the forecast deviates significantly from the actual shipment data. Second, the errors between the demand and the forecast are also large, especially in weeks 5, 7, 9, 13, 14, 15, 16, 17, 18 and 19. This means that the forecast fails to capture the variations in the demand. Third, the percentage error shows that there are some weeks where the forecast overestimates or underestimates the demand by more than 50%, such as weeks 5, 7, 9, 13, 15 and 17. This means that the forecast is not reliable for planning purposes. The average percentage error is 46%, which is relatively high compared to other products.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | **MAD** | **MSE** | **MAPE** |
| **Week Number** | **DEMAND** | **SHIPMENT** | **4 MOVING AVERAGE** | **STANDARD ERROR** | **ERROR** | **|ERROR|** | **Error^2** | **|%Error|** |
| 1 | 725 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 2 | 567 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 3 | 496 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 4 | 487 | 00 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| 5 | 390 | 225 | 00 | #N/A | 225 | 225 | 50,625,000 | 100% |
| 6 | 335 | 00 | 56 | #N/A | (56) | 56 | 3,164,063 | #DIV/0! |
| 7 | 310 | 28 | 56 | #N/A | (29) | 29 | 813,961 | 103% |
| 8 | 299 | 261 | 63 | 91 | 198 | 198 | 39,132,752 | 76% |
| 9 | 240 | 62 | 128 | 112 | (66) | 66 | 4,356,924 | 106% |
| 10 | 237 | 595 | 88 | 75 | 508 | 508 | 257,712,838 | 85% |
| 11 | 227 | 77 | 237 | 192 | (160) | 160 | 25,486,765 | 207% |
| 12 | 279 | 359 | 249 | 210 | 110 | 110 | 12,081,142 | 31% |
| 13 | 385 | 648 | 273 | 204 | 375 | 375 | 140,298,377 | 58% |
| 14 | 351 | 00 | 420 | 233 | (420) | 420 | 176,257,229 | #DIV/0! |
| 15 | 351 | 00 | 271 | 201 | (271) | 271 | 73,424,741 | #DIV/0! |
| 16 | 349 | 420 | 252 | 221 | 168 | 168 | 28,318,158 | 40% |
| 17 | 349 | 179 | 267 | 230 | (88) | 88 | 7,807,490 | 49% |
| 18 | 339 | 00 | 150 | 201 | (150) | 150 | 22,398,116 | #DIV/0! |
| 19 | 339 | 547 | 150 | 166 | 397 | 397 | 157,783,728 | 73% |
| 20 | 278 | 00 | 286 | 169 | (286) | 286 | 82,013,504 | #DIV/0! |
| 21 | 312 | 00 | 181 | 176 | (181) | 181 | 32,898,704 | #DIV/0! |
| 22 | 370 | 571 | 137 | 188 | 435 | 435 | 188,859,776 | 76% |
|  |  |  |  |  |  |  |  | 56% |

### 4.2.10 110-02888-N BUTTON AND 110-02898-T BUTTON, THIN

#### Table 4.2.9 110-02888-N BUTTON AND 110-02898-T BUTTON, THIN: Demand vs. Shipment Error Analysis

The table reveals some insights about the performance of the forecast and the shipment. First, the forecast accuracy is low, as indicated by the large standard errors in most weeks. This means that the forecast deviates significantly from the actual shipment data. Second, the errors between the demand and the forecast are also large, especially in weeks 5, 8, 9, 10, 11, 13, 14, 15, 16, 18 and 19. This means that the forecast fails to capture the variations in the demand. Third, the percentage error shows that there are some weeks where the forecast overestimates or underestimates the demand by more than 50%, such as weeks 5, 7, 9, 10, 11, 14 and 15. This means that the forecast is not reliable for planning purposes. The average percentage error is 56%, which is relatively high compared to other products.

## 4.3. Hypothesis checking about the average total planned production per week

### 4.3.1 Dataset preparation

By Applying the same way as section 4.1.1, we can get the data of the total planned production per week.

