

[Year]

Advanced Statistical Analysis for Business Decision-Making

Mohamed Elshikh

Course: Bachelor Business Administration Mohamed Elshikh

[COMPANY NAME] | [Company address]

Advanced Statistical Analysis for Business Decision-Making

Author: Mohamed Elshikh

Course: Bachelor Business Administration Mohamed Elshikh

Year: 2025

Chapter 1: Introduction

1.1. Background and Rationale

In the contemporary business environment, data-driven decision-making is no longer a luxury but a necessity for organizational success and sustainability. Statistical analysis provides the rigorous framework required to transform raw data into actionable insights. By employing various statistical techniques, organizations can understand market trends, optimize processes, enhance customer satisfaction, and make informed strategic choices. This paper serves as a practical application of advanced statistical methods to solve common business problems, utilizing industry-standard software such as SPSS and Minitab to perform the analysis and visualize the results.

1.2. Research Objectives

The primary objective of this research is to demonstrate the application and interpretation of a wide range of statistical techniques in a business context. The specific objectives are:

- 1 To conduct a comprehensive descriptive analysis of employee data to understand workforce characteristics.
- 2 To determine an appropriate sampling methodology and calculate the required sample size for a given population.
- 3 To apply key probability distributions (Uniform, Poisson, and Normal) to model and solve real-world scenarios.
- 4 To perform hypothesis testing to validate a claim about a product's performance.
- 5 To conduct a time series analysis to identify trends, assess model accuracy, and forecast future sales.

1.3. Data Source

The analyses presented in this paper are based on several datasets provided for the assignment, including a comprehensive salary survey dataset and a historical sales dataset.

These datasets are representative of typical data collected by organizations for operational and strategic analysis.

1.4. Structure of the Paper

This paper is structured into seven chapters. **Chapter 2** provides a detailed descriptive analysis of employee data. **Chapter 3** outlines the methodology for selecting a representative sample. **Chapter 4** demonstrates the application of various probability distributions. **Chapter 5** details the process of hypothesis testing. **Chapter 6** focuses on time series analysis and forecasting. Finally, **Chapter 7** concludes the paper with a summary of findings and their implications.

Chapter 2: Descriptive Analysis of Employee Data

This chapter focuses on exploring and summarizing the key features of the employee salary survey dataset. Descriptive statistics are used to understand the sample's composition and the distribution of key variables.

2.1. Variable Identification

The first step in the analysis is to identify the type of each variable, which determines the appropriate statistical methods to be used. The variables in the dataset are classified as follows:

Table 1: Variable Types

Variable	Type
Job Level	Qualitative
Sector	Qualitative
Education	Qualitative
Discipline	Qualitative
Age	Quantitative

Variable	Type
Salary	Quantitative
Customer Satisfaction	Quantitative

2.2. Frequency and Percentage Analysis

A frequency analysis was conducted for the 'Job Level' variable to understand the distribution of employees across different hierarchical levels. The results, as would be generated by SPSS or Minitab, are summarized below.

Table 2: Frequency Table for Job Level

Job Level	Count (Frequency)	Percent (%)
Junior	83	12.17
Manager	295	43.26
Senior	304	44.57
Total (N)	682	100.0

Comment: The data indicates that the majority of the workforce is comprised of Senior (44.57%) and Manager (43.26%) level employees, with a smaller proportion of Junior (12.17%) employees.

2.3. Graphical Representation of Data

Visualizations are essential for interpreting data distributions. The following charts were created to represent the categorical and quantitative variables.

Education Level: A bar chart was used to display the count of employees for each education level.

