Implementing Stepwise Regression in Studying the Dynamics of Listed Establishments in the Philippine Manufacturing Sector — Insights from the Past and Potential of the Future

Jomari Manto   
*College of Computer Studies & Engineering*  
*Jose Rizal University*Mandaluyong City, Philippines  
[jomari.manto@my.jru.edu](mailto:jomari.manto@my.jru.edu)

***Abstract*—This research provides analysis on the dynamics of listed establishments in the Philippine manufacturing sector. Stepwise regression is applied to identify the most influential factor that dictates the overall trend of the Philippine manufacturing sector in terms of listed establishments. The significance of this analysis to real world context in the Philippines is also included in the report. (***Abstract***)**

***Keywords—Philippine manufacturing sector, trends in the Philippine manufacturing, Manufacturing data MSMEs Philippines, Manufacturing industry establishment prediction, Manufacturing sector statistical analysis Philippines (****keywords****)***

# Introduction

## *Background of the Study*

The manufacturing industry plays a vital role in a country’s economic development. Despite facing numerous global challenges and setbacks over the years, it continues to be a cornerstone of economic growth. Su and Yao [1] demonstrated that advancements in manufacturing can generate ripple effects across other sectors by enhancing human capital—through skills, knowledge, experience, and education—as well as strengthening economic institutions, such as laws, regulations, and structural frameworks. This underscores the manufacturing sector’s enduring status as a key engine of economic progress.

In the context of the Philippines, the manufacturing industry’s contribution to its 2023 Gross Domestic Product (GDP) amounts to 3.78 trillion Philippine pesos, making it the second largest sector. According to Statista, semiconductors and other electronics are the primary exports of the manufacturing sector [2]. Despite this milestone, the Philippine sector continues to face drawbacks due to various factors.

## *Statement of the Problem*

Various studies have been conducted to identify the challenges in the Philippine economic sector. A study by Aldaba [3] on multiple performance indicators underscores the stagnant growth of the Philippine manufacturing sector. Statista [2] also reports that one of the major challenges of the sector is its reliance on imported raw materials, leading to trade deficits.

A Monthly Integrated Survey of Selected Industries (MISS) report published by the Philippine Statistics Authority (PSA) on April 8, 2025 shows a decline in two performance indicators—Value of Production Index (VaPI) and Value of Net Sales Index (VaNSI) [4].

Furthermore, a summary of a report by Christy Balita on the number of establishments in the manufacturing sector covering 2016-2022 reported a decline [5].

With these findings, it is crucial to study the factors that affect the sector's performance. This research will use historical data on the number of listed establishments, henceforth referred to as LE, to identify trends and predict future values using stepwise regression.

## *Hypotheses of the Study*

## The Philippine Statistics Authority classifies manufacturing establishments into four categories based on total employment. The categories of LEs are as follows:

## Micro: 1 to 9 employees

## Small: 10 to 99 employees

## Medium: 100 to 199 employees

## Large: 200 or more employees

Based on these categories, the hypotheses of these research are as follows:

**Ho1:** The micro-scale manufacturing sector (1–9 employees) comprises the majority of manufacturing establishments in the Philippines, reflecting the dominance of small-scale industrial activities.

**Ho2:** The small-scale manufacturing sector (10–99 employees) shows a steady presence across regions, suggesting its role as a backbone of local industrial development.

**Ho3:** The number of medium-scale manufacturing establishments (100–199 employees) remains limited, indicating potential barriers to enterprise growth beyond the small-scale level.

**Ho4:** The large-scale manufacturing sector (200 or more employees) accounts for the fewest number of establishments, emphasizing the challenges in scaling operations to a large employment size.

Lastly, in conjunction with our problem statement, the fifth hypothesis is:

**Ho5:** The overall number of manufacturing establishments in the Philippines has been declining over time, suggesting a downward trend in the growth of the sector.

*D. Scope and Limitation of the Study*

This study focuses on the number of listed establishments. Other performance indicators are not included. The number of observations is based on available data.

## *E. Significance of the Study*

This study is significant to ordinary Filipinos as it aims to raise awareness about the importance of supporting the country’s domestic manufacturing sector. By increasing their understanding of the sector’s impact, this research encourages Filipinos to make more informed decisions—such as supporting local businesses, advocating for stronger government support, and playing an active role in the growth of the manufacturing industry.

# Methodology

## *Data*

## This research utilizes official data on the number of listed establishments in the Philippine manufacturing sector. The dataset was provided by the PSA FOI Team upon the researcher’s request (Reference No. FOI-REQ-2025-3951). It includes raw data on the number of establishments by region, section, and employment groupings (Micro, Small, Medium, Large) in the Philippines from 1993 to 2023, along with information from other sectors such as agriculture and fisheries.

## For the purpose of this study, data specific to the manufacturing sector was extracted and organized into a final dataset containing the following variables:

1. DataFrame Variables And Descriptions

| **Variable** | **Description** |
| --- | --- |
| Year | Reference year (1999–2023) |
| Total | Total number of listed manufacturing establishments |
| Micro | Number of micro establishments (1–9 employees) |
| Small | Number of small establishments (10–99 employees) |
| Medium | Number of medium establishments (100–199 employees) |
| Large | Number of large establishments (200+ employees) |

1. Data frame variables and descriptions.

## *Exploratory Data Analysis (EDA)*

## As a prerequisite for linear regression, stepwise regression, and predictive modeling, exploratory data analysis (EDA) was applied to the dataset. The objectives are as follows:

## Identify the trend in the total number of establishments from 1999 to 2023

## Detect outliers or significant changes over time

## Examine the distribution and proportion of establishments by size category

## Identify significant relationships between each category vs total number of establishments

## Assess datasets' readiness for regression modeling.

## The following tables and figures illustrate the results of the EDA and their significance in the context of the research problem statement.



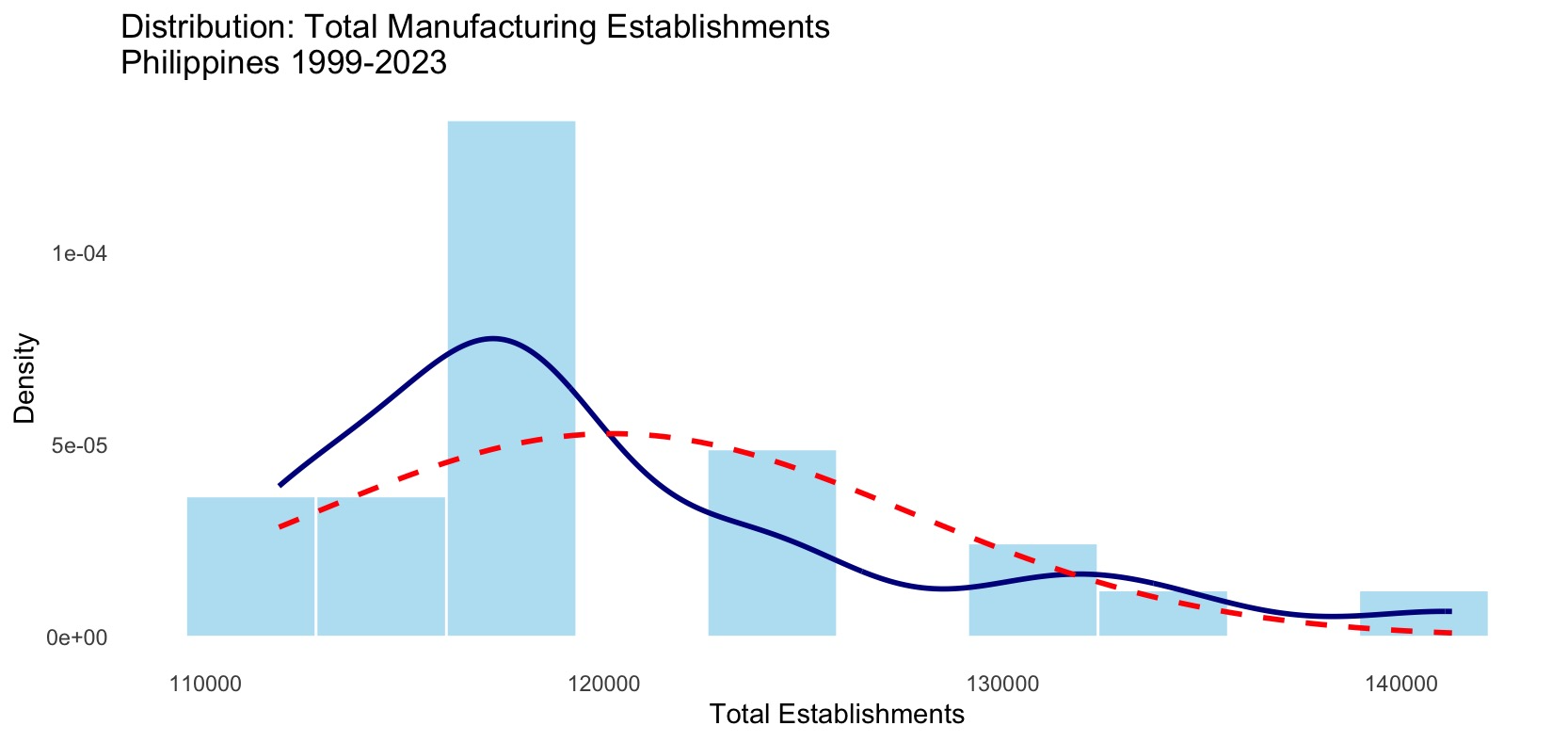
1. Trend of Listed Manufacturing Establishments in the Philippines (1999–2023).

