Statistics

Statistics (Data Science)

Data:

- Collection of facts / information
- List of facts / information.
- → List of numbers or numerical information.

Height of all the students in your class : [Data]

S1 : 150 cm.
S2 : 153 cm.
S3 : 149 cm.
⋮
S10 — 151 cm.

Data

↳ Two types :
- Primary Data (Data collected personally, for specific purpose).
- Secondary. (collected by someone else for some other purpose)

Data collect in the very beginning is kla Raw Data

Organisation of Data

↳ organised in table (Rows and columns.)

Raw data is organised in tabular form.

Raw data can be organised using
- alphabetical order
- ascending order
- descending order
- serial nos.

The raw data when arranged in ascending or descending order of magnitude is called an array or arrayed data

Eg: Height of the 10 students

[149, 132, 145, 151, 146, 154, 135, 136, 151, 142]

Data
or
Raw Data

(observations)

151 → frequency is 2

frequency: no. of times an observation occurs in the dataset.

If Data set (Data)

↓

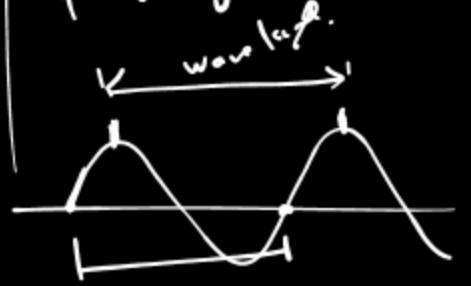
large no. of observation

For this we use

↓

Frequency distribution

Frequency Distribution Table



Frequency: No. of occurance of a given observation in the data set

41, 41, 42, 43, 44, 41, 49, 36, 35, 43, 41, 48, 45, 46, 45, 36, 39, 41, 45, 42 ...

41, 41, 42, 45, 44, 41, 49, 36, 35, 43, 41, 48, 45, 46, 45, 36, 39, 41, 45, 42 ...

Ascending order / Descending

Marks	Tally Bars	No. of students
35	I	1
36	II	2
39	I	1
41	IIII	5
42	II	2
43	II	2
44	I	1
45	III	3
46	I	1
48	I	1
49	I	1

Frequency Distribution

Arrangement

Central Values

- Mean
- Median
- Mode

Representative values of the given Dataset.

(2) (5) (6) (7)

$$\text{Mean} = \frac{21}{4} = 5.$$

They may or may not be actually present in the dataset.

Mean (Arithmetic mean)

$$\text{Arithmetic mean} = \frac{\text{Sum of all the observation}}{\text{No. of observations.}}$$

If $x_1, x_2, x_3, \cdots, x_n$ are n observations.

$$\boxed{\text{A.M.} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}}$$

Range of a Data

Range = Value of largest observation − Value of lowest observation

Example 1: If the heights of 5 students are 144 cm, 152 cm, 151 cm, 158 cm and 155 cm respectively. Find the mean height and range of the data.

$$\text{Mean height} = \frac{144 + 152 + 151 + 158 + 155}{5}$$

$$= \frac{760}{5} = 152 \text{ cm}$$

$$\text{Range} = 158 - 144$$

$$= 14.$$

Q. Find the mean of first five whole nos.

First five whole no. = 0, 1, 2, 3, 4

Q. Find the mean of first five prime numbers.

2, 3, 5, 7, 11

$$\text{mean} = \frac{2+3+5+7+11}{5} = \frac{28}{5} = \boxed{5.6}$$

Range : $11-2 = 9$

Q. If the mean of 6, 4, 7, p and 10 is 8, find the value of p.

$$\frac{6+4+7+p+10}{5} = 8$$

$$6+4+7+p+10 = 8\times 5$$

$$27+p = 40$$

$$p = 40-27$$

$$\boxed{p = 13}$$

Q. The mean of 10 numbers is 20. If 5 is subtracted from every number, what will be the new mean?

Let $x_1, x_2, x_3, \ldots, x_{10}$ be the 10 numbers/observations.

$$\frac{x_1 + x_2 + x_3 + \cdots + x_{10}}{10} = 20$$

$$x_1 + x_2 + x_3 + \cdots + x_{10} = 20 \times 10$$

$$x_1 + x_2 + x_3 + \cdots + x_{10} = 200$$

New numbers would be: $(x_1 - 5), (x_2 - 5), \ldots, (x_{10} - 5)$

$$\text{New mean} = \frac{(x_1 - 5) + (x_2 - 5) + (x_3 - 5) + \cdots + (x_{10} - 5)}{10}$$

$$= \frac{x_1 - 5 + x_2 - 5 + x_3 - 5 + \cdots + x_{10} - 5}{10} = \frac{(x_1 + x_2 + x_3 + \cdots + x_{10}) - 5 - 5 - 5 \cdots - 5}{10}$$

$$\text{New mean} = \frac{200 - 5 \times 10}{10} = \frac{200 - 50}{10}$$

$$= \frac{150}{10}$$

New mean = 15

Q. The mean of 16 numbers is 8. If 2 is added to every number, what will be the new mean.

8 +2

10

each +2

Q. If the mean of five observations x, $x+2$, $x+4$, $x+6$, $x+8$ is 11, find the mean of first three observations

Sol:

$$\frac{x+(x+2)+(x+4)+(x+6)+(x+8)}{5} = 11$$

$$x+x+2+x+4+x+6+x+8 = 55$$

$$5x+20 = 55$$

$$5x = 35$$

$$\boxed{x = 7}$$

$$\text{New mean} = \frac{x+x+2+x+4}{3} = \frac{3x+6}{3} = \frac{3\times 7+6}{3} = \frac{27}{3} = 9$$

$$\frac{\left(\frac{1}{4}\right)}{4}$$

$$\frac{1}{4} \div \frac{4}{1}$$

$$\frac{1}{4} \times \frac{1}{4}$$

$$\frac{\left(\frac{1}{4}\right)}{\left(\frac{1}{3}\right)} = \frac{1}{4} \div \frac{1}{3}$$

$$= \frac{1}{4} \times \frac{3}{1}$$

$$= \frac{3}{4}$$

$$\frac{5.3 \overset{\times 10}{}}{2.5 \overset{\times 10}{}} = \frac{53}{25}$$

$$\frac{\left(\frac{53}{10}\