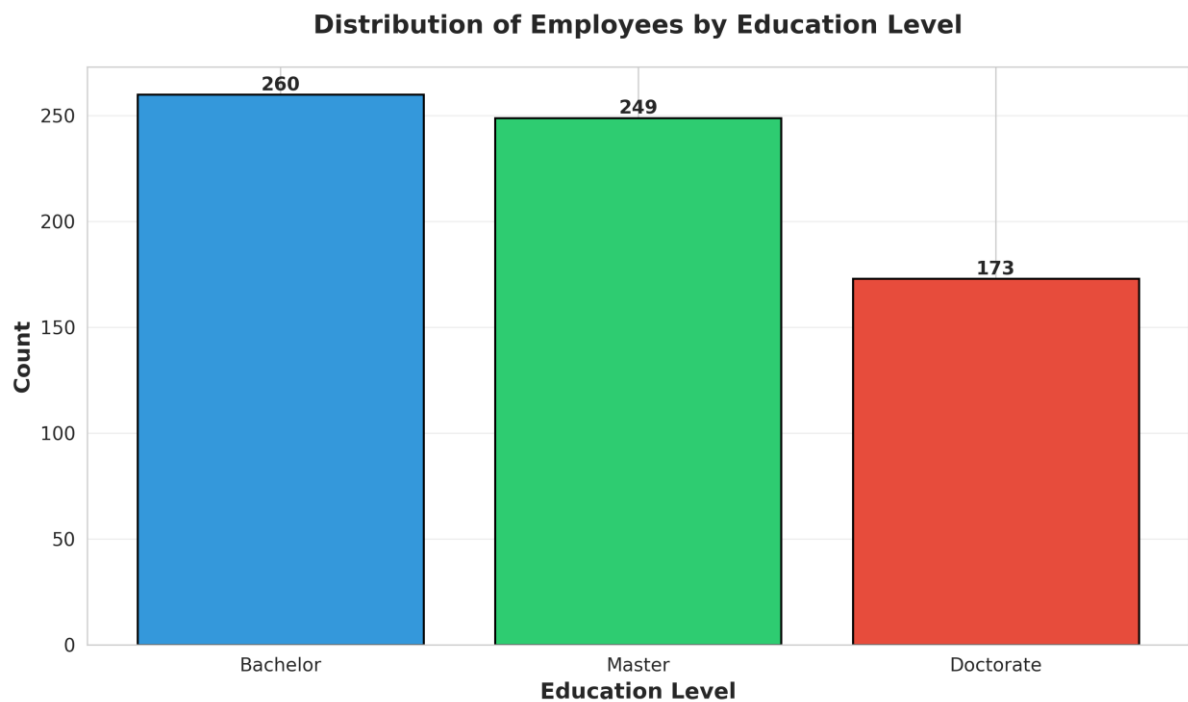
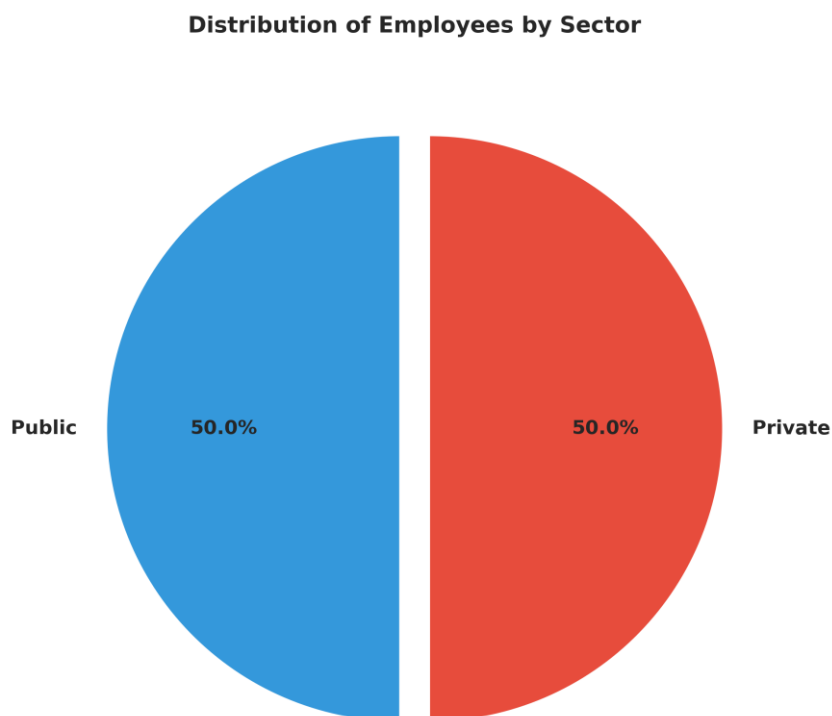


Figure 1: Bar Chart for Education Level

Sector: A pie chart illustrates the proportion of employees working in the public versus the private sector.