#### Table 4.3 Total planned production in each week

|  |  |
| --- | --- |
| **Week** | **Total Planned Production (x-axis)** |
| 1 | 39 |
| 2 | 29 |
| 3 | 24 |
| 4 | 22 |
| 5 | 22 |
| 6 | 21 |
| 7 | 18 |
| 8 | 16 |
| 9 | 18 |
| 10 | 17 |
| 11 | 16 |
| 12 | 21 |
| 13 | 24 |
| 14 | 22 |
| 15 | 22 |
| 16 | 23 |
| 17 | 23 |
| 18 | 22 |
| 19 | 22 |
| 20 | 20 |
| 21 | 19 |
| 22 | 21 |
| 23 | 22 |
| 24 | 23 |
| 25 | 23 |
| 26 | 34 |
| 27 | 38 |
| 28 | 30 |
| 29 | 31 |
| 30 | 37 |
| 31 | 36 |
| 32 | 35 |
| 33 | 36 |
| 34 | 36 |
| 35 | 31 |
| 36 | 30 |
| 37 | 29 |
| 38 | 38 |
| 39 | 47 |
| 40 | 54 |
| 41 | 55 |
| 42 | 51 |
| 43 | 50 |
| 44 | 46 |
| 45 | 52 |
| 46 | 57 |
| 47 | 51 |
| 48 | 49 |
| 49 | 40 |
| 50 | 31 |
| 51 | 39 |
| 52 | 37 |

### 4.3.2 Hypothesis Checking

The data is used to conduct a one sample test for mean to assess whether the average total planned production per week meets a specific criterion (mean >= 30). The null hypothesis (H0) assumes that the average of total planned production per week is greater than or equal to 30. The alternative hypothesis (H1) assumes that the average of total planned production per week is less than 30. After that, we calculate all the value needed to perform the hypothesis checking

Average:

The average total planned production per week was calculated by summing up all the values and dividing by the sample size. We found the average to be 31.71 units.

Standard Deviation:

To understand the variability in the planned production, we calculated the standard deviation. The computed standard deviation was 11.82 units.

Degrees of Freedom:

Based on the sample size, the degrees of freedom for this analysis were determined to be 51.

Significance Level (α):

We set the significance level (α) at 0.05, indicating a 5% chance of making a Type I error.

Test Statistic:

The test statistic represents the standardized value used to assess the hypotheses. In our analysis, the calculated test statistic was 1.0432, obtained by dividing the difference between the sample mean and the hypothesized mean by the standard deviation.

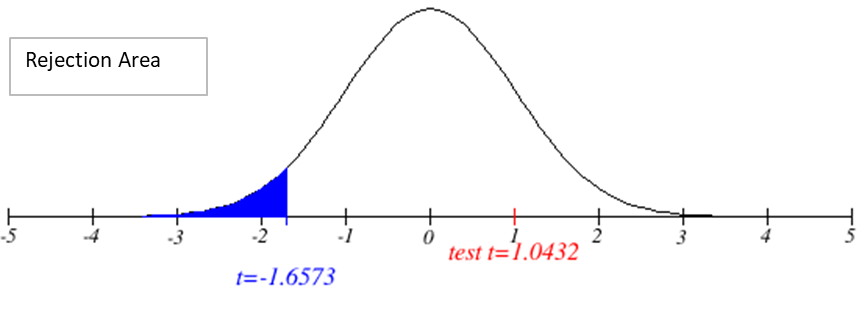
Critical Value:

Using the significance level and degrees of freedom, we determined the critical value for a one-tailed test. In this case, the critical value for α=0.05 and 51 degrees of freedom was found to be 1.6753 and we have adjusted the critical value to negative value, -1.6753. In this case:

* Null Hypothesis (H0): Average of total planned production per week >= 30
* Alternative Hypothesis (H1): Average of total planned production per week < 30

Since the alternative hypothesis states that the average is less than 30, we are conducting a lower tailed test. In a lower tailed test, we are looking for evidence to reject the null hypothesis and support the alternative hypothesis when the t-stat is lower than critical value. The t-distribution is a symmetric distribution, and in a lower tailed test, we are interested in the lower end of the distribution where the values are smaller. Hence, the critical value for the lower tailed test is negative.