## Fig. 2 shows the trend of listed manufacturing establishments from 1999 to 2023. The red dashed trend line reflects minimal upward movement, indicating stagnant growth in the number of establishments within the industry.



1. Trend of Listed Manufacturing Establishments in the Philippines (1999–2023) per Category

## Fig. 3 shows the trend of listed manufacturing establishments by category from 1999 to 2023. Except for Micro establishments (first row, third column), all categories—Small, Medium, and Large—exhibit a decline in numbers. Overall, these illustrations highlight a clear downward trend in the number of listed establishments.



1. Distribution of Total Manufacturing Establishments in the Philippines (1999–2023)

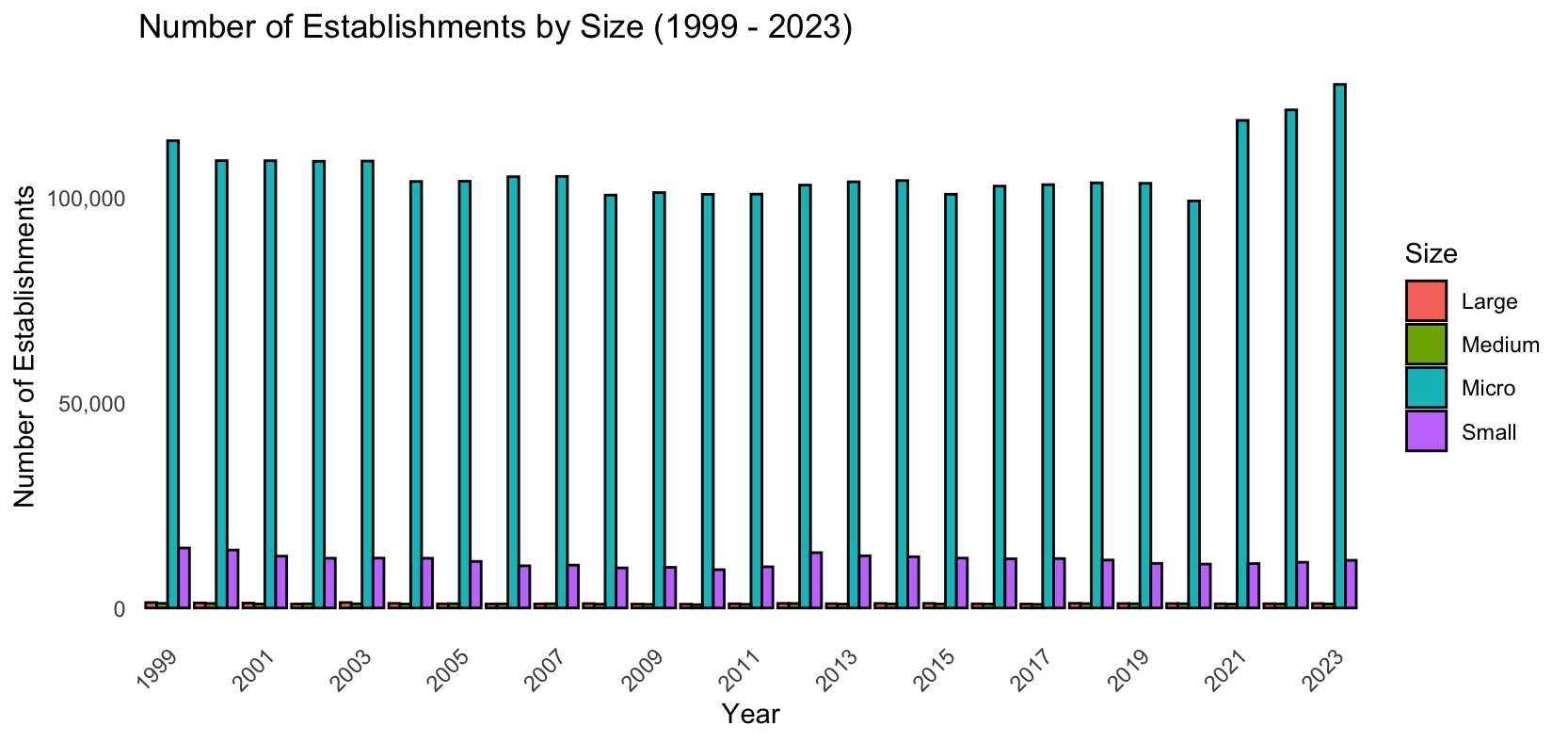
## Fig. 4 shows the number of manufacturing businesses in the Philippines from 1999 to 2023. Most years had between 117,000 and 122,000 establishments, with fewer years having much higher numbers. This means the average is pulled up by a few outliers, but in general, the number of businesses stayed on the lower side. This could mean that growth in the manufacturing sector hasn’t been steady and may be affected by deeper issues or changing conditions over time.

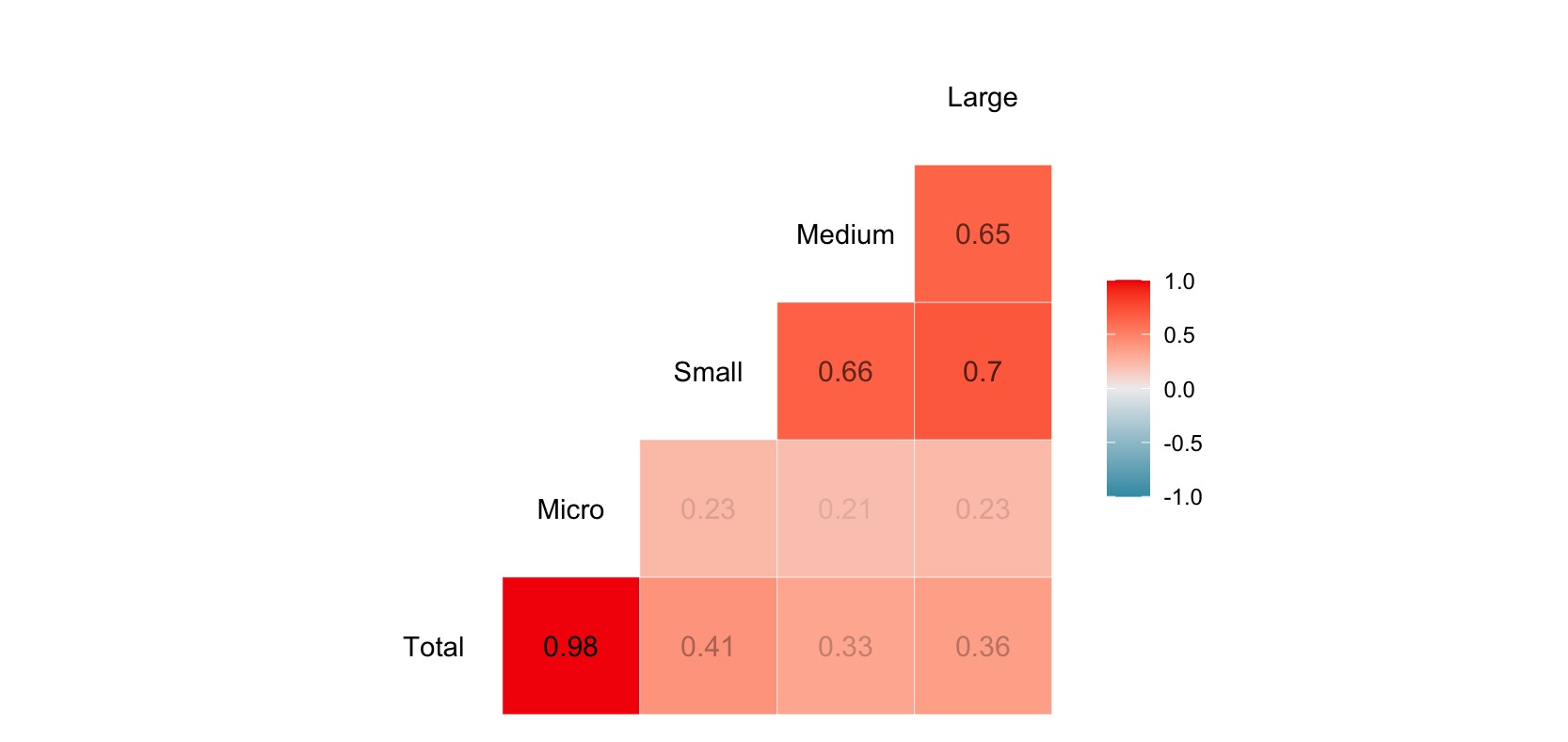
## 

1. Box Plot of Manufacturing Establishments in the Philippines by Total and Size Categories (1999–2023)

## Fig. 5 presents box plots of manufacturing establishments in the Philippines from 1999 to 2023, showing the total, and size categories.