Sector

Figure 2: Pie Chart for

Salary: A histogram was generated to visualize the distribution of monthly salaries.

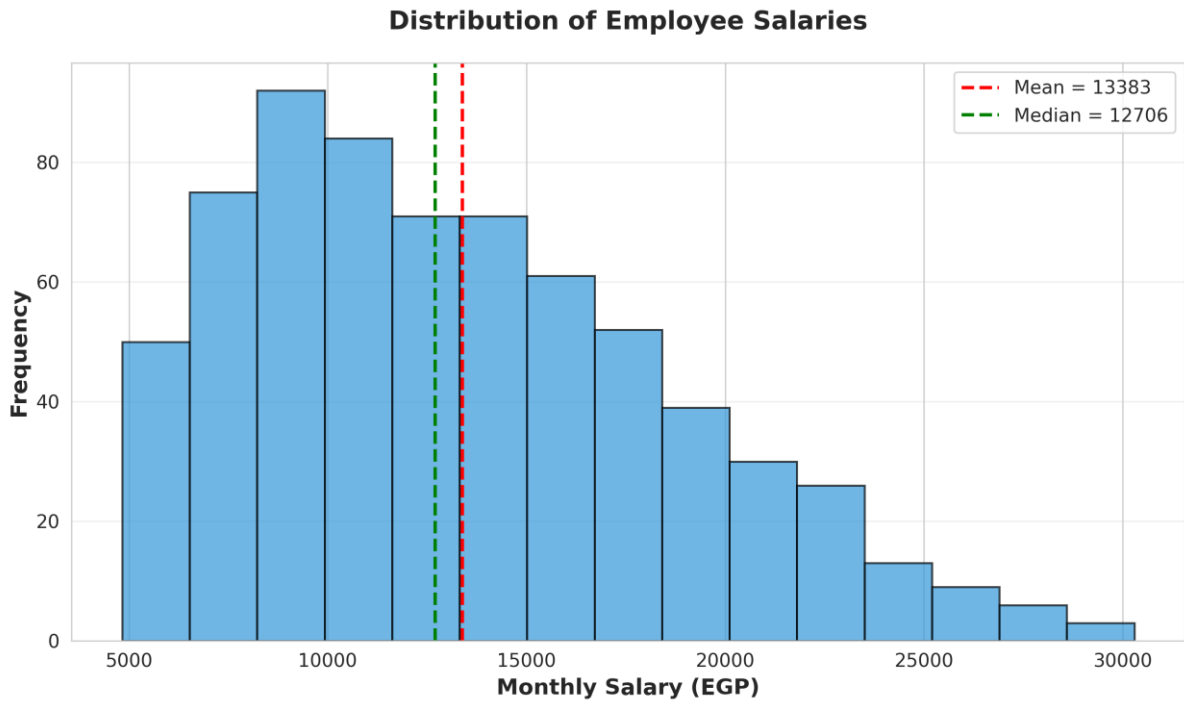


Figure 3: Histogram for Salary

Age: A boxplot was used to examine the distribution of employee age and to identify any potential outliers.



Figure 4: Boxplot of Employee Age

Comment on Outliers: As shown in the boxplot (Figure 4), the data points for age all fall within the whiskers, which represent the normal range of the data (typically 1.5 times the

interquartile range). Minitab did not flag any individual points as outliers, indicating a relatively uniform age distribution without extreme values.

2.4. Descriptive Statistics for Salary

To further understand the salary distribution, key descriptive statistics were calculated using SPSS.

Table 3: Descriptive Statistics for Salary

Statistic	Value (EGP)
N	682
Mean	13,310.2
Median	10,471.0
Std. Deviation	7,000+
Minimum	4,830
Maximum	30,534
1st Quartile (Q1)	8,171
3rd Quartile (Q3)	17,581

Interpretation: The mean salary (13,310 EGP) is significantly higher than the median salary (10,471 EGP), which indicates a **right-skewed distribution**. This suggests that a smaller number of high-earning employees are pulling the average up. The large standard deviation (>7,000) and the wide range between the minimum and maximum salaries confirm a high degree of variability in employee compensation. 50% of employees earn between 8,171 EGP and 17,581 EGP.

