
Lecture 2: Data Detectives

Part 1: Describing the Center and Spread

A campus coffee shop recorded the number of drinks sold during 9 different half-hour time blocks:

6, 8, 9, 10, 10, 11, 12, 13, 20

1. Calculate the mean of the dataset. Then explain what the mean is telling us.
2. Calculate the median of the dataset. Then explain what the median is telling us.
3. Find the mode of the dataset. Then explain what the mode is telling us.
4. Calculate the range difference of the dataset.
5. Calculate the standard deviation of the dataset.

-
6. Compare the mean and median. Do they tell a similar story about the “typical” number of sales in a half-hour block, or a different story? Briefly explain.

Part 2: Thinking About Outliers

7. Is there a possible outlier in this dataset? If so, which value seems unusual, and why?
8. Remove the possible outlier and recalculate the mean of the remaining data.
9. Compare the new mean to the original mean. What does this suggest about how outliers affect the mean?
10. Without recalculating it exactly, do you think the median would change a lot, a little, or not at all if the outlier were removed? Explain your reasoning.

Part 3: Interpreting a z-score

Use the original dataset.

11. Calculate the z-score for an hour with 7 sales.
12. Is 7 sales above or below the mean? How can you tell from the z-score?
13. Suppose another hour had a z-score of 1.8. What does that tell us about that half-hour's sales compared to the rest of the dataset?

Part 4: Visualizing the Data

14. Sketch a rough histogram for the sales data. You do not need to make it perfect, but the overall shape should make sense.
15. Based on your sketch, does the distribution look fairly symmetric, skewed left, or skewed right? Explain your answer.
16. If you were trying to explain this dataset to the manager of the business, which measure of center would you use more: the mean or the median? Explain why.

Part 5: Categorical Data

A second store recorded the department where each sale happened during one afternoon:

Clothing, Electronics, Clothing, Grocery, Grocery, Grocery
Electronics, Clothing, Grocery, Toys, Clothing, Grocery

17. Create a count table for the departments.
18. Which department had the most sales?
19. What type of graph would be most appropriate for this categorical data?
20. Briefly explain why a histogram would not be appropriate for this data.

Part 6: Reflection

21. In a few sentences, explain the difference between describing the center of a dataset and describing the spread of a dataset.
22. Which measure from today's activity still feels the least clear to you: mean, median, mode, range, standard deviation, or z-score? Briefly explain what is confusing.