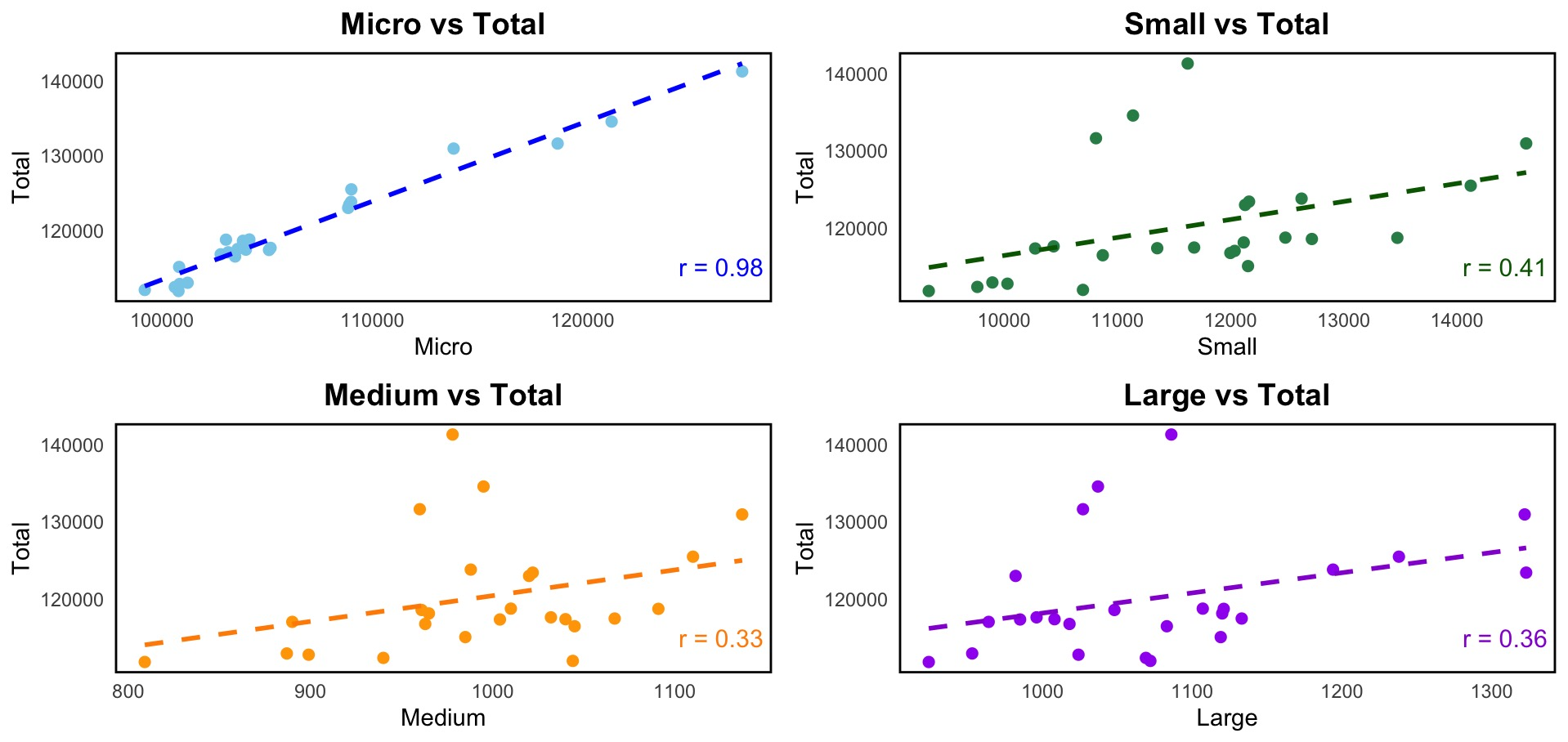
## The box plots reveal several trends in the number of businesses over the years. First, there were more years with fewer businesses, especially for smaller establishments, as shown by the boxes being pushed toward the top, indicating most numbers were on the lower side. However, some years saw spikes in the number of small businesses, marked by dots above the normal range. For medium and large businesses, the data was more steady, with smaller and more stable boxes, and fewer surprising jumps. Micro businesses, on the other hand, dominated in terms of numbers, with their values reaching over 110,000, making them the largest group. Lastly, large businesses remained small and steady, with little change over the years, reflected by a tight, low box in the plot.

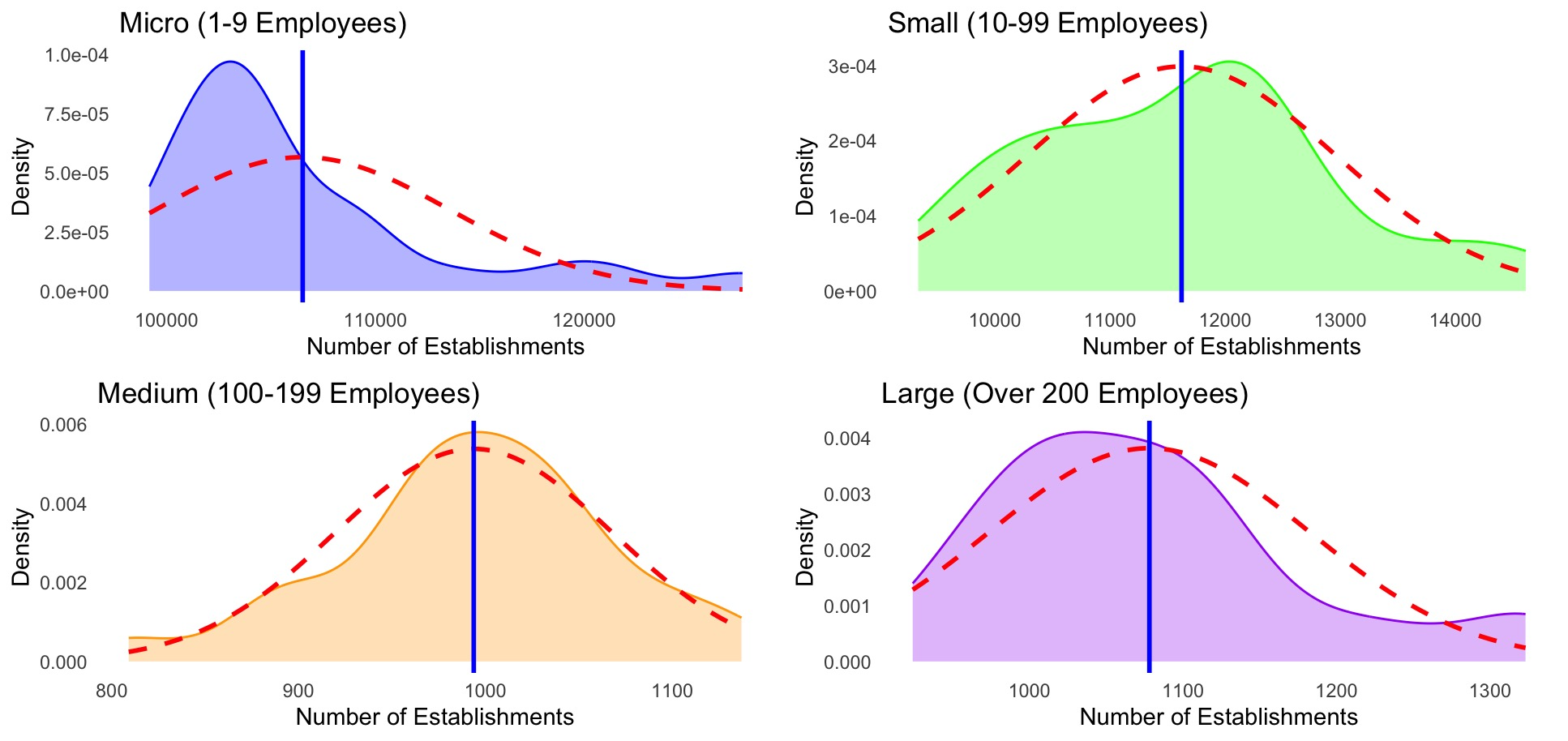




1. Side-by-Side Bar Chart of Philippine Manufacturing Establishments by Size (1999–2023) & Correlation Matrix

## Fig. 6 shows the dominance of micro establishments, suggesting their significant influence on the total number of manufacturing establishments in the Philippines. The bar charts highlight the towering height of the Micro category compared to the Small, Medium, and Large categories. Meanwhile, the correlation matrix indicates that the Micro category has the strongest positive relationship with the total number of listed establishments, with a correlation value of 0.98.