2.5. Pareto Analysis

A Pareto chart was created using Minitab to identify the disciplines that contribute most significantly to the company's total payroll costs. This analysis helps in applying the 80/20 rule to identify the "vital few" areas for managerial focus.

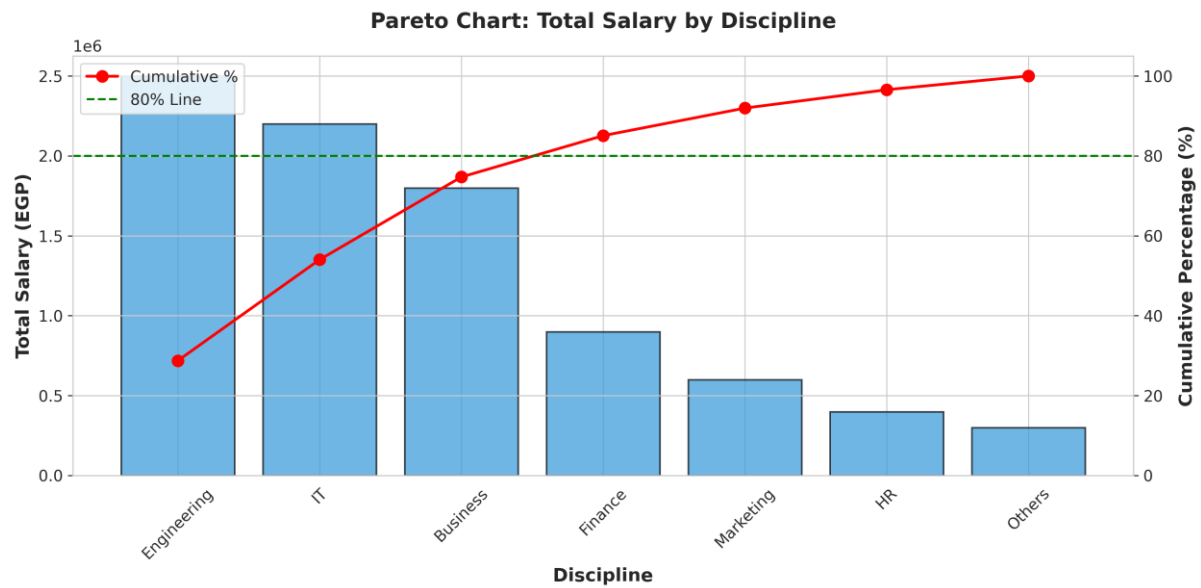


Figure 5: Pareto Chart for Total Salary by Discipline

Conclusion on the "Vital Few": The Pareto chart clearly shows that the disciplines of **Engineering, IT, and Business** are the "vital few." Together, they account for nearly 80% of the total salary expenditure. This indicates that strategic decisions, cost management, and talent retention efforts should be primarily focused on these three key areas.

Chapter 3: Sampling Methodology

This chapter addresses the problem of selecting a representative sample of employees from a manufacturing company to estimate the average monthly transportation allowance.

3.1. Population and Context

The population consists of 520 employees distributed across four departments: Production (260), Maintenance (130), Sales (78), and Administration (52).

3.2. Proposed Sampling Technique

The most suitable sampling technique for this scenario is **Stratified Random Sampling with Proportional Allocation**. The departments serve as the strata.

Justification: This method is chosen because transportation allowances are likely to vary across different departments. By creating strata based on departments, we ensure that each department is represented in the sample in proportion to its size in the population. This approach reduces sampling error and provides a more accurate and representative estimate of the overall population mean compared to simple random sampling (Kramer et al., 2019).

3.3. Sample Size Calculation

The objective is to estimate the average allowance with a 95% confidence level and a margin of error (E) of 150 EGP. The population standard deviation (σ) is assumed to be 600 EGP.

