

E-commerce Business Analysis Using Statistical Modeling Tools

Course Title: CS504049 – Business Intelligence (BI)

Instructor: Phuc H. Duong

Location: Tan Phong Campus

Context

An e-commerce company wants to analyze customer purchase patterns to improve its marketing strategy. The company has collected data on daily sales (in dollars) over the past 6 months. By applying statistical modeling tools, they aim to gain insights into customer behavior and optimize their promotional efforts.

Application of Statistical Modeling Tools

1. Mean (Average Sales)

- **Application:** The company calculates the mean daily sales to understand the average revenue generated on a typical day.
- **Advantages:** Provides a general sense of the company's sales performance over time.
- **Disadvantages:** The mean is sensitive to outliers, such as days with exceptionally high or low sales (e.g., holiday sales spikes), which might not represent typical sales.

2. Median (Middle Value of Sales)

- **Application:** To mitigate the impact of outliers, the company also calculates the median daily sales.
- **Advantages:** The median provides a more accurate reflection of central tendency when the data contains outliers.
- **Disadvantages:** It doesn't account for the spread of the data and may not capture the full sales distribution.

3. Mode (Most Frequent Sales Value)

- **Application:** The mode is used to find the most common sales amount during the observed period.
- **Advantages:** Helps identify the most typical sales figure, which can be useful for inventory and staffing purposes.
- **Disadvantages:** The mode may not always exist or may not be useful if the data has no repeating values or if multiple modes exist.

4. Variance (Measure of Spread in Sales)

- **Application:** The company computes the variance to understand the variability in daily sales.
- **Advantages:** Captures the degree of dispersion in sales, helping the business understand fluctuations in revenue.
- **Disadvantages:** Since variance is in squared units, it may be harder to interpret directly in terms of dollars.

5. Standard Deviation (Sales Variation in Dollar Terms)

- **Application:** The company uses the standard deviation to assess how much the sales deviate from the mean, in dollar terms.
- **Advantages:** It provides an easily interpretable measure of sales volatility, allowing the company to prepare for large fluctuations.
- **Disadvantages:** Like the variance, it is sensitive to outliers and may not fully reflect the data's spread if extreme sales values are present.

6. Mean Absolute Deviation (MAD)

- **Application:** To provide a robust measure of dispersion, the company calculates the MAD, which is less sensitive to extreme values compared to variance and standard deviation.
- **Advantages:** Offers a more interpretable and less skewed measure of sales variability when outliers exist.
- **Disadvantages:** It doesn't account for the direction of deviation (whether above or below the mean).

7. Box-and-Whiskers Plot (Visual Representation of Sales Distribution)

- **Application:** The company uses a Box-and-Whiskers Plot to visually assess the distribution of sales, identify outliers, and understand the spread of sales data.
- **Advantages:** It effectively shows the central tendency, spread, and outliers in a single diagram, making it easier to communicate findings.
- **Disadvantages:** While it is visually informative, it can oversimplify data distribution, and some stakeholders may find it challenging to interpret without statistical knowledge.

Insights Gained from the Analysis

- The **mean sales** were high, but the **median** revealed that most days had moderate sales, indicating a few days with extremely high sales (likely due to promotions).
- The **mode** highlighted that certain sales days (during weekends) were more frequent.
- The **variance** and **standard deviation** showed significant fluctuations in sales, prompting the company to focus on stabilizing revenue.

- Using **MAD**, the company recognized that most deviations from the mean were moderate, but occasional spikes contributed to the overall variability.
- The **Box-and-Whiskers Plot** identified several outliers, such as Black Friday and Cyber Monday sales, which strongly influenced the overall sales analysis.

Conclusion

Each statistical tool offered unique insights into the company's sales patterns. However, the company realized that no single tool could fully explain the data, and the combination of these tools provided a more comprehensive view of the business's performance.

Discussion Questions

1. How does the mean differ from the median in providing insights into the e-commerce company's sales performance, and why might it be important to use both?
2. Why might the company prefer using Mean Absolute Deviation (MAD) over variance or standard deviation when assessing sales volatility?
3. In what other real-world applications might the Box-and-Whiskers Plot be particularly useful, and why?
4. What are the limitations of using only descriptive statistics (such as mean, median, mode) in making business decisions, and how might these limitations be addressed?