1. Scatter Plots of Size Categories vs Total Number of Establishments and Density Plot of Size Categories

## Fig. 7 shows both scatter plots and density plots. The scatter plots illustrate the linear relationship between each size category and the total number of listed establishments. Among them, the Micro category shows the strongest positive relationship, backed by a high correlation value of 0.98 and a tight clustering of data points along the trendline. This means that the number of micro establishments closely follows the overall trend with little variation. Meanwhile, the density plots show how the values are distributed. The Small, Medium, and Large categories appear closest to a normal distribution, with curves that are fairly symmetrical. In contrast, the Micro category is right-skewed, meaning most of its values fall below the mean, with fewer but higher values stretching the curve to the right.

## In summary, our Exploratory Data Analyses have provided the following:

## **Trend of the Data**: There is a downward trend in the total number of listed establishments over time..

## **Linear Relationship**: A strong positive linear relationship was found between the different size categories and the total number of establishments, with the Micro size category showing the most significant relationship to the total.

1. **Data Distribution**: Although the data is not perfectly normally distributed, it exhibits sufficient normality for it to be considered appropriate for linear regression modeling.

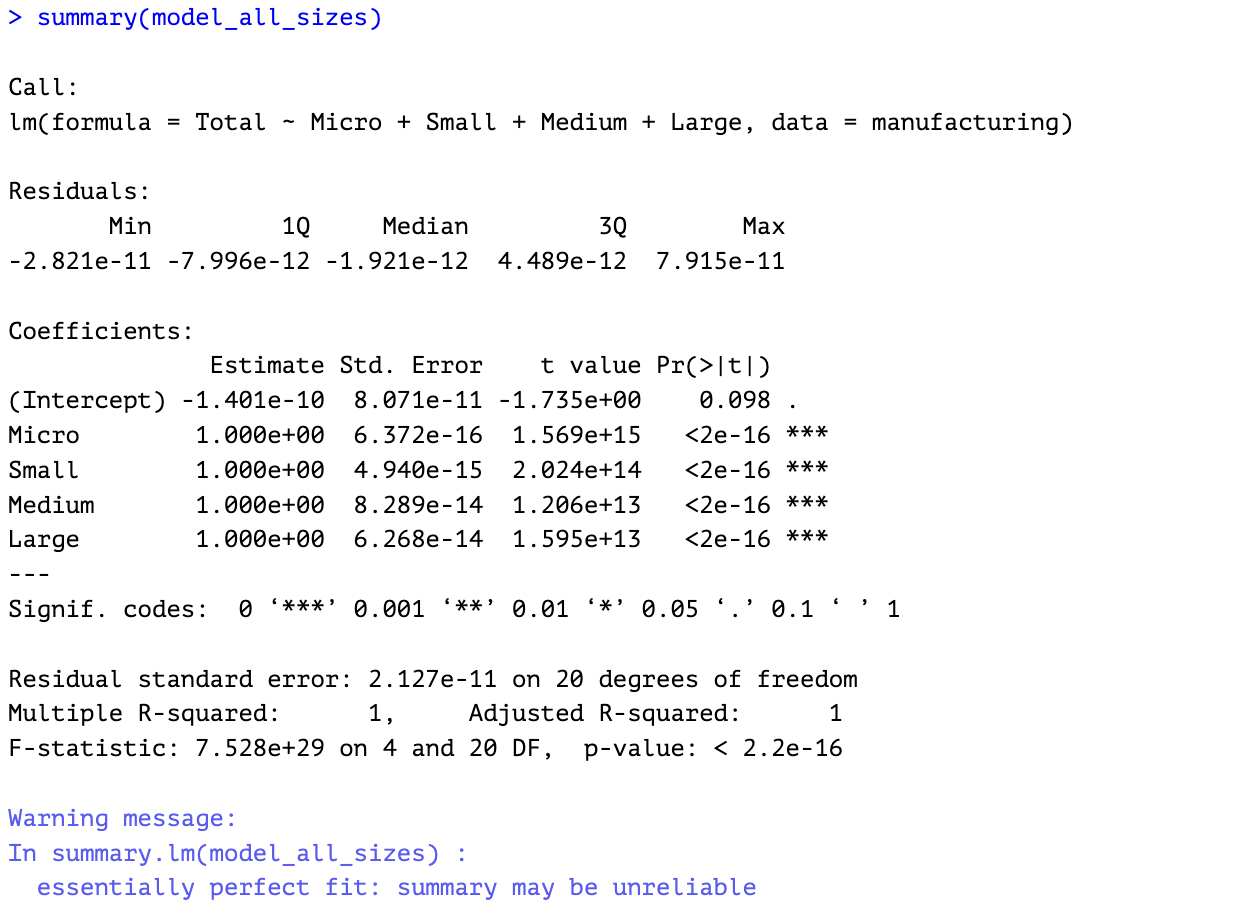
## 

## *Linear Regression with Stepwise Variable Selection*

This section outlines the process of building the predictive model, starting with variable selection, followed by the creation of a base model, application of stepwise regression to refine the model, and concluding with model selection, testing, and evaluation.

The response variable is Total, while the predictor variables are the size categories: Micro, Small, Medium, and Large.

For the initial linear regression base model, all predictor variables were included and processed in R using the *lm()* function.



1. Summary of Linear Base Model Using All Predictors.

The results of our initial model show that all the size categories (Micro, Small, Medium, Large) are important predictors, and they have a strong relationship with the total number of establishments. The model has very small errors, and the R-squared value is 1, meaning it fits the data almost perfectly. However, this perfect fit can be a red flag because it might mean the model is too closely matched to the current data and may not work as well with new data.

The Adjusted R-squared value, which adjusts for the number of predictors in the model, is also 1, confirming that the model is overly tailored to the data. Additionally, the very low p-values (less than 0.001) suggest that the predictors are statistically significant, meaning they are likely important, but this doesn’t guarantee the model’s effectiveness outside the current dataset.

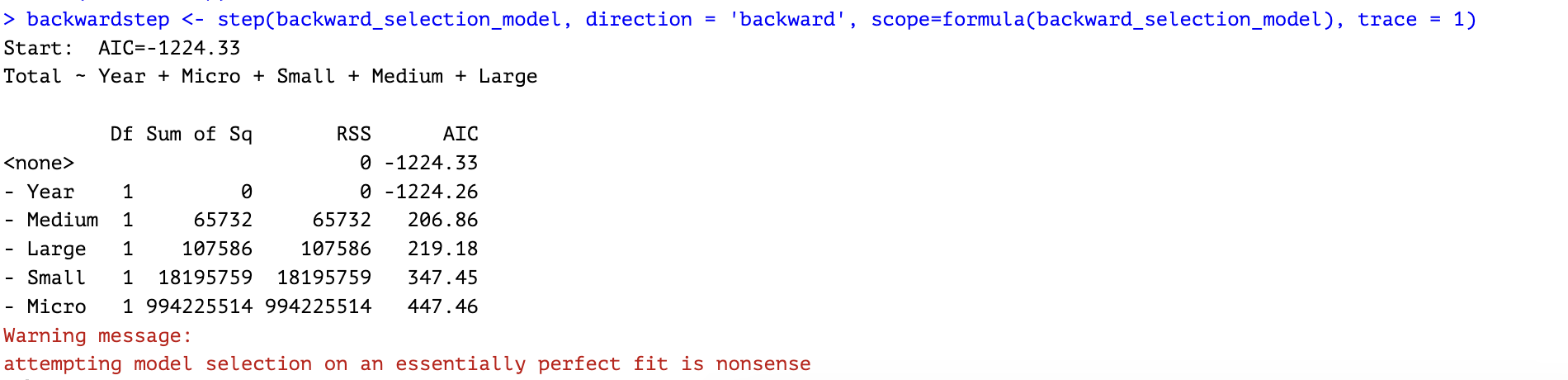
To make sure the model works well with new data and to focus on the most important predictors, stepwise regression will be used. This approach will help improve the model by selecting only the most significant factors, ensuring the final model is simpler and more reliable.

