# Statistics

**Statistics** is one of the parts of mathematics in which we study about the collecting, organizing, analyzing, interpreting and presenting data.

Statistics is very helpful in real life situations as it is easy to understand if we represent a data in a particular number which represents all numbers. This number is called the **measure of central tendency**. Some of the central tendencies commonly in use are mean, median and mode.

## Mean

It is a ratio of sum of all observation with total number of observation.it is denoted by $\overline{X}$

**It is the average of "n" numbers, which is calculated by dividing the sum of all the numbers by n. It is also called as "Arithmetic mean" or simply "Average".**

## For Raw data

If $x_1, x_2, x_3, \ldots\ldots x_n$ are n observation then mean of n values $x_1, x_2, x_3, \ldots\ldots x_n$ is given by,

$$\overline{X} = \frac{x_1 + x_2 + x_3 + \ldots\ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

**Example**: Find mean of following data.

50, 90, 120, 40, 180, 200, 80.

Solution:- $\overline{X} = \frac{50 + 90 + 120 + 40 + 180 + 200 + 80}{7}$

$$= \frac{760}{7}$$

$$\overline{X} = 108.57$$

## For Ungrouped Frequency data

**Frequency of observation:** it is number of repetition of observation in given statistical data. It is denoted by "fi"

**If $x_1, x_2, x_3 \ldots x_n$ are the observations**

**$f_1, f_2, f_3, \ldots\ldots f_n$ are the respective frequencies of the given observations**

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \cdots + f_nx_n}{f_1 + f_2 + \cdots + f_n}$$

$$= \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i}$$

**Example:**

Find mean of following data,

| Observation (xi) | 20 | 40 | 60 | 80 | 100 |
|---|---|---|---|---|---|
| **Frequency (**fi) | 40 | 60 | 30 | 50 | 20 |

Solution:

| **Table for Mean** | | |
|---|---|---|
| xi | fi | fixi |
| 20 | 40 | 800 |
| 40 | 60 | 2400 |
| 60 | 30 | 1800 |
| 80 | 50 | 4000 |
| 100 | 20 | 2000 |
| | 200 | ∑fixi = 11000 |

Here, $x_1$, $x_2$, $x_3$, $x_4$, $x_5$ are 20, 40, 60, 80, 100 respectively and $f_1$, $f_2$, $f_3$ , $f_4$, $f_5$ are 40, 60, 30, 50, 20 respectively.

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \cdots + f_nx_n}{f_1 + f_2 + \cdots + f_n}$$

$$\bar{x} = \frac{11000}{200} = 55$$

## For Grouped Frequency Data

When the data is grouped in the form of class interval then we use a midpoint of each class interval as "xi" which represents the whole class. It is called the **class mark or mid value**. It is the average of the upper limit and the lower limit.

$$Classmark = Midvalue = \frac{Upper\lim it + Lower\lim it}{2}$$

**Mean of grouped frequency data is given by**, $\overline{X} = \frac{\sum fi.xi}{\sum fi}$ where "xi" is mid value of class.

### Example

A teacher marks the test result of the class of 55 students for mathematics. Find the mean for the given group.

| **Marks of Students** | 0 – 10 | 10 – 20 | 20 – 30 | 30 – 40 | 40 – 50 | 50 – 60 |
|---|---|---|---|---|---|---|
| **Frequency** | 27 | 10 | 7 | 5 | 4 | 2 |

Solution:

To find the mean we need to find the mid-point or class mark for each class interval which will be the x and then by multiplying frequency and midpoint we get fixi.

| **Marks of students** | **Frequency(fi)** | **Midpoint(xi)** | **fixi** |
|---|---|---|---|
| 0 – 10 | 27 | 5 | 135 |
| 10 – 20 | 10 | 15 | 150 |
| 20 – 30 | 7 | 25 | 175 |
| 30 – 40 | 5 | 35 | 175 |
| 40 – 50 | 4 | 45 | 180 |
| 50 – 60 | 2 | 55 | 110 |
| | | ∑fi = 55 | ∑fixi = 925 |

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \cdots + f_nx_n}{f_1 + f_2 + \cdots + f_n}$$

$$\bar{x} = \frac{925}{55} = 16.8 \text{ marks}$$

# Measure of Dispersion

**Dispersion means variation or scatterness of observations.**

**1) Range**: It is simplest measurement of dispersion. It is maximum distance between any two observations.

For Raw data: -

Range = Largest value – Smallest value

= L – S

Coefficient of Range $= \frac{L-S}{L+S}$

Example:

Find the Range and coefficient of Range for the following data:

21, 25, 30, 35, 32, 27

Solution: Range $= L - S$

$= 35 - 21$

$= 14$

Coefficient of Range $= \frac{L-S}{L+S}$

$= \frac{35-21}{35+21}$

$= \frac{14}{56}$

$= \frac{1}{4}$ OR 0.25

For Ungrouped data: -

Range = Largest value of $x_i$ – Smallest value of $x_i$

= L – S

Coefficient of Range = $\frac{L-S}{L+S}$

Example:

Find the range and coefficient of range for the following:

| $x_i$ | 30 | 40 | 50 | 60 | 70 |
|---|---|---|---|---|---|
| $f_i$ | 15 | 20 | 17 | 22 | 18 |

Solution: Range = $L-S$

$= 70-30$

$= 40$

Coefficient of Range $= \frac{L-S}{L+S}$

$= \frac{70-30}{70+30}$

$= \frac{40}{100} = 0.4$

For Grouped data: -

Range = Upper class boundary of last class – Lower class boundary of first class

= U – L

Coefficient of Range = $\frac{U-L}{U+L}$

Example:

Find the range and coefficient of range of the following:

| Marks | 21-25 | 26-30 | 31-35 | 36-40 | 41-45 |
|---|---|---|---|---|---|
| No. of students | 4 | 16 | 38 | 12 | 10 |

Range = U – L

=45.5-20.5

=25

Coefficient of range = $\frac{U-L}{U+L}$

$= \frac{45.5-20.5}{45.5+20.5}$

=0.379

**2) Mean deviation about mean:** It is average distance between any observation and mean of observation.

- For raw data:

$$\text{Mean deviation about mean} = \frac{\sum|x_i - \bar{x}|}{N} = \frac{\sum|d_i|}{N}$$

where $\bar{x}$ = mean of $N$ observations

Example:

Find mean deviation about mean of the following data:

2. 3, 4, 5, 6, 7

Solution:

$$\bar{x} = \frac{\sum x_i}{N} = \frac{2+3+4+5+6+7}{6}$$

$$= 4.5$$

| $x_i$ | $\lvert d_i \rvert = \lvert x_i - \bar{x} \rvert$ |
|---|---|
| 2 | 2.5 |
| 3 | 1.5 |
| 4 | 0.5 |
| 5 | 0.5 |
| 6 | 1.5 |
| 7 | 2.5 |
| | $\sum \lvert d_i \rvert = 9$ |

$$\text{M.D.} = \frac{\sum \lvert d_i \rvert}{N} = \frac{9}{6}$$

$$= 1.5$$

- For ungrouped data:

  Mean deviation about mean= $\frac{\sum f_i|x_i-\bar{x}|}{\sum f_i}=\frac{\sum f_i|d_i|}{N}$

- Example:

  Calculate mean deviation from mean for the following data:

| Marks | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|
| No. of students | 1 | 3 | 7 | 5 | 2 | 2 |

Solution:

| $x_i$ | $f_i$ | $f_ix_i$ | $\|d_i\|=\|x_i-\bar{x}\|$ | $f_i\|d_i\|$ |
|---|---|---|---|---|
| 3 | 1 | 3 | 2.5 | 2.5 |
| 4 | 3 | 12 | 1.5 | 4.5 |
| 5 | 7 | 35 | 0.5 | 3.5 |
| 6 | 5 | 30 | 0.5 | 2.5 |
| 7 | 2 | 14 | 1.5 | 3.0 |
| 8 | 2 | 16 | 2.5 | 5.0 |
| | $\sum f_i=20$ | $\sum f_ix_i=110$ | | $\sum f_i\|d_i\|=21$ |

$$\text{Mean}=\bar{x}=\frac{\sum f_ix_i}{N}=\frac{110}{20}=5.5$$

$$\text{M.D.}=\frac{\sum f_i|d_i|}{N}=\frac{21}{20}=1.05$$

- For grouped data:

$$\text{M.D. about mean} = \frac{\sum f_i |x_i - \bar{x}|}{\sum f_i} = \frac{\sum f_i |d_i|}{N}$$

where $x_i$ = Mid-value

- Example:

  Find mean deviation of the following data:

| Class | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 |
|---|---|---|---|---|---|
| Frequency | 1 | 2 | 4 | 2 | 1 |

