MALMS8

TAILORED LEARNING, PERSONALIZED PROGRESS

LEARNING MODULE 1

Determine Measures of Central Tendency of Ungrouped Data

GRADE 8

malms8.online

3 DETERMINE MEASURES OF CENTRAL TENDENCY OF UNGROUPED DATA

Welcome!

Welcome, adventurers, to our data journey! In this main course, we'll embark on an exciting expedition through the world of ungrouped data and its central tendencies. We'll interpret real-life circumstances, solve entertaining issues, and arm ourselves with the tools to analyze and evaluate data like seasoned statisticians during our journey. Put on your explorer hats and join me as we navigate this course, seeking hidden riches within numbers!

LEARNING OBJECTIVES

1. Evaluate and compare the applicability of mean, median, and mode in describing central tendencies within ungrouped data;
2. Enhance their computational skills in calculating measures of central tendency, manipulating datasets and generating data subsets
3. Explore the significance of measures of central tendency in real-world context.

Mean

The mean, also known as the , is a fundamental measure of central tendency in statistics. It represents the sum of all values in a dataset divided by the total number of values. However, its susceptibility to extreme values, or outliers, can distort its accuracy in reflecting the central tendency of the data.

Understanding Outliers

Outliers are data points that significantly differ from the rest of the dataset. They can arise due to measurement errors, natural variations, or exceptional circumstances. When calculating the mean, outliers hold the potential to heavily influence its value, particularly in smaller datasets.

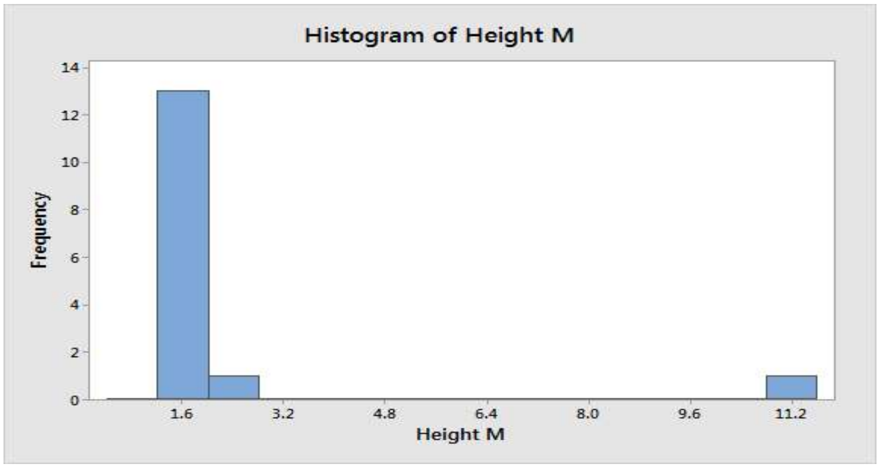


Figure 1. Example of an Outlier in the Height Frequency

In the figure above, the height in meters which is 1.6 is the outlier since it has extremely number of frequency compared to the 2 measurements.

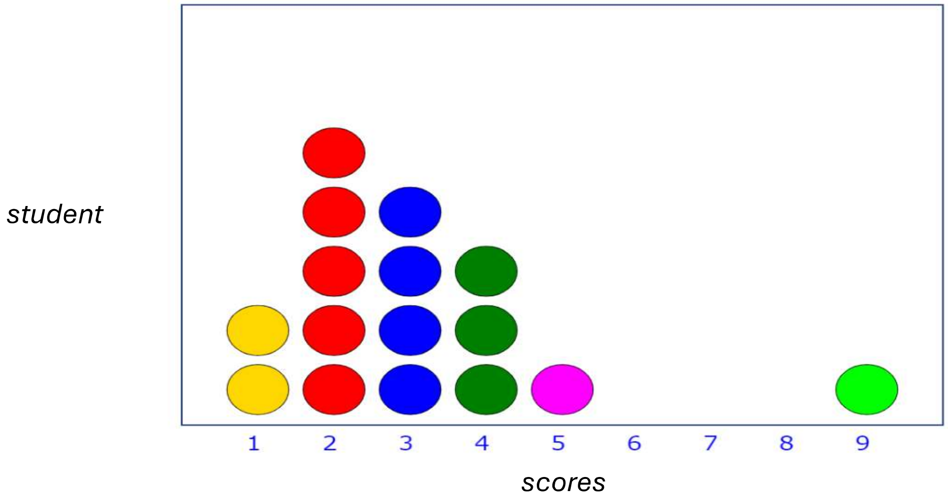


Figure 2. Example of an Outlier in the Students' Scores

Consider the dot plot below. We would call 9 an outlier as it is well above the distribution of students' scores.