Stepwise regression will be applied in the following order:

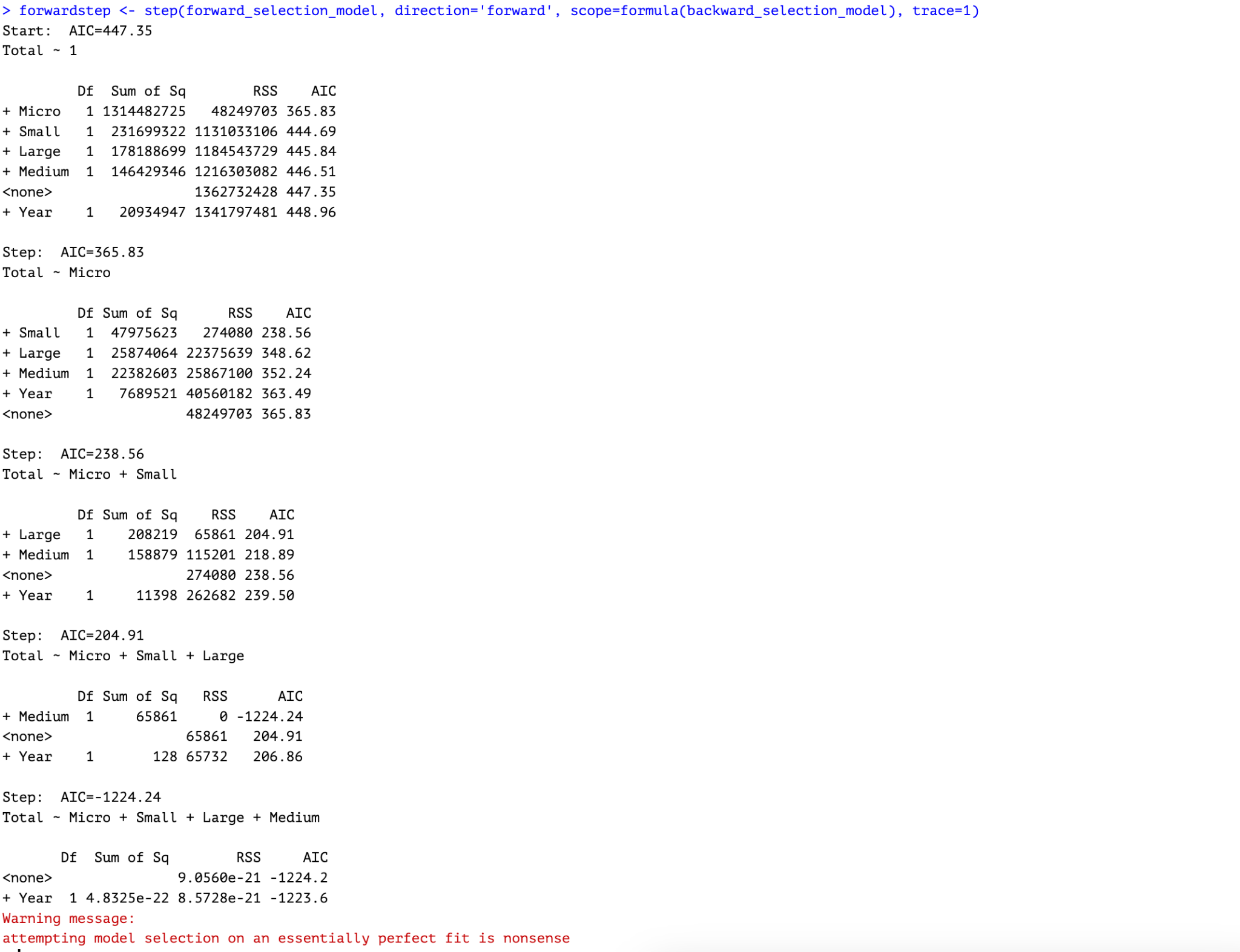
Backward — eliminating the least significant predictors one-by-one on the base model.

Forward — iteratively adding predictors to a null model, i.e. the model is created using the coefficient only, without any predictors.

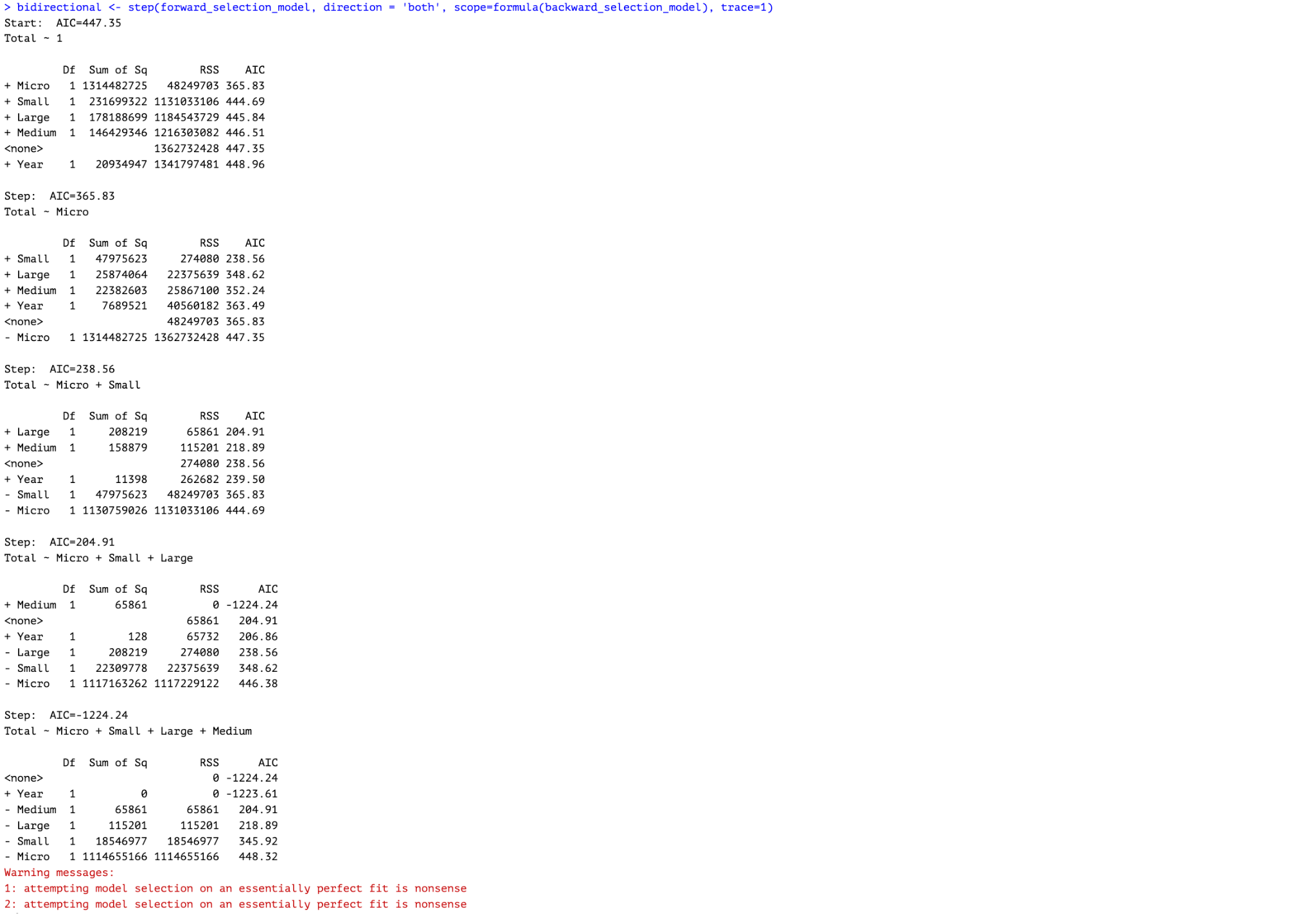
Bidirectional — simultaneously adding and removing predictors to the model of the data set.



1. Results of Backward Elimination Using *step()* Function



1. Results of Forward Selection Using *step()* Function

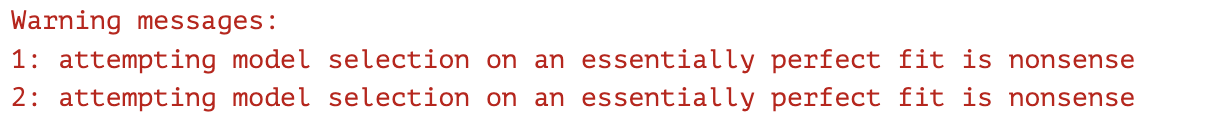


1. Results of Bidirectional Elimination Using *step()* Function

Figures 9, 10, and 11 show the results of the stepwise regression analysis conducted using the step() function in R. The models used are stored in the variables *backward\_selection\_model* (the model including all predictors for the Total variable) and *forward\_selection\_model* (the model created based on coefficients alone, without including any predictors).

