
📊 Statistical Business Analysis Report

1. Introduction

This report analyzes sales data across multiple products, regions, and customers. The dataset includes **Date, Product, Quantity, Price, Customer_ID, Region, and Total_Sales**. The goal is to evaluate sales performance, identify regional differences, and assess the impact of product pricing and quantity on overall revenue.

---

2. Descriptive Statistics

- **Average Sales:** ₹123,650
- **Median Sales:** ₹97,956
- **Mode Sales:** ₹6,540
- **Standard Deviation:** ₹100,161

Interpretation:
Sales values vary widely, with a high standard deviation indicating significant differences between transactions. The mean is higher than the median, suggesting the presence of large sales outliers (e.g., high-value laptop purchases).

---

3. Data Distribution

- Histogram and density plots show **Total_Sales** is **not normally distributed**.
- Shapiro-Wilk test statistic = 0.899, p-value ≈ 1.26e-06 → Reject normality assumption.

Interpretation:
Sales data is skewed, likely due to high-value product categories (e.g., laptops, phones). This skewness should be considered when applying parametric tests.

---

4. Correlation Analysis

- **Quantity vs Total_Sales:** r = 0.69 (Strong Positive)
- **Price vs Total_Sales:** r = 0.65 (Moderate Positive)

Interpretation:
Both quantity and price strongly influence total sales. Quantity has a slightly stronger correlation, meaning sales volume is a key driver of revenue.

---

5. Hypothesis Testing

- **T-test (North vs South Sales):** t = 0.130, p = 0.897 → No significant difference.
- **ANOVA (North, South, East):** F = 0.051, p = 0.950 → No significant regional differences.

Interpretation:
Regional sales differences are not statistically significant. Sales performance is consistent across regions.

---

6. Confidence Intervals

- **95% CI for Average Sales:** ₹103,776 – ₹143,525
- **Margin of Error:** ± ₹19,874

Interpretation:
We are 95% confident that the true average sales lie within this range. The margin of error is relatively large, reflecting high variability in transaction values.

---

7. Regression Analysis

- **Model:** Total_Sales ~ Price + Quantity
- **$R^2 = 0.884$** → Price and Quantity explain 88% of sales variation.
- **Coefficients:**
    - Price: 4.61 ($p < 0.001$) → Each ₹1 increase in price adds ~₹4.61 to sales.
    - Quantity: 26,430 ($p < 0.001$) → Each additional unit sold adds ~₹26,430 to sales.

Interpretation:
The regression model is highly predictive. Quantity has a stronger impact than price, confirming that increasing sales volume is the most effective way to boost revenue.

---

8. Business Insights & Recommendations

- **Sales drivers:** Quantity and price are both significant predictors of revenue, with quantity being the stronger factor.
- **Regional consistency:** No significant differences across regions, suggesting uniform performance.
- **Strategic focus:**
    - Increase sales volume through promotions, bundling, and upselling.
    - Optimize pricing strategies carefully, as price changes significantly affect revenue.
    - Since regional differences are minimal, focus on product-level strategies rather than region-specific campaigns.

---

9. Conclusion

This analysis confirms that **sales volume and product pricing are the primary drivers of revenue**, while regional differences are negligible. Future work should explore customer segmentation and product-level profitability to refine business strategies further.