Solution:

| Class | $f_i$ | $x_i$ | $f_i x_i$ | $\lvert d_i \rvert = \lvert x_i - \bar{x} \rvert$ | $f_i \lvert d_i \rvert$ |
|---|---|---|---|---|---|
| 0-10 | 1 | 5 | 5 | 20 | 20 |
| 10-20 | 2 | 15 | 30 | 10 | 20 |
| 20-30 | 4 | 25 | 100 | 0 | 0 |
| 30-40 | 2 | 35 | 70 | 10 | 20 |
| 40-50 | 1 | 45 | 45 | 20 | 20 |
| | $\sum f_i = 10$ | | $\sum f_i x_i = 250$ | | $\sum f_i \lvert d_i \rvert = 80$ |

$$\text{Mean} = \bar{x} = \frac{\sum f_i x_i}{N} = \frac{250}{10} = 25$$

$$\text{M.D. about mean} = \frac{\sum f_i |d_i|}{N}$$

$$= \frac{80}{10} = 8$$

**3) Standard Deviation ($\sigma$):** It is standard distance between any observation and mean of observation. It is defined as square root of arithmetic mean of square of deviation taken from mean of observation.

- For raw data:

$$S.D. = \sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}} = \sqrt{\frac{\sum d_i^2}{N}}$$

- Example:

  Find standard deviation for:

  9, 10, 12, 13, 18, 20, 21, 25

Solution: $\bar{x} = \frac{\sum x_i}{N} = \frac{9+10+12+13+18+20+21+25}{8} = 16$

| $x_i$ | $d_i = x_i - \bar{x}$ | $d_i^2$ |
|---|---|---|
| 9 | -7 | 49 |
| 10 | -6 | 36 |
| 12 | -4 | 16 |
| 13 | -3 | 9 |
| 18 | 2 | 4 |
| 20 | 4 | 16 |
| 21 | 5 | 25 |
| 25 | 9 | 81 |
| | | $\sum d_i^2 = 236$ |

$$\sigma = \sqrt{\frac{\sum d_i^2}{N}}$$

$$= \sqrt{\frac{236}{8}} = 5.431$$

- For ungrouped data:

$$S.D. = \sigma = \sqrt{\frac{\sum f_i d_i^2}{N}}$$

Where, $d_i = \left( x_i - \bar{x} \right)$

Example:

Calculate standard deviation for the following:

| $x_i$ | 27 | 28 | 29 |
|---|---|---|---|
| $f_i$ | 1 | 7 | 2 |

Solution:

| $x_i$ | $f_i$ | $f_i x_i$ | $d_i = x_i - \bar{x}$ | $d_i^2$ | $f_i d_i^2$ |
|---|---|---|---|---|---|
| 27 | 1 | 27 | -1.1 | 1.21 | 1.21 |
| 28 | 7 | 196 | -0.1 | 0.01 | 0.07 |
| 29 | 2 | 58 | 0.9 | 0.81 | 1.62 |
| | $\sum f_i = 10$ | $\sum f x_i = 281$ | | | $\sum f_i d_i^2 = 2.9$ |

$$\bar{x} = \frac{\sum f_i x_i}{N}$$

$$= \frac{281}{10} = 28.1$$

$$S.D. = \sigma = \sqrt{\frac{\sum f_i d_i^2}{N}}$$

$$= \sqrt{\frac{2.9}{10}} = 0.539$$

- For grouped data:

$$\text{S.D.} = \sigma = \sqrt{\frac{\sum f_i \left(x_i - \bar{x}\right)^2}{\sum f_i}} = \sqrt{\frac{\sum f_i d_i^2}{N}}$$

where $x_i$ = mid-value

Example:

The following table shows the chest measurement of 100 students. Calculate the standard deviation.

| Chest in cm | 68-74 | 75-81 | 82-88 | 89-95 | 96-102 | 103-109 |
|---|---|---|---|---|---|---|
| No. of students | 5 | 31 | 40 | 20 | 3 | 1 |