After getting all the value, we compare the test statistic value and the critical value. Since we conducted a lower-tailed test, we compared the test statistic to the critical value. In this case, the calculated t-statistic (1.0432) was greater than the critical value (-1.6753) and did not fall in the rejection area as shown in the below graph. In conclusion, we do not have enough evidence to reject the null hypothesis. The data does not provide strong evidence to conclude that the average of total planned production per week is more than or equal to 30. In statistical terms, we fail to reject the null hypothesis (H0: Average of total planned production per week >= 30) at the 0.05 significance level. Even though the sample mean of 31.71 is well above 30, we have too much sampling error to conclude the that the true population mean is greater 30.

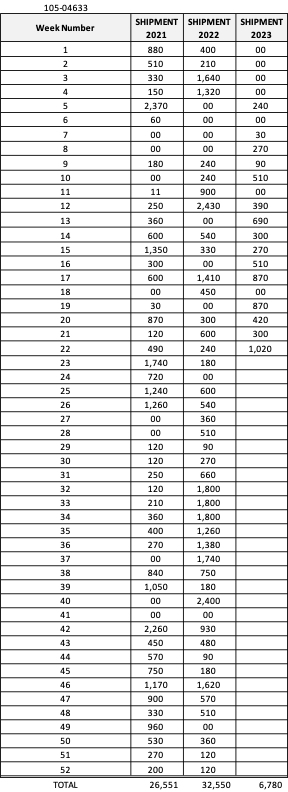
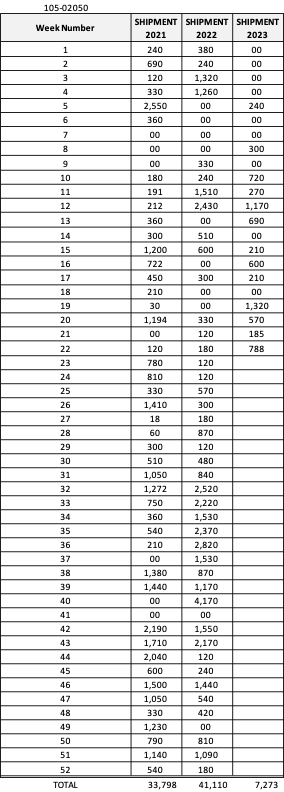


##### Figure 4.4 Normal distribution of lower tailed rejection area

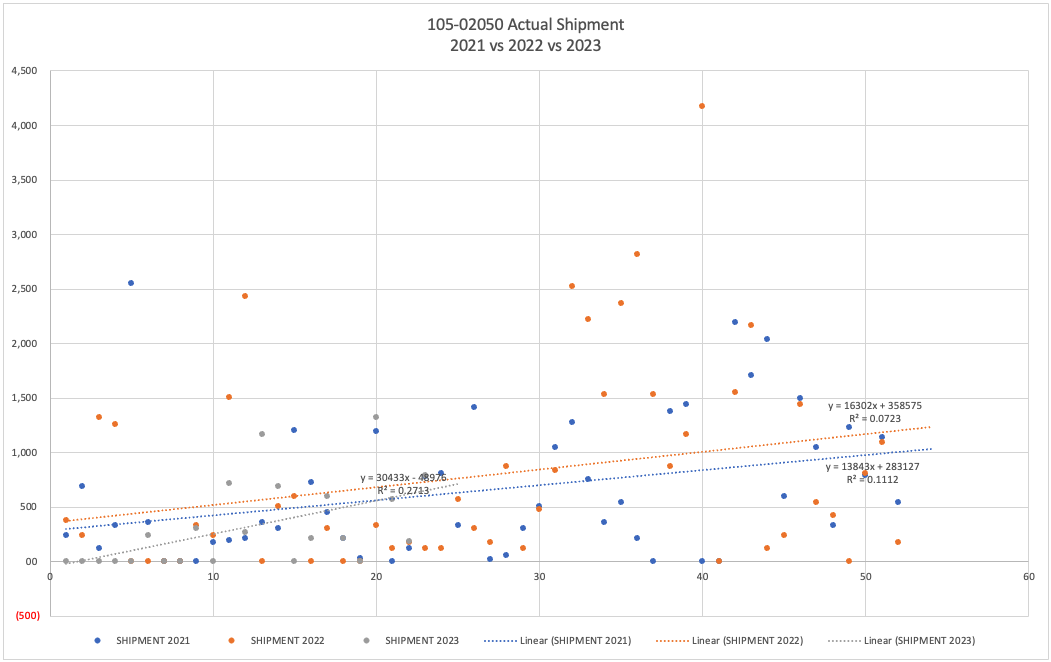
## 4.4 Regression analysis for Actual shipment 2021 vs 2022 vs 2023

