COMPREHENSIVE END-TO-END STATISTICS REPORT

Sales Performance & Decision Making Using Statistics

PART 1: Descriptive Statistics (Chapter 2)

Task 1.1 – Central Tendency

The median is the best measure of central tendency for revenue because:

The median is the most appropriate measure of central tendency for monthly revenue because it is not influenced by extreme values caused by seasonal spikes or promotional periods. While the mean provides an overall average, it may overestimate typical performance when outliers are present. The mode is not reliable for continuous revenue data, as exact values rarely repeat.

Task 1.2 – Dispersion

A high standard deviation indicates:
A high standard deviation in monthly revenue indicates that sales fluctuate significantly from month to month. This suggests low sales stability, making it difficult for the company to accurately forecast revenue, manage inventory, and plan cash flow. Large variations may be caused by seasonal demand, promotional campaigns, or market changes. Management should plan cash flow carefully due to volatility.

Task 1.3 – Shape of Distribution

Observations

- Histogram shows positively skewed revenue distribution. There are few very high revenue values
- Most transactions cluster at lower values
- Long right tail caused by high-value transaction

Skewness & Kurtosis

The Skewness is positive as it shows very few high revenue values
The Kurtosis is leptokurtic as there is a sharp peak and heavy tails
Revenue is not normally distributed, but skewed by high-performing sales periods.

PART 2: Data Visualization

Line Chart – Revenue Over Time

Reveals: Overall revenue fluctuates across dates
Possible seasonal or campaign-driven spikes
Business Insight:
Revenue is not constant, timing and strategy matter.

Bar Chart – Revenue by Store Type

Clear comparison between Online and Physical stores
Online stores clearly outperformed physical stores

Box Plot – Revenue by Region

Regional revenue spread
Outliers per region and regions with higher median performance
Identifies high-performing regions and unstable markets.

Scatter Plot – Marketing Spend vs Revenue

Direction of relationship (positive / weak / none)
Whether higher marketing investment aligns with higher revenue
This supports evaluation of marketing ROI.

PART 3: Sampling and Bias (Chapter 3)

Task 3.1 – Population vs Sample

- **Population**: All sales transactions across all stores and regions
- **Sample**: The dataset used for analysis

Task 3.2 – Sampling Bias

Bias Identified

- Urban-only sampling bias

Effect on Conclusions

- Overestimates revenue performance
- Ignores rural purchasing behavior
- Misleads expansion and marketing decisions

Better Sampling Method

- **Stratified random sampling**
  - Include urban & rural stores
  - Proportional representation by region

PART 4: Law of Large Numbers & CLT

Task 4.1 – Law of Large Numbers

As sample size increases (n = 10 → 500):

- Sample mean converges to population mean
- Variability decreases

 Interpretation: Larger datasets give more reliable revenue estimates.

Task 4.2 – Central Limit Theorem

- Distribution of **sample means (n = 30)** becomes approximately normal
- Holds even though original revenue distribution is skewed
- CLT states that the sampling distribution of the mean becomes normal as sample size increases.
  Why this matters:
   Enables valid t-tests and confidence intervals.

PART 5: Hypothesis Testing (Chapter 4)

Task 5.1 – Hypotheses

- H_0: μ_campaign = μ_no_campaign
- H_1: μ_campaign > μ_no_campaign
- **Test type:** One-tailed
- **Confidence level:** 95%
- **α:** 0.05

Task 5.2 – t-Test (Conceptual Outcome)

- If **p-value < 0.05** → Reject H_0
- If **p-value ≥ 0.05** → Fail to reject H_0

Interpretation
- A significant result supports the marketing strategy
- the marketing campaign significantly increases revenue,
- A non-significant result does not prove no effect — it may indicate low power

PART 6: Errors and Interpretation (Chapter 5)

Type I Error

- Rejecting H_0 when it is true
- Company thinks marketing works when it doesn't

Business impact:
 The company continues funding an ineffective campaign → wasted resources.

Type II Error

- Failing to reject H_0 when it is false
- Company misses a profitable strategy

Business impact:
 Company abandons a profitable strategy → lost revenue.

PART 7: Effect Size and Power

Task 7.1 – Effect Size (Cohen's d)
Interpretation

- **Small (~0.2)**: Minor practical impact
- **Medium (~0.5)**: Meaningful business difference
- **Large (≥0.8)**: Strong practical significance

 Key insight:
 Effect size explains how much revenue changes, not just whether it changes.
A statistically significant result with a small effect may not justify high marketing costs.

Task 7.2 – Power Discussion

Why an insignificant result may still matter

- Sample size too small
- High revenue variability
- Real effect exists but wasn't detected

Should the company collect more data?

Yes:
- Increases power
- Reduces Type II errors
- Improves confidence in decisions

8. Final Business Recommendations

- Use median revenue for reporting
- Expand sampling beyond urban stores
- Evaluate campaigns using both p-values and effect size
- Collect more data before rejecting strategic changes