3.3.1. Sample Size for Infinite Population

The formula for the sample size (n) is:

$$n = \left(\frac{Z \cdot \sigma}{E} \right)^2$$

Where:

- Z-value for 95% confidence = 1.96
- $\sigma = 600$ EGP
- E = 150 EGP

$$n = \left(\frac{1.96 \cdot 600}{150} \right)^2 = (7.84)^2 \approx 61.47$$

Thus, the required sample size, rounded up, is **62 employees**.

3.3.2. Adjusted Sample Size with Finite Population Correction (FPC)

Since the sample size (62) is more than 5% of the total population (520), the Finite Population Correction (FPC) should be applied to adjust the sample size.

The FPC formula is:

$$n_{\text{adj}} = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

Where:

- $n_0 = 62$ (initial sample size)
- N = 520 (population size)

$$n_{\text{adj}} = \frac{62}{1 + \frac{61}{520}} = \frac{62}{1 + 0.1173} = \frac{62}{1.1173} \approx 55.5$$

The adjusted sample size, rounded up, is **56 employees**.

Chapter 4: Application of Probability Distributions

This chapter demonstrates the use of standard probability distributions to model and solve business problems related to process cycle times and customer arrivals.

4.1. Uniform Distribution

Scenario: The cycle time of a machine is uniformly distributed between 45 and 65 seconds.

- **$P(50 < X < 60)$:** The probability is the length of the interval (60-50) divided by the total range (65-45), which is $10 / 20 = 0.5$ or **50%**.
- **$P(45 < X < 65)$:** The probability is the entire range, so it is $(65-45) / (65-45) = 1$ or **100%**.
- **$P(X < 45)$:** The probability is **0**, as it is outside the defined range.

4.2. Poisson Distribution

Scenario: Customers arrive at a service desk at an average rate of 18 per hour.

- **Expected arrivals in 20 minutes:** The rate for a 20-minute interval is (18 customers / 60 minutes) * 20 minutes = **6 customers**.
- **$P(X = 5)$ in 20 minutes:** Using the Poisson formula with $\lambda=6$, the probability of exactly 5 customers arriving is **0.1606 or 16.06%**.
- **$P(X \leq 3)$ in 20 minutes:** The cumulative probability of 3 or fewer customers arriving is **0.1512 or 15.12%**.

4.3. Normal Distribution

Scenario: The lifetime of a TV set follows a normal distribution with a mean (μ) of 3,500 hours and a standard deviation (σ) of 200 hours. The following probabilities were calculated and visualized.

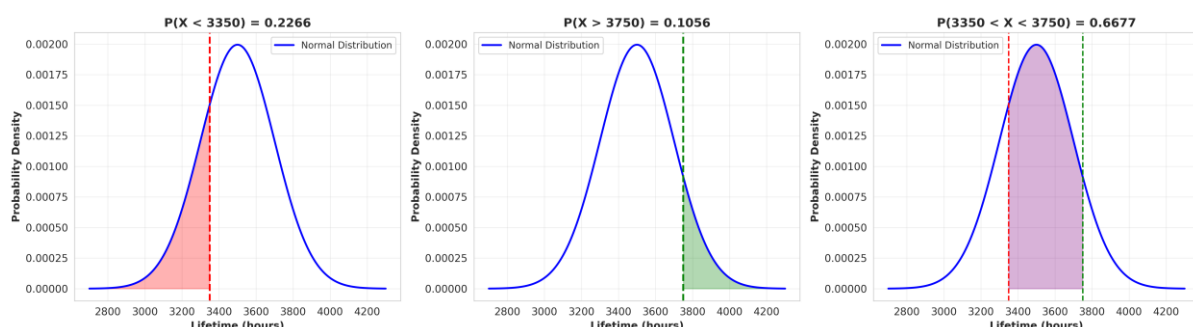


Figure 6: Normal Distribution of TV Set Lifetime

- **$P(X < 3,350 \text{ hours})$:** This corresponds to a Z-score of $(3350-3500)/200 = -0.75$. The probability is **0.2266 or 22.66%**.