In an ideal scenario, stepwise regression is expected to iteratively add or remove predictor variables in order to refine the model and retain only those that contribute significantly to predicting the response variable. However, in this case, the stepwise regression process—whether performed in a forward, backward, or bidirectional manner—ultimately retained all predictors (Micro, Small, Medium, and Large) in the final model. This lack of variable elimination suggests that all predictors are strongly correlated with the response variable.

Furthermore, R consistently issued the following warning message during the model selection process:



1. Warning message, indicating that the model is overfitting, and may be unreliable.

This warning is a clear indication of overfitting, where the model fits the dataset too perfectly, potentially compromising its ability to generalize to new data.

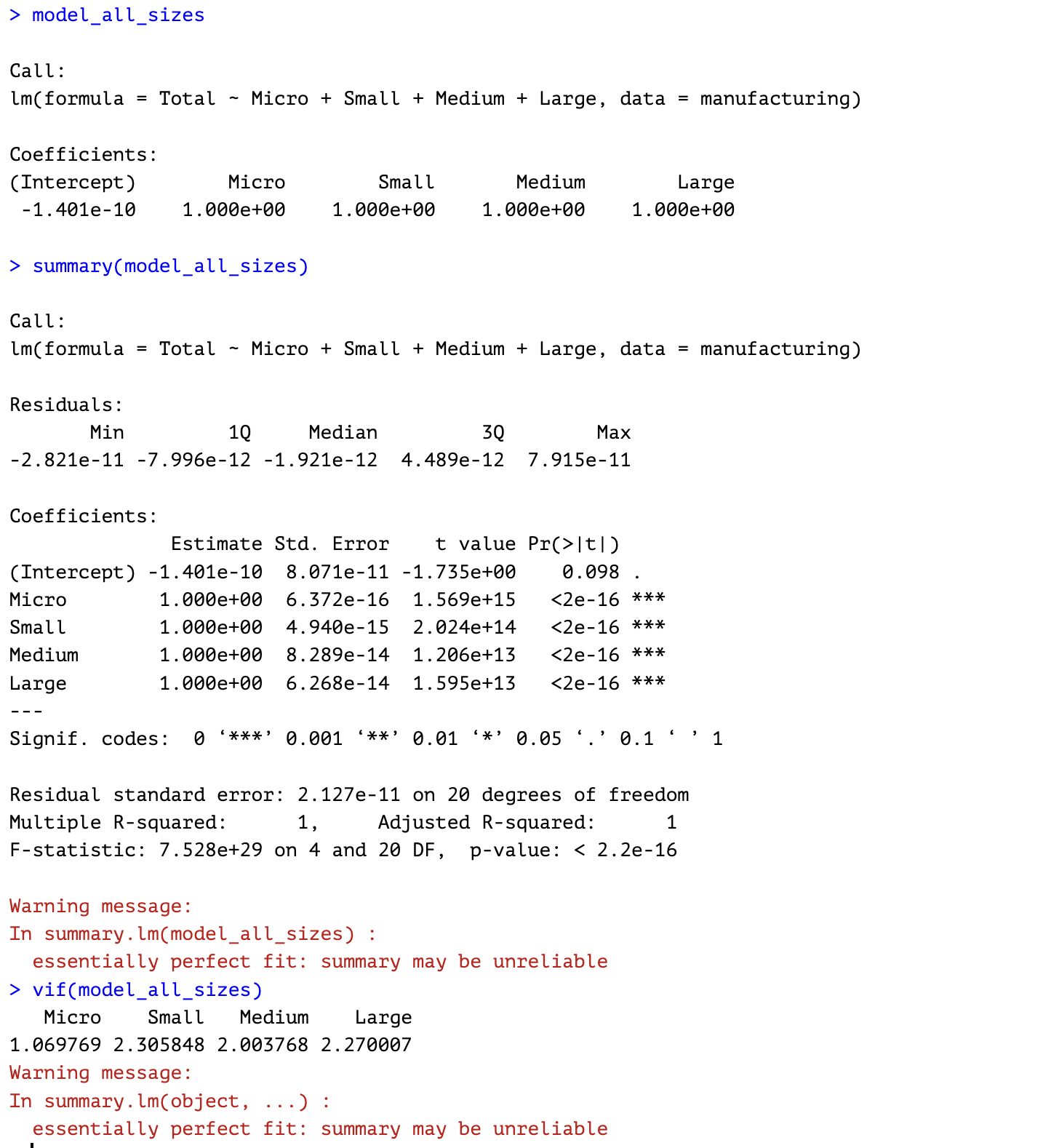
1. Comparison of Model Assessment Criteria Across Stepwise Regression Methods

| **Metric** | **Backward Selection** | **Forward Selection** | **Bidirectional Selection** |
| --- | --- | --- | --- |
| Model | Total ~ Micro + Small + Medium + Large | Total ~ Micro + Small + Large + Medium | Total ~ Micro + Small + Medium + Large |
| MSE (Mean Squared Error) | 9.06E-21 | 0.0 (final step: perfect fit) | 0.0 (final step: perfect fit) |
| RMSE (Root MSE) | 3.01E-11 | 0.0 (very small) | 0.0 (very small) |
| R-squared | 1.00 (Perfect Fit) | 1.00 (Perfect Fit) | 1.00 (Perfect Fit) |
| Adjusted R-squared | 1.00 (Perfect Fit, but unreliable) | 1.00 (Perfect Fit, but unreliable) | 1.00 (Perfect Fit, but unreliable) |
| AIC | -1224.33 | 365.83 (first step, Micro only) | 447.35 (initial null model) |
| Final Model | Total ~ Micro + Small + Medium + Large | Total ~ Micro + Small + Large + Medium | Total ~ Micro + Small + Medium + Large |

1. Performance of the three stepwise regression methods.

The final model from the stepwise regression retained all predictors, suggesting that each variable significantly contributes to explaining the total number of listed establishments. However, the model has already been identified as overfitted and therefore unreliable, highlighting the need to explore alternative modeling approaches.

Before exploring alternative models, a final check will be conducted to calculate the Variance Inflation Factor (VIF) to check for multicollinearity among the predictors that may help explain the results.



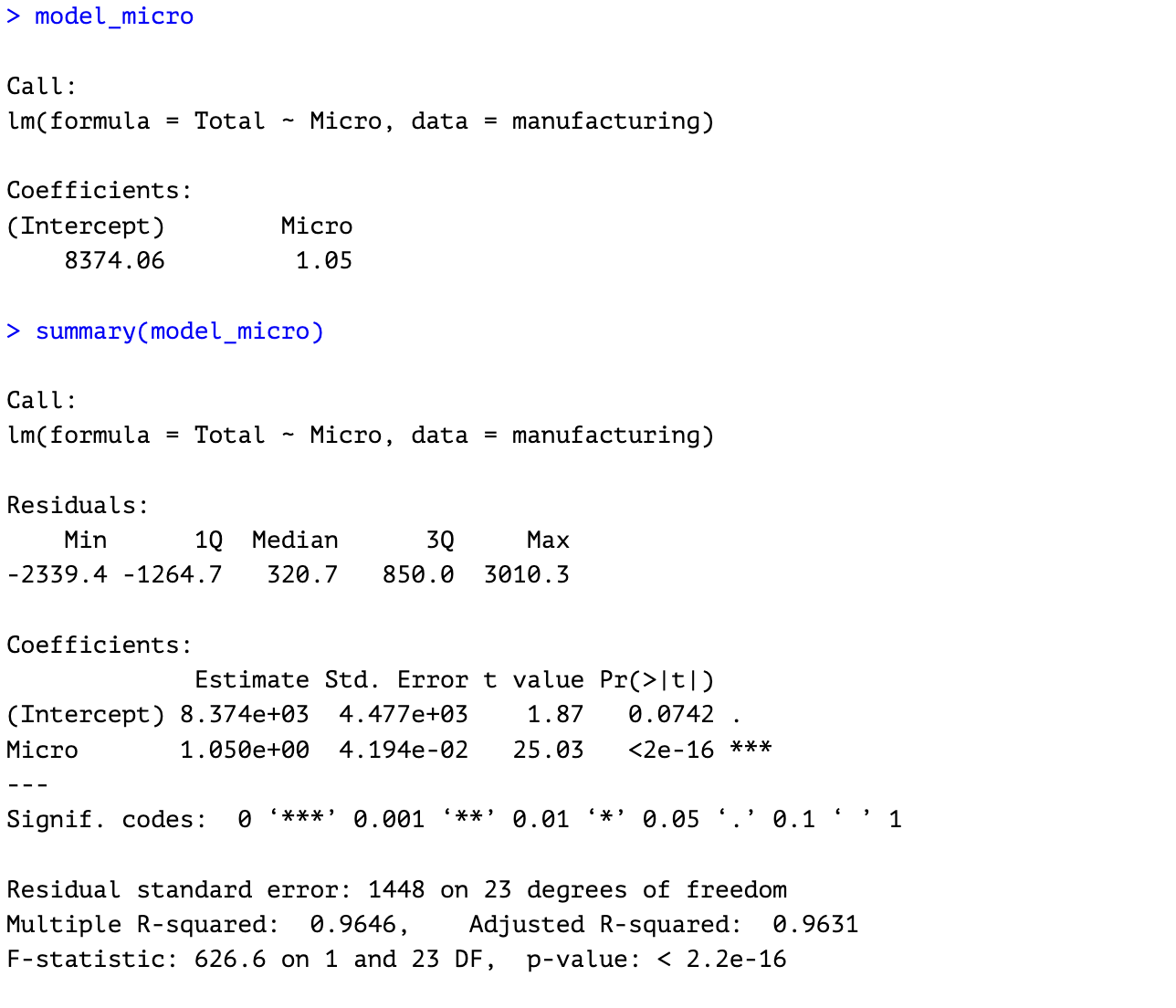
1. Checking for multicollinearity using *vif()* function in R.
2. Vif Values Of Predictors

