MALMS8

TAILORED LEARNING, PERSONALIZED PROGRESS

SUPPLEMENTARY MODULE 1

Determine Measures of Central Tendency of Ungrouped Data

GRADE 8

malms8.online

3 DRAW CONCLUSIONS FROM STATISTICAL DATA USING THE MEASURES OF CENTRAL TENDENCY

WELCOME, LEARNERS!

Hey there, data detectives! Fret not if you're feeling a bit puzzled after our main course! Mastering measures of central tendency takes practice, and it's okay to take your time. This supplementary session is here to be your guide, helping those still unlocking the secrets of ungrouped data. Remember, every step you take towards understanding is a stride closer to mastering this skill. You've got this!

LEARNING OBJECTIVES

1. Critically analyze the impact of outliers on measures of central tendency, discerning between mean, median, and mode through comparative analysis.
2. Synthesize understanding of central tendency measures, employing statistical techniques and enhancing ability to communicate analytical insights coherently.
3. Recognize the subtle differences in determining measures of central tendency, cultivating a sense of confidence in interpreting data variations.

The choice of using mean, median, or mode depends on the characteristics of the data and the specific insights you're seeking. Here are scenarios where each measure might provide different insights:

1. **Symmetrical Distribution:**
    - **Mean:** In symmetrical distributions (like a normal distribution), the mean, median, and mode coincide, offering a central tendency estimate that represents the typical value well.
    - **Median:** Useful when dealing with skewed data or distributions with outliers. It's robust to extreme values, providing a more accurate representation of the center when outliers exist.
    - **Mode:** Helpful when identifying the most frequent value(s) in a dataset, like in categorical data or distributions with clear peaks. However, it might not fully represent the central tendency in more complex distributions.

Consider the following data set. 4; 5; 6; 6; 6; 7; 7; 7; 7; 7; 7; 8; 8; 8; 9; 10

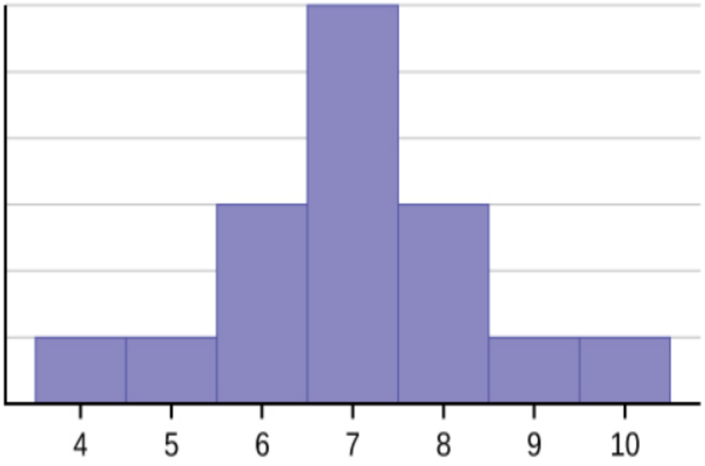

Figure 1. Example of Symmetrical Distribution

HISTOGRAM

Histograms show balanced data distribution. A distribution is symmetrical if a vertical line can be formed in the histogram such that the shapes to the left and right of it are mirror images. These data have a seven-point mean, median, and mode. In a fully symmetrical distribution, mean and median are equal. One mode (unimodal) matches the mean and median in this example. In a bimodal symmetrical distribution, the modes are different from the mean and median.

2. **Skewed Data:**
   - **Mean:** Easily influenced by extreme values in skewed datasets, giving a misleading representation of the average. For positively skewed data (long tail on the right), the mean tends to be higher than the median.
   - **Median:** Ideal for skewed distributions. It accurately represents the central value, being less affected by outliers or extreme values.
   - **Mode:** May not reflect the overall shape of the distribution in skewed data; it could be at a lower or higher value compared to the central tendency.

The histogram for the data: 4; 5; 6; 6; 6; 7; 7; 7; 7; 8 is not symmetrical. The right-hand side seems "chopped off" compared to the left side. A distribution of this type is called skewed to the left because it is pulled out to the left.

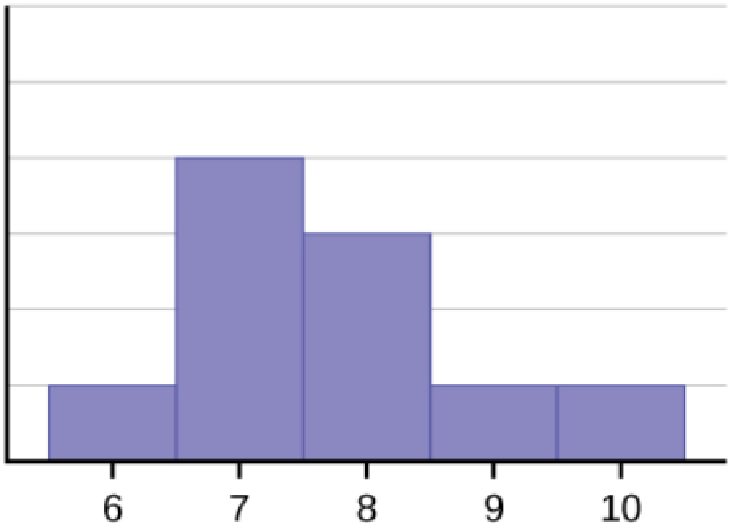

Figure 2. Positively Skewed

The histogram for the data: 6; 7; 7; 7; 7; 8; 8; 8; 9; 10, is also not symmetrical. It is **skewed to the right**.