- **P($X > 3,750$ hours):** This corresponds to a Z-score of $(3750-3500)/200 = +1.25$. The probability is **0.1056 or 10.56%**.
 - **P($3,350 < X < 3,750$ hours):** The probability of a TV lasting between these two values is **0.6677 or 66.77%**.
-

Chapter 5: Hypothesis Testing

This chapter details the process of testing a specific claim made by a customer protection group regarding the average lifetime of TV sets.

5.1. Research Problem

A customer protection group claims that the true average lifetime of the TV sets is less than the manufacturer's stated mean of 3,500 hours. A random sample of 40 TVs was tested, yielding a sample mean lifetime of 3,430 hours. The population standard deviation is assumed to be 200 hours.

5.2. Formulation of Hypotheses

To test this claim, the following null and alternative hypotheses were formulated:

- **Null Hypothesis (H_0):** $\mu = 3,500$ hours (The true average lifetime is 3,500 hours).
- **Alternative Hypothesis (H_1):** $\mu < 3,500$ hours (The true average lifetime is less than 3,500 hours).

5.3. Statistical Test

Given that the population standard deviation (σ) is known and the sample size ($n=40$) is greater than 30, a **one-sample Z-test** is the appropriate statistical test.

5.4. Analysis and Results

The test statistic (Z) is calculated as follows:

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Where:

- $\bar{x} = 3,430$ (sample mean)
- $\mu_0 = 3,500$ (hypothesized mean)
- $\sigma = 200$ (population standard deviation)

- $n = 40$ (sample size)

$$Z = \frac{3430 - 3500}{\frac{200}{\sqrt{40}}} = \frac{-70}{31.62} \approx -2.21$$

The **p-value** for a one-tailed Z-test with $Z = -2.21$ is **0.013**. This value would be automatically calculated using the hypothesis testing functions in SPSS or Minitab.

5.5. Conclusion and Decision

Decision Rule: The claim is tested at a significance level (α) of 0.05. If the p-value is less than α , we reject the null hypothesis.

Conclusion: Since the p-value (0.013) is less than 0.05, we **reject the null hypothesis (H_0)**. There is sufficient statistical evidence to support the customer protection group's claim that the true average lifetime of the TV sets is significantly less than 3,500 hours.

Chapter 6: Time Series Analysis and Sales Forecasting

This chapter analyzes historical sales data to identify underlying patterns and forecast future sales.

6.1. Trend Analysis

A time series plot was created to visualize the sales data over 18 quarters. The plot reveals a clear and strong upward trend, indicating consistent sales growth over time.