### 4.4.1 Original data

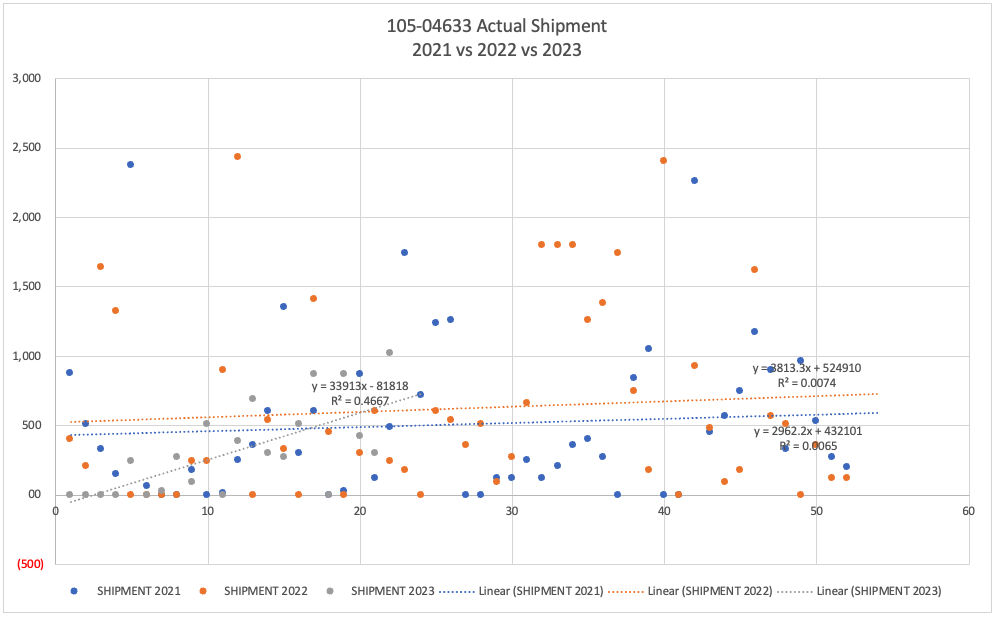
#### Table 4.4 Shipment data for part number 105-02050 and 105-04633 over three years



### 4.4.2 Trendline analysis



##### Figure 4.5 Trendline for shipment data of part number 105-02050 over three years



##### Figure 4.6 Trendline for shipment data of part number 105-04633 over three years

### 4.4.3 Result Analysis

We can interpret the trendline for 105-04633 (year 2022) as following:

1. Trendline equation:
   1. The equation y = 3813.3x + 524910 represents the estimated relationship between the independent variable (week number) and the dependent variable (year 2022 actual shipment).
   2. In this case, it suggests that for every unit increase in x (week number), the estimated value of y (shipment volume) increases by 3813.3 units and is then further increased by 524910 units.
2. Positive trend:
   1. The positive coefficient (3813.3) indicates that there is a positive relationship between the week number (x) and the shipment volume (y).
   2. As the week number increases, the estimated shipment volume tends to increase as well.
3. Intercept:
   1. The intercept of 524910 suggests that the estimated shipment volume at the starting point (x = 0) is 524910 units.
   2. This value represents the initial shipment volume when the week variable is zero or not applicable.
4. R-squared value:
   1. The R-squared value of 0.0074 indicates that only approximately 0.74% of the variation in the dependent variable (shipment volume) can be explained by the independent variable (week number) in the regression model.
   2. This implies that the week number variable has *limited predictive power* in explaining the changes in shipment volume, as most of the variation remains unexplained by the regression model.
5. Low explanatory power:
   1. With an R-squared value of 0.0074, the regression model does not provide a strong fit to the data.
   2. This suggests that other factors, not accounted for in the model, may have a more significant influence on the shipment volume.
   3. It's essential to consider other variables or factors that may impact the data, such as market conditions, economic factors, or specific events.