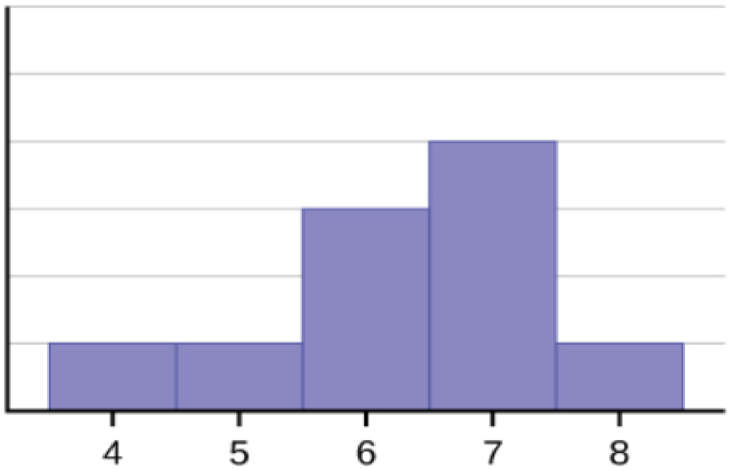

Figure 3. Negatively Skewed

3. **Bimodal or Multimodal Distributions:**
    - **Mean:** Might not be meaningful in cases where the data has multiple peaks or modes. It could fall between peaks, providing a value that doesn't represent the data well.
    - **Median:** Provides a clear central value, even in multimodal distributions, by identifying the middle value.
    - **Mode:** Useful for identifying individual peaks or clusters in the data but might not capture the entire central tendency when multiple modes exist.

A **bimodal mode** is a set of data that has two modes. This indicates that the data values with the highest frequencies are two.

Set A = {2,2,2,3,4,4,5,5,5} has a mode of 2 and 5, because both 2 and 5 are repeated three times in the provided set.

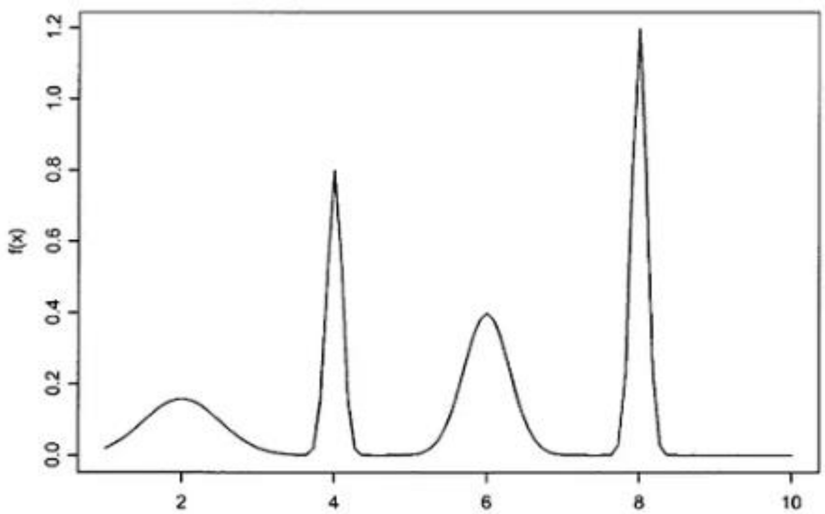

Figure 4. Bimodal Mode

A **multimodal mode** is a set of data that contains four or more modalities.

Because all four values in the given set recur twice, the mode of data set A = 100, 80, 80, 95, 95, 100, 90, 90,100,95 is 80, 90, 95 and 100. As a result, it's a multimodal dataset.

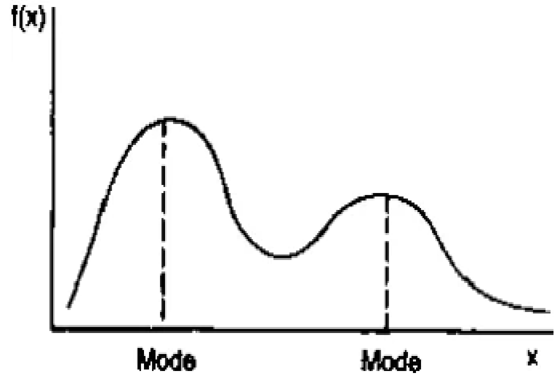

Figure 5. Multimodal mode

Datasets where the mean, median, and mode significantly differ in an ungrouped data.

Consider a dataset representing the ages of people in a neighborhood:

15,18,20,22,24,26,5015,18,20,22,24,26,50

Calculating Mean, Median, and Mode:

Mean=15+18+20+22+24+26+50=175

175/ 7= 25

Median: Since there are 7 values, the median is the fourth value:

Median= 22

Mode: In this dataset, there is no value that repeats, so there is no mode.

Data Set	Mean	Median	Mode
15,18,20,22,24,26,50	25	22	No mode

In this dataset, the mean (25) is significantly influenced by the outlier (50). The median (22) gives a better representation of the typical age in the neighborhood, as it's less affected by extreme values. Since no value repeats, there's no mode in this dataset.

This example illustrates how an outlier (50 years old person) significantly affects the mean, pulling it higher, while the median remains closer to the central tendency of the majority of values.

.