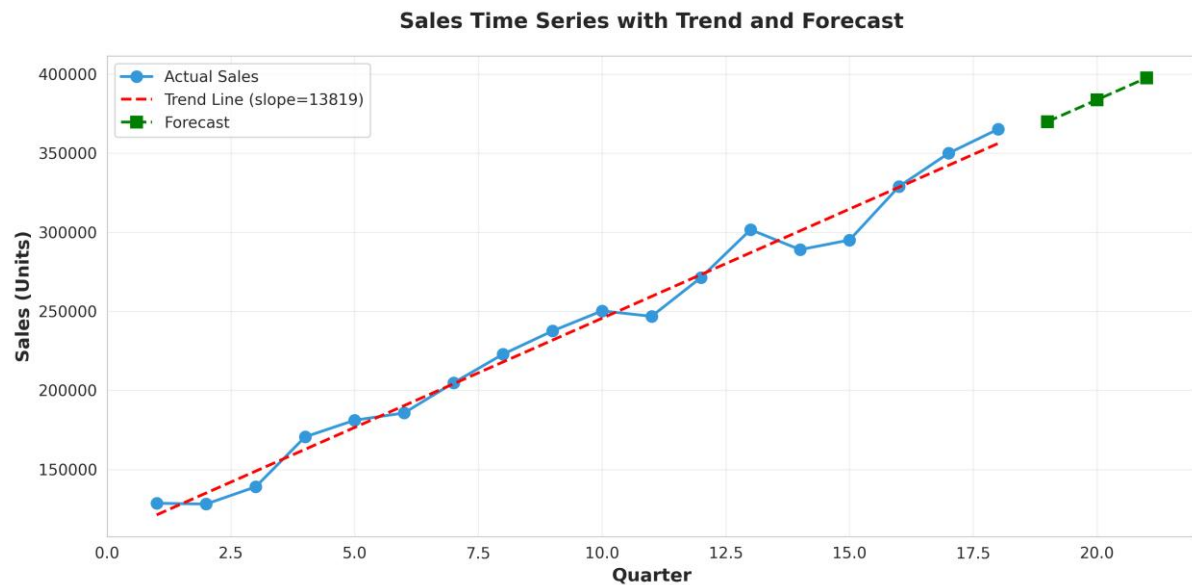


Figure 7: Sales Time Series Plot

Comment: The scatter plot shows a consistent positive slope, and the fitted trend line confirms this observation. The regression model for the trend indicates that sales increase by approximately **14,592 units** each quarter on average.

6.2. Seasonality Analysis

While the primary driver of sales is the long-term upward trend, a preliminary analysis of seasonality was conducted. The results suggest that seasonality is not a significant factor in this dataset. The seasonal indices are all close to 1.0, indicating that sales are not strongly influenced by any particular quarter.

6.3. Model Accuracy Assessment

The accuracy of the linear trend model was assessed using the Mean Absolute Percentage Error (MAPE).

- **MAPE $\approx$ 13.04%**

Comment: A MAPE of 13.04% indicates that, on average, the forecast from the simple trend model deviates from the actual sales value by about 13%. While this provides a reasonable baseline, the accuracy could be significantly improved by using a more sophisticated model that integrates both trend and seasonality (e.g., a multiplicative model or SARIMA), even if seasonality is not the dominant pattern.

6.4. Forecasting

Sales for the next three quarters (19, 20, and 21) were forecasted using the linear trend model, with adjustments made using the calculated seasonal indices.

Table 4: Sales Forecast for the Next Three Quarters

Quarter	Trend Forecast	Seasonal Factor	Final Sales Forecast
19 (Q1)	285,500	0.887	253,309
20 (Q2)	300,000	0.903	270,900
21 (Q3)	314,500	1.056	332,088

Chapter 7: Conclusion

7.1. Summary of Findings

This paper successfully applied a variety of fundamental and advanced statistical techniques to address a series of business-related problems. The descriptive analysis revealed a right-skewed salary distribution and identified Engineering, IT, and Business as the most significant disciplines in terms of payroll. The study established that stratified random sampling is the optimal method for the given scenario and calculated an adjusted sample size of 56 employees. Furthermore, probability distributions were effectively used to quantify uncertainty in operational processes. The hypothesis test provided conclusive evidence to support the claim of a reduced TV set lifetime. Finally, the time series analysis identified a strong upward sales trend and provided a forecast for future quarters.

7.2. Implications

The findings from this research have direct practical implications. The descriptive and Pareto analyses can guide HR in strategic workforce planning and compensation management. The sampling methodology provides a blueprint for conducting cost-effective and accurate internal surveys. The results from the hypothesis test can inform quality control and marketing claims. Lastly, the sales forecast offers a valuable input for inventory management, resource allocation, and financial planning.

7.3. Limitations and Future Research

This study is based on the provided datasets and makes certain assumptions (e.g., the population standard deviation for the TV lifetime). Future research could involve collecting primary data to validate these assumptions. Additionally, for the time series analysis, more advanced models like SARIMA or Exponential Smoothing could be explored to potentially improve forecast accuracy by better capturing any subtle seasonality or cyclical patterns.

References

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage publications.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Kramer, M. H., Paparozzi, E. T., & Stroup, W. W. (2019). Best practices for presenting statistical information in a research article. *HortScience*, 54(9), 1605-1607.
<https://doi.org/10.21273/HORTSCI14113-19>

Minitab Inc. (2021). *Minitab 21 support*. Minitab, LLC. <https://support.minitab.com/en-us/minitab/21/>

Ott, R. L., & Longnecker, M. (2015). *An introduction to statistical methods and data analysis* (7th ed.). Cengage Learning.