Solution:

| Class | Continuous class | $f_i$ | $x_i$ | $f_i x_i$ | $d_i = \lvert x_i - \bar{x} \rvert$ | $d_i^2$ | $f_i d_i^2$ |
|---|---|---|---|---|---|---|---|
| 68-74 | 67.5-74.5 | 5 | 71 | 355 | 13.16 | 173.1856 | 865.928 |
| 75-81 | 74.5-81.5 | 31 | 78 | 2418 | 6.16 | 37.9456 | 1176.3136 |
| 82-88 | 81.5-88.5 | 40 | 85 | 3400 | 0.84 | 0.7056 | 28.224 |
| 89-95 | 88.5-95.5 | 20 | 92 | 1840 | 7.84 | 61.4656 | 1229.312 |
| 96-102 | 95.5-102.5 | 3 | 99 | 297 | 14.84 | 220.2256 | 660.6768 |
| 103-109 | 102.5-109.5 | 1 | 106 | 106 | 21.84 | 476.9856 | 476.9856 |
| | | $\sum f_i = 100$ | | $\sum f_i x_i = 8416$ | | | $\sum f_i d_i^2 = 4437.44$ |

$$\bar{x} = \frac{\sum f_i x_i}{N} = \frac{8416}{100} = 84.16$$

$$S.D. = \sqrt{\frac{\sum f_i d_i^2}{N}}$$

$$= \sqrt{\frac{4437.44}{100}} = 6.661$$

## Variance

The square of a standard deviation is called the variance.

Variance = $(S.D.)^2 = \sigma^2$

Coefficient of variance = $\frac{\sigma}{\bar{x}} \times 100$

Example:

Find variance and coefficient of variance of the following data:

| Marks | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 |
|---|---|---|---|---|---|
| No. of students | 14 | 23 | 27 | 21 | 15 |

Solution:

| C.I. | $f_i$ | $x_i$ | $f_i x_i$ | $d_i = \lvert x_i - \bar{x} \rvert$ | $d_i^2$ | $f_i d_i^2$ |
|---|---|---|---|---|---|---|
| 0-10 | 14 | 5 | 70 | 20 | 400 | 5600 |
| 10-20 | 23 | 15 | 345 | 10 | 100 | 2300 |
| 20-30 | 27 | 25 | 675 | 0 | 0 | 0 |
| 30-40 | 21 | 35 | 735 | 10 | 100 | 2100 |
| 40-50 | 15 | 45 | 675 | 20 | 400 | 6000 |
| | $\sum f_i = 100$ | | $\sum f_i x_i = 2500$ | | | $\sum f_i d_i^2 = 16000$ |

$$\text{Mean} = \bar{x} = \frac{\sum f_i x_i}{N} = \frac{2500}{100} = 25$$

$$\text{S.D.} = \sigma = \sqrt{\frac{\sum f_i d_i^2}{N}} = \sqrt{\frac{16000}{100}}$$

$$= 12.649$$

$$\text{Variance} = \sigma^2 = (12.649)^2$$

$$= 159.997$$

$$\text{Coefficient of variance} = \frac{\sigma}{\bar{x}} \times 100$$

$$= \frac{12.649}{25} \times 100$$

$$= 50.596$$

Ex: If coefficient of variance is 75% and standard deviation are 24. Find mean of data.

Solution: coefficient of variance = 75

Standard Deviation ($\sigma$) = 24

Mean ($\bar{x}$) =?

$$\text{Coefficient of variance} = \frac{\sigma}{\bar{x}} \times 100$$

$$75 = \frac{24}{\bar{x}} \times 100$$

$$\bar{x} = \frac{24}{75} \times 100$$

$$mean = \bar{x} = 32$$

Ex: If mean is 12 and coefficient of variation is 45%. Find standard deviation

Solution: coefficient of variance = 45

Mean ($\bar{x}$) = 12

Standard Deviation ($\sigma$) =?

$$\text{Coefficient of variance} = \frac{\sigma}{\bar{x}} \times 100$$

$$45 = \frac{\sigma}{12} \times 100$$

$$\frac{45 \times 12}{100} = \sigma$$

Standard Deviation ($\sigma$) = **5.4**

## Comparison of two statistical data

Let 'A' & 'B' are two statistical data under study.

Let, $V_1$ = Coefficient of variation of data A.

$V_2$ = Coefficient of variation of data B.

Then,

1) If $V_1 \prec V_2$, Data B is more variable than A

(Data A is more consistent than B)

2) If $V_1 \succ V_2$, Data A is more variable than B

(Data B is more consistent than A)

**For Example:**

The mean and S.D. of runs scored by two batsman in 10 innings are

| Batsman | Mean | S.D. |
|---|---|---|
| A | 50 | 15 |
| B | 12 | 2 |

Which batsman is more consistent?