| **Predictor** | **VIF Value** | **Interpretation** |
| --- | --- | --- |
| Micro | 1.07 | No multicollinearity concern |
| Small | 2.31 | Low multicollinearity |
| Medium | 2 | Low multicollinearity |
| Large | 2.27 | Low multicollinearity |

1. Summary and interpretation of VIF results.

The results show that there is no multicollinearity to low multicollinearity with the predictors. Yet, the model is still not reliable due to overfitting.

In this section, we examine our Exploratory Data Analysis results, which show that the Micro category exhibits the strongest positive linear relationship with the response variable, Total.



1. Summary of Selected Model, using Micro as predictor for Total.

The linear regression equation based on the selected model is:

*Total Establishments = 8,374.06 + 1.05 × (Micro Establishments)*

Which means that for every additional micro establishment in the Philippines, the total number of listed manufacturing establishments increases by about 1.05. When there are (hypothetically) zero micro establishments, the model predicts 8,374 total establishments—an artifact here, since in reality micro firms are always present.

## *Testing the Model*

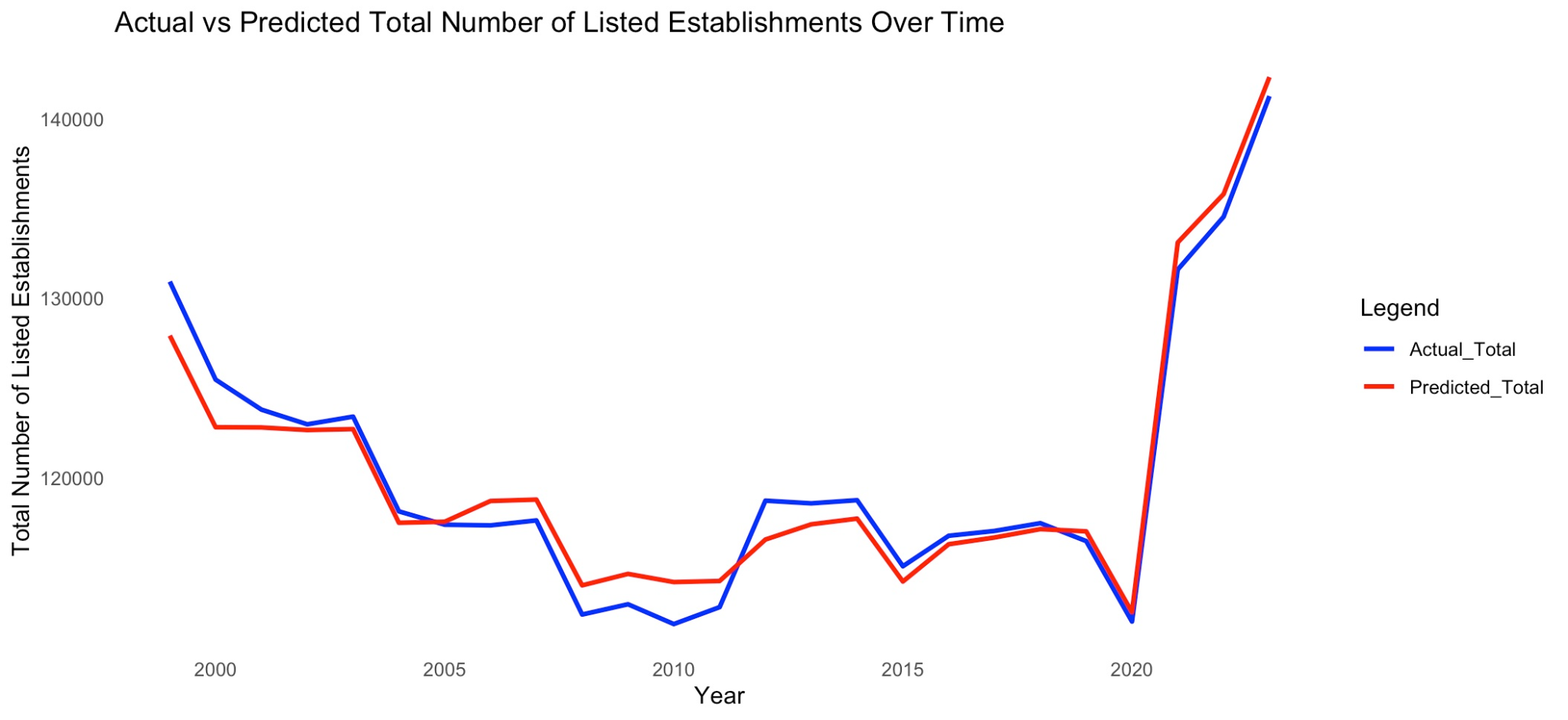
## Due to the limited availability of data, all observations were used in the creation of the model. To assess the reliability of the model, a prediction method will be employed, where the predicted values will be compared to the actual observed values.

## This will allow us to evaluate the accuracy and performance of the model in forecasting the total number of listed establishments in the Philippines' manufacturing sector.

1. Actual Values, Predicted Total, & Percent Error

| **Year** | **Actual Total** | **Predicted Total** | **Percent Error** |
| --- | --- | --- | --- |
| 1999 | 130,931 | 127,921 | 2.30% |
| 2000 | 125,467 | 122,815 | 2.11% |
| 2001 | 123,795 | 122,802 | 0.80% |
| 2002 | 122,977 | 122,656 | 0.26% |
| 2003 | 123,406 | 122,710 | 0.56% |
| 2004 | 118,127 | 117,490 | 0.54% |
| 2005 | 117,382 | 117,548 | -0.14% |
| 2006 | 117,346 | 118,704 | -1.16% |
| 2007 | 117,622 | 118,782 | -0.99% |
| 2008 | 112,377 | 114,003 | -1.45% |
| 2009 | 112,950 | 114,642 | -1.50% |
| 2010 | 111,846 | 114,185 | -2.09% |
| 2011 | 112,789 | 114,246 | -1.29% |
| 2012 | 118,722 | 116,556 | 1.82% |
| 2013 | 118,572 | 117,405 | 0.98% |
| 2014 | 118,749 | 117,723 | 0.86% |
| 2015 | 115,068 | 114,218 | 0.74% |
| 2016 | 116,766 | 116,294 | 0.40% |
| 2017 | 117,035 | 116,670 | 0.31% |
| 2018 | 117,468 | 117,137 | 0.28% |
| 2019 | 116,470 | 117,012 | -0.47% |
| 2020 | 111,988 | 112,502 | -0.46% |
| 2021 | 131,604 | 133,113 | -1.15% |
| 2022 | 134,542 | 135,807 | -0.94% |
| 2023 | 141,266 | 142,326 | -0.75% |

1. Comparison of actual vs predicted values of Total Listed Establishments using the selected model.



1. Time Series Graph, comparing the values of predicted data vs actual data (Total Listed Establishments)

Figures 17 and 18 display the comparison on actual value vs the predicted value. In evaluating the forecasting model's performance, the average percent error was calculated to be approximately -0.02%. This indicates that, on average, the model's predictions slightly overestimate the actual values. The near-zero average error reflects the model's high accuracy in forecasting, with only minimal bias. This insight is valuable for understanding the model's predictive behavior and can guide further refinements to enhance forecasting precision.