### 4.4.4 Result Conclusion

Regression analysis is a statistical method used to model the relationship between a dependent variable and one or more independent variables. In the context of actual shipment data for the years 2021, 2022, and 2023, a regression analysis can help identify patterns and trends, as well as make predictions or forecasts.

The regression analysis above compares the actual shipment data for the years 2021, 2022, and 2023 for two specific parts, namely 105-02050 and 105-04633. The analysis was limited to these parts due to the availability of data. Upon examining the scatter plot representing the shipment data for these parts, we observed a consistent trend between them. In particular, the trendlines for the years 2021 and 2022 exhibited a similar pattern, indicating a certain level of consistency in the shipment numbers. However, it is worth noting that the total shipments for 2022 were higher compared to 2021. This suggests a positive growth or an increased demand for these parts during that period.

However, the trendline pattern took a noticeable shift when we analysed the data for the year 2023. It is possible that this change in trend is attributed to economic regression or a downturn in the overall economic conditions. While we cannot confirm the exact factors contributing to this shift based solely on the regression analysis, it is plausible to consider the influence of economic factors on the shipment patterns. To gain a more comprehensive understanding of the situation, further monitoring and analysis are needed.

# 5. Conclusion

In conclusion, the data analysis conducted in this project encompassed various phases, including data collection, data pre-processing, data analysis, and data visualization. The collected data sets consisted of information related to the manufacturing process of a plastic & rubber factory in 2023, as well as customer forecast data and actual shipping plans to predict future shipping requirements. The analysis focused on regression analysis, hypothesis testing, and forecasting techniques to achieve the objectives of uncovering the association between planned production and manpower needed, examining the variability between predicted and actual shipping data, investigating the hypothesis regarding average total planned production per week, and exploring the relationships within the actual shipment data across the years 2021, 2022, and 2023.

The findings of the analysis are as follows:

I: The regression analysis revealed a positive association between total planned production and the number of manpower needed. The results indicated that for every unit increase in total planned production, there is an approximate increase of 3.77 (rounded up to 4) in the total manpower required. This model provides reliable insights that can support informed decision-making.

II: In general, the average percentage of error for most of the part is above 50%, some even goes beyond 100%. This indicates that the forecasted data is not so accurate, due to many unstable external factors such as production or delivery delay, market changes and etc. All these factors need to be resolved or alternate data analysis approach like multiple regression should be implemented instead to increase the accuracy of forecasting result.

III: The hypothesis testing results indicated that there is insufficient evidence to reject the null hypothesis, which suggests that the average total planned production per week is not greater than or equal to 30. Although the sample mean was above 30, the presence of significant sampling error prevents us from concluding that the true population mean exceeds 30.

IV: The regression analysis conducted on the actual shipment data for the years 2021, 2022, and 2023 compared the shipment patterns of specific parts, namely 105-02050 and 105-04633. The trendlines for 2021 and 2022 exhibited a consistent pattern, indicating a level of stability and growth in shipments. However, in 2023, there was a noticeable shift in the trendline pattern, suggesting a potential influence of economic factors or other external variables. Further investigation and monitoring are necessary to gain a comprehensive understanding of this shift.

In this project, data visualization played a crucial role in transforming the analysed data into various illustrations, such as histograms, line charts, and scatter charts. This visual representation facilitated easier interpretation and extraction of relevant information from the data, enhancing the clarity and effectiveness of communication.

In conclusion, the findings from the data analysis provide valuable insights and recommendations to support managerial decision-making. It is recommended to further refine the forecasting methodology and techniques to improve accuracy, particularly in capturing demand variations. Additionally, continuous monitoring, evaluation, and adjustment of the forecast performance are essential to enhance operational efficiency and customer satisfaction. By utilizing the power of data analysis and visualization, organizations can drive innovation, improvement, and informed decision-making, ultimately leading to enhanced performance and competitiveness in the manufacturing industry.

# References

*Kelley, K. (2023, June 6). Mastering data analysis: Process, types, methods, and Techniques. Simplilearn.com.* [*https://www.simplilearn.com/data-analysis-methods-process-types-article*](https://www.simplilearn.com/data-analysis-methods-process-types-article)