Let $v_1$ and $v_2$ be coefficient of variations for Batsman A and B

$$v_1 = \frac{\sigma}{\bar{x}} \times 100$$

$$= \frac{15}{50} \times 100$$

$$= 30$$

$$v_2 = \frac{2}{12} \times 100$$

$$= 16.67$$

$\therefore v_1 > v_2$

$\therefore$ Batsman B is more consistent.

# Exercise

Find the range of the following data:
50, 90, 120, 40, 180, 200, 80

Find the range of the following data :
2, 3, 1, 6, 10, 17, 20, 24, 31

Calculate the range and coefficient of range of the following data:

| Marks | No. of Students |
|---|---|
| 10-20 | 5 |
| 20-30 | 10 |
| 30-40 | 12 |
| 40-50 | 25 |
| 50-60 | 8 |
| 60-70 | 10 |

Find the range and coefficient of range for the following data:

| Marks | No. Of Students |
|---|---|
| 10-19 | 6 |
| 20-29 | 10 |
| 30-39 | 16 |
| 40-49 | 14 |
| 50-59 | 8 |
| 60-69 | 4 |

Find the mean deviation of the following:
17, 20, 9, 13, 22, 15, 18, 17, 13

Calculate M.D. about mean for:

| Marks | No. of students |
|---|---|
| 3 | 4 |
| 4 | 9 |
| 5 | 10 |
| 6 | 8 |
| 7 | 6 |
| 8 | 3 |

Find M.D. for the following:

| Marks | No. of Students |
|---|---|
| 0-10 | 5 |
| 10-20 | 8 |
| 20-30 | 15 |
| 30-40 | 16 |
| 40-50 | 6 |

Find the standard deviation of the following:
1, 2, 3, 4, 5, 6, 7, 8, 9

If S.D.= 7.01 then Variance=?

Calculate the standard deviation of the following:

| Class | Frequency |
|---|---|
| 0-20 | 20 |
| 20-40 | 130 |
| 40-60 | 220 |
| 60-80 | 70 |
| 80-100 | 60 |

The following table shows the chest measurement of 100 students. Calculate the mean and standard deviation.

| Chest in cm | No. Of Students |
|---|---|
| 68-74 | 5 |
| 75-81 | 31 |
| 82-88 | 40 |
| 89-95 | 20 |
| 96-102 | 3 |
| 103-109 | 1 |

Coefficient of variation of a distribution is 75% and standard deviation is 24.What is its mean?

$v_1$ and $v_2$ are coefficients of variations for Set I and Set II.
$v_1$=10.25 and $v_2$=12.5 , which set is more consistent?

If mean is 82.5 and S.D.= 7.2, then coefficient of variance is?

Coefficient of variation of a certain distribution is 5 and mean is 60.Find S.D.

If the mean of the data is 12 and coefficient of variation is 45%, then find the standard deviation.

If the mean of the data is 12 and coefficient of variation is 45%, then find the standard deviation.

Find variance and coefficient of variance of the following data:
49, 63, 46, 59, 65, 52, 60, 54

The data of runs scored by two batsman A and B in five one day matches is given below:

| Batsman | Average runs scored | S.D. |
|---|---|---|
| A | 44 | 5.1 |
| B | 54 | 6.31 |

State which batsman is more consistent?

Find variance of the following :

| C.I. | frequency |
|---|---|
| 0-10 | 14 |
| 10-20 | 23 |
| 20-30 | 27 |
| 30-40 | 21 |
| 40-50 | 15 |

In the two factories A and B engaged in the same industry, the average weekly wages and S.D. are as follows:

| Factory | A | B |
|---|---|---|
| Avg. | 34.5 | 28.5 |
| S.D. | 5.0 | 4.5 |

Which factory is more consistent?

H. An analysis of monthly wages paid to the workers in two firms A and B belonging to the industry gives the following results:

| | Firm A | Firm B |
|---|---|---|
| Avg. wages | 186 | 175 |
| Variance | 81 | 100 |

Which firm has greater variability?

- Statistics is a science of collection, presentation, analysis and interpretation of numerical data.
- Statistics has a wide range of applications in the modern age of technology.
- Statistical information enables organizations to frame policies and guidelines to improve the overall working of the system.
- Statistics also helps in understanding various economic problems.
- In industries, statistics helps in the field of Quality Control and also provides information in making critical decisions .