# Discussion of Findings

This section delves into the implications of the research findings and revisits the study's hypotheses. By interpreting the results, we aim to understand their significance and how they contribute to the existing body of knowledge. Additionally, we will assess whether the findings support or challenge the initial hypotheses, providing a comprehensive analysis of the study's outcomes

## *Revisting the Hypotheses of the Study*

***Ho1:*** *The micro-scale manufacturing sector (1–9 employees) comprises the majority of manufacturing establishments in the Philippines, reflecting the dominance of small-scale industrial activities.*

The research results confirm that the initial assumption was correct, as shown in Figure 6. This means that the model's predictions were accurate and aligned with the actual data, supporting the hypothesis made at the beginning of the study.​

***Ho2:*** *The small-scale manufacturing sector (10–99 employees) shows a steady presence across regions, suggesting its role as a backbone of local industrial development.*

The research results indicate that small-scale manufacturing businesses have a minimal impact on the overall growth of all manufacturing establishments. This conclusion is supported by Figures 6 and 7, which show a correlation value of only 0.41. A correlation of 0.41 suggests a weak positive relationship, meaning that as small-scale manufacturing establishments increase, the growth in the total number of manufacturing establishments also increases slightly, but the effect is not strong.

***Ho3:*** *The number of medium-scale manufacturing establishments (100–199 employees) remains limited, indicating potential barriers to enterprise growth beyond the small-scale level.*

The research findings indicate that the number of medium-scale manufacturing businesses (those with 100–199 employees) has decreased over time. This trend is clearly depicted in Figure 3, which shows a consistent decline in establishments within this category.

***Ho4:*** *The large-scale manufacturing sector (200 or more employees) accounts for the fewest number of establishments, emphasizing the challenges in scaling operations to a large employment size.*

The research findings indicate that large-scale manufacturing businesses, defined as those with 200 or more employees, represent a relatively small portion of the total manufacturing establishments. With an average of 1,078.12 establishments, they have the second-lowest number, following medium-scale establishments, which average 993.68 establishments. This is supported by the descriptive summary statistics of the data set included in the appendix of this research.

***Ho5:*** *The overall number of manufacturing establishments in the Philippines has been declining over time, suggesting a downward trend in the growth of the sector.*

The research findings indicate that the number of manufacturing establishments in the Philippines declined from 1999 to 2020. However, starting in 2020, there was an upward spike in the data, suggesting a period of stagnant growth rather than a continued decline. This is illustrated in Figure 2 of this research.

1. Summary of Hypotheses Evaluation

| **Hypothesis** | **Evaluation** |
| --- | --- |
| *Ho1* | True |
| *Ho2* | Partially True |
| *Ho3* | True |
| *Ho4* | False |
| *Ho5* | Partially False |

## *Summary and Significance of Findings in the Context of the Domain*

This research has identified both the strengths and weaknesses in the dynamics of listed establishments within the Philippine manufacturing sector. A significant highlight is the noticeable increase in the number of listed establishments from 2020 onwards, indicating positive growth for the sector. If this momentum is maintained, it could substantially contribute to the Philippines' economic progress, including its GDP. However, the data also reveals that, despite recent advancements, the sector experienced a prolonged period of decline prior to 2020. Furthermore, three out of the four categories lag significantly behind the micro-size category, suggesting that micro businesses in the country face challenges in scaling and expanding their operations.

While the number of listed establishments serves as a key indicator of the manufacturing sector's performance, its impact extends beyond mere numbers. A decline in these establishments can signal a reduction in production levels, employment opportunities, and overall economic contribution. Such a downturn may adversely affect the sector's contribution to the Philippines' economic progress, including its Gross Domestic Product (GDP).

Therefore, monitoring and addressing the dynamics of listed establishments is crucial for sustaining the sector's role in national economic development.​

# Conclusion & Reflection

This section presents the researcher's reflections on the stepwise regression process employed in this study, rather than focusing on the findings related to the research problem statement.

While stepwise regression is a widely used method for model selection, it has been critiqued for potential limitations, such as the risk of overfitting and the exclusion of important variables. This exact critique on overfitting is what occurred in this research, leading to the decision not to use the model obtained from the stepwise regression. Overfitting may cause the model to perform well on the training data but poorly on new, unseen data. This undermines the model's generalizability and predictive accuracy.

A significant issue with the dataset in this research pertains to the relationship between the response variable and the predictors. The response variable, "Total," is an arithmetic sum of the predictor variables representing different size categories. This inherent correlation among predictors led to overfitting, despite the Variance Inflation Factor (VIF) tests indicating no multicollinearity.

Nonetheless, stepwise regression remains a valuable method for identifying key predictors, particularly when the response variable is not a direct result of arithmetic operations involving the predictors. In such cases, the risk of overfitting is mitigated, allowing the model to more accurately capture the underlying relationships between variables.

Therefore, it is advisable to apply stepwise regression when the response variable is independent of the predictors, ensuring more reliable and interpretable results.​

# Acknowledgements

This following entities have been instrumental in the completion of this report:

*Rodolfo C. Raga Jr., PhDCS* – course instructor.

*Innovators and Technologists* – responsible for developing the tools and platforms that made research, data analysis, and visualization possible.

*Researchers and Academics* – for their dedication to expanding human knowledge. Their studies and insights have provided the foundation and direction for this work.

*Online Resources & Communities* – for making information accessible to all.

*ChatGPT* – In the absence of a human collaborator, the artificial intelligence tool has served as the researcher's partner. While AI-powered tools assisted in various tasks, all decision-making in this research was performed by the researcher particularly on analyzing and translating results into digestible insights.

Appendix

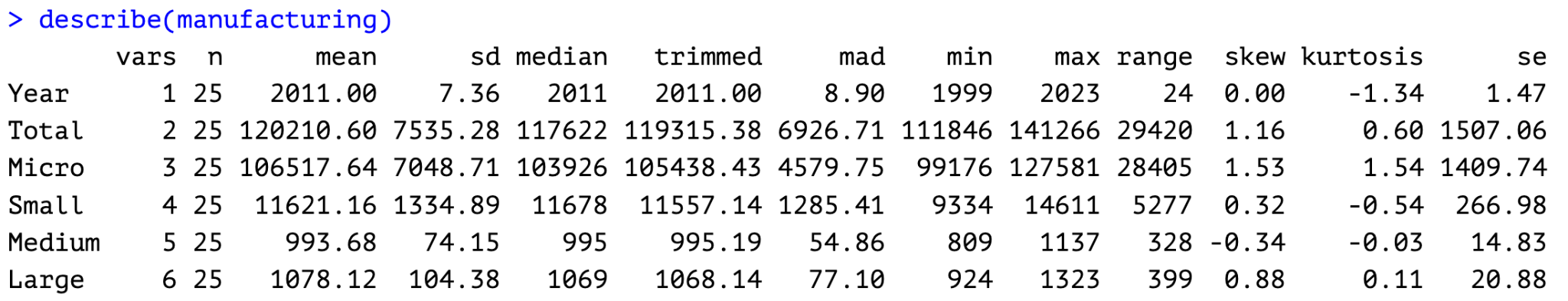


Fig. A1. Descriptive Statistics of the Data Set

References

1. D. Su and Y. Yao, Manufacturing as the Key Engine of Economic Growth for Middle-Income Economies, Asian Development Bank Institute (ADBI) Working Paper No. 573, May 2016. [Online]. Available: <https://www.adb.org/publications/manufacturing-key-engine-economic-growth-middle-income-economies/>
2. C. Balita, "Manufacturing sector in the Philippines - statistics & facts," Statista, Dec. 3, 2024. [Online]. Available: https://www.statista.com/topics/11457/manufacturing-in-the-philippines/. [Accessed: Apr. 20, 2025].
3. R. M. Aldaba, *The Philippine Manufacturing Industry Roadmap: Agenda for New Industrial Policy, High Productivity Jobs, and Inclusive Growth*, Philippine Institute for Development Studies, Discussion Paper Series No. 2014-32, Jun. 2014. [Online]. Available: <https://pids.gov.ph/publication/discussion-paper-series/the-philippine-manufacturing-industry-roadmap-agenda-for-new-industrial-policy-high-productivity-jobs-and-inclusive-growth>
4. Philippine Statistics Authority, "Seasonally Adjusted Production Index," *psa.gov.ph*, [Online]. Available:<https://psa.gov.ph/statistics/manufacturing/seasonally-adjusted-production-index>. [Accessed: Apr. 20, 2025].
5. Statista, Philippines: number of manufacturing establishments 2022, Statista Research Department. [Online]. Available: <https://www.statista.com/statistics/1331181/philippines-number-of-manufacturing-establishments/>. [Accessed: Apr. 20, 2025].
6. Philippine Statistics Authority, *Establishments by Region, Section, and Employment Groupings (Micro, Small, Medium, Large), Philippines: 1993–2023*. Google Sheets, Apr. 2025. [Online]. Available:<https://docs.google.com/spreadsheets/u/0/d/1YJEZqBIeoW3N38eg3MlHkNX_tc3IXcrm/htmlview>