Example:

Illustrating Sensitivity to Outlier

Let's consider an example problem to illustrate this sensitivity:

Problem: Suppose a class of students' test scores in Mathematics is as follows: 75,80,82,85,90,95,100,150

Calculating the Mean:

First, let's calculate the mean of the dataset:
Mean=75+80+82+85+90+95+100+150=677

757/8 = 94.63

The calculated mean is 94.63.

Impact of Outlier:

The outlier in this dataset is the score of 150. When this extreme value is included in the calculation of the mean, it significantly affects the result. Let's recalculate the mean without the outlier:

Mean (excluding outlier)=75+80+82+85+90+95+100=607

607/7 ≈86.71

The mean without the outlier is approximately 86.71, significantly lower than the mean including the outlier (94.63).

Importance of Examining Data Distribution

Exploring the distribution through methods like boxplots, histograms, or scatter plots allows us to better understand the data's behavior, identify outliers, and choose appropriate measures of central tendency.

The mean is a valuable statistical tool, but its susceptibility to outliers requires caution. Evaluating data distribution and considering alternative measures of central tendency (like the median or mode) can provide a more robust understanding of the dataset's characteristics.

Median

Constructing strategies to manage skewed distributions by emphasizing the use of the median as a measure of central tendency.

Understanding Skewed Distributions

Skewness is a measure of the asymmetry of a distribution. A distribution is asymmetrical when its left and right side are not mirror images.

A distribution can have right (or positive), left (or negative), or zero skewness. A right-skewed distribution is longer on the right side of its peak, and a left-skewed distribution is longer on the left side of its peak. Skewness can have a significant impact on the mean's accuracy as a measure of central tendency, especially in asymmetric datasets.

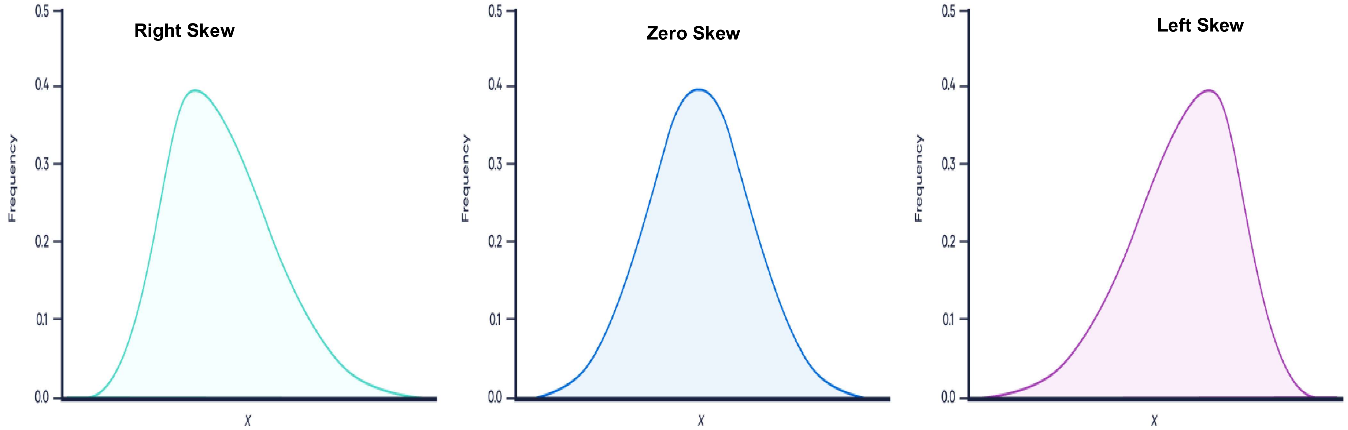


Figure 3. Types of Skewness

DID YOU KNOW?

Importance of Median in Skewed Distributions

The median, unlike the mean, is resistant to extreme values and outliers. In skewed distributions, the median tends to provide a more accurate representation of the central value. It is the middle value when the data is arranged in ascending or descending order, effectively dividing the dataset into two equal halves.

Nature of Skewness

Skewness refers to the lack of symmetry in a distribution. A perfectly symmetrical distribution exhibits no skewness, where the mean, median, and mode align at the center. However, in skewed distributions, the data is asymmetrical, and the mean, median, and mode may not coincide.

Positive Skewness

In a positively skewed distribution, the tail of the data points extends towards higher values, pulling the mean toward the larger values, making it greater than the median.

Negative Skewness

Conversely, in a negatively skewed distribution, the tail stretches toward lower values, dragging the mean towards the smaller values, causing it to be less than the median.

