

---

## 2. Data Description

### 1. Measures of Central Tendency

### 2. Measures of Variation

#### 2.1 Measures of central tendency

After the process of classification and tabulation the next important objective of statistical analysis is to determine various numerical measures which measure inherent characteristics of the data. It can be achieved through the statistical techniques such as measures of central tendency and dispersion.

The concentration of values around central value of a distribution is known as **central tendency**.

The different measures that are used to study the characteristics of the distribution are known as measures of central tendency.

The following are some common measures of central tendency.

1. Arithmetic Mean (AM) – Simple and Weighted
2. Geometric Mean (GM)
3. Harmonic Mean (HM)
4. Median (M)
5. Mode (Z)

**Arithmetic Mean (AM)** is also known as simple average. It is denoted by  $\bar{x}$ .  $\bar{x}$  can be calculated by dividing the sum (total) of all observations by number of observations.

To calculate the arithmetic mean of a set of data we must first add up (sum) all of the data values ( $x$ ) and then divide the result

by the number of values (n). Since  $\Sigma$  is the symbol used to indicate that values are to be summed.

$$\bar{x} = \frac{\sum x}{n}$$

**Example: Find the mean of: 6, 8, 11, 5, 2, 9, 7, 8**

$$\bar{X} = \frac{\sum x}{n} = \frac{6+8+11+5+2+9+7+8}{8} = 7$$

**Example:**

Consider the data set: 17, 10, 9, 14, 13, 17, 12, 20, 14

$$\text{Mean} = \frac{\sum x_i}{n} = \frac{17 + 10 + 9 + 14 + 13 + 17 + 12 + 20 + 14}{9} = \frac{126}{9} = 14$$

The mean of this data set is 14.

**Example:** Calculation of the mean – birthweights

Consider the following five birthweights in kilograms recorded to 1 decimal place selected randomly from the Simpson (2004) study of low birthweight babies.

1.2, 1.3, 1.4, 1.5, 2.1

$$\text{Mean} = \frac{\sum x_i}{n} = \frac{1.2 + 1.3 + 1.4 + 1.5 + 2.1}{5} = \frac{7.5}{5} = 1.50 \text{ Kg}$$

It is usual to quote 1 more decimal place for the mean than the data recorded.

**Median:** The median value of a set of data is the middle value of the ordered data

**Step 1** Arrange the observed values of variable in a data in increasing order.

**Step 2** If the total number of items  $n$  is an odd number, then the number on the  $n+1/2$  position is the median; If  $n$  is an even number, then the average of the two numbers on the  $n/2$  and  $(n/2)+1$  positions is the median. (For ordinal level of data, choose any one on the two middle positions).

**Example: find the median of the following:**

**A) 11,4,9,7,10,5,6**

Ordering the data gives: 4,5,6,7,9,10,11

$n=7$  (Odd number)

The median position is  $(n+1)/2$

The Median Position =  $(7+1)/2 = 4$

4,5,6, 7,9,10,11

Median=7

**B) 11,4,9,9,7,10,5,6**

Ordering the data gives: 4,5,6,7,9,9,10,11

$n=8$  (Even number)

The median position is  $n/2$  ,  $n/2+1$

Median Position:  $8/2=4$  ,  $(8/2) +1=5$

4,5,6,7,9, 9,10,11

Here there is a middle pair 7 and 9. The median is between these 2 values i.e. the mean of them

Median=  $(7+9)/2=8$

**Example:**

Consider the data set: 17, 10, 9, 14, 13, 17, 12, 20, 14

Step 1: Put the data in order from smallest to largest. 9, 10, 12, 13, 14, 14, 17, 17, 20

Step 2: Determine the absolute middle of the data. 9, 10, 12, 13, 14, 14, 17, 17, 20

Note: Since the number of data points is odd, choose the one in the very middle.

The median of this data set is 14.

**Mode**

**Mode:** The data entry that occurs with the greatest frequency. A data set may have one mode, more than one mode, or no mode. If no entry is repeated the data set has no mode

To determine the mode:

1. Put the data in order from smallest to largest, as you did to find your median.
2. Look for any value that occurs more than once.
3. Determine which of the values from Step 2 occurs most frequently.

**Example:**

Consider the data set: 17, 10, 9, 14, 13, 17, 12, 20, 14

Step 1: Put the data in order from smallest to largest. 9, 10, 12, 13, 14, 14, 17, 17, 20

Step 2: Look for any number that occurs more than once. 9, 10, 12, 13, 14, 14, 17, 17, 20

Step 3: Determine which of those occur most frequently. 14 and 17 both occur twice. The modes of this data set are 14 and 17.

## Mean (Average)

*The total of the numbers divided by how many numbers there are.*

11 7 11 18 9 7 6 23 7  
 $11+7+11+18+9+7+6+23+7=99$   
 $99 / 9 = 11$   
**Mean: 11**

## Median (Middle)

*The number which is in the middle or the middle value.*

11 7 11 18 9 7 6 23 7  
 6 7 7 7 9 11 11 18 23  
**Median: 9**

## Mode (Most)

*The number that appears the most.*

11 7 11 18 9 7 6 23 7  
 6 7 7 7 9 11 11 18 23  
**Mode: 7**

**Exercises**/The age (in years) of a sample of 20 motorcyclists killed in road traffic accidents is given below.

18 41 24 28 71 52 15 20 21 31

16 24 33 44 20 24 16 64 24 32

Calculate the mean, median and mode.