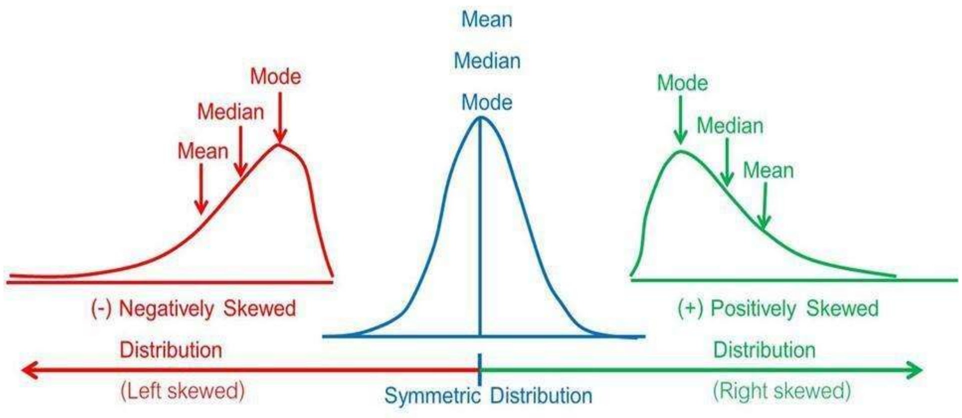


Figure 4. Positive and Negative Skewness

Impact on Mean and Median

Positive Skewness:

In a positively skewed distribution:

- The mean is pulled in the direction of the skew by the presence of larger values, increasing its value.
- The median remains less influenced by extreme values and typically lies closer to the smaller values, being less than the mean.

Negative Skewness:

In a negatively skewed distribution:

- The mean is dragged towards the smaller values due to the presence of lower values, decreasing its value.
- The median, being less sensitive to extreme values, remains relatively stable and tends to be greater than the mean.

Example

Calculating the mean and median for each dataset and observing how the presence of extreme values influences the measures of central tendency.

Dataset 1 (Positive Skewness):

5,7,9,11,13,15,20,25,35,80

Mean Calculation:

Mean=5+7+9+11+13+15+20+25+35+80 = 220

220/ 10 = **22**

Median Calculation:

Since the data is in ascending order:

Median= 13+15 = 28

28/2= **14**

In this positively skewed dataset, the mean is influenced by the larger outlier (80), causing it to be higher than the median.

Dataset 2 (Negative Skewness):

90,85,80,75,70,65,30,20,15,10

Mean Calculation:

Mean=90+85+80+75+70+65+30+20+15+10 = 540

540/10 = **54**

Median Calculation:

Since the data is in ascending order:

Median= 70+65= 135

135/2 = **67.5**

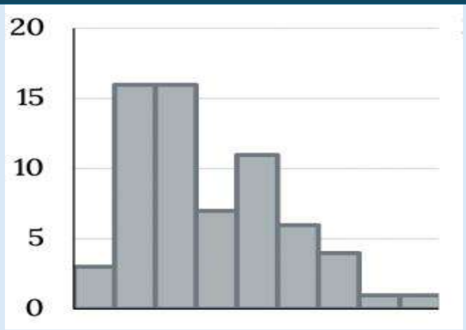
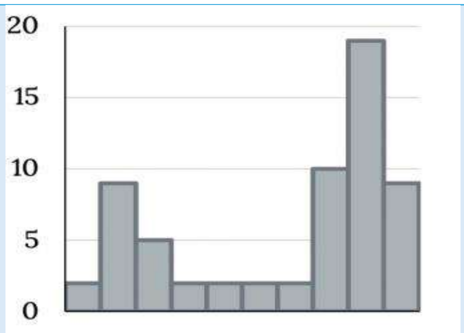
In this negatively skewed dataset, the mean is influenced by the smaller values (10), causing it to be lower than the median.

Observations:

- **Dataset 1:** The presence of the larger outlier (80) significantly affects the mean, pulling it upwards from the rest of the data. However, the median remains less impacted, providing a more stable representation of central tendency.
- **Dataset 2:** The smaller values (especially 10) in the dataset have a notable influence on the mean, pulling it downwards, while the median remains less affected, giving a clearer picture of the central value.

Mode

The mode represents the most frequently occurring value or values in a dataset. It is particularly valuable in categorical data where items are grouped into distinct categories or in datasets where values occur discretely. Unlike the mean and median, the mode is not affected by extreme values or outliers.

Types of Mode	Definition	Visual Representation
Unimodal	Describes a distribution with a single peak or mode, indicating that one value occurs more frequently than any other value in the dataset.	
Bimodal	Refers to a distribution with two distinct peaks or modes, indicating that there are two values that occur with relatively high frequency in the dataset.	
Multimodal	Characterizes a distribution with three or more peaks or modes, indicating the presence of multiple values that occur with relatively high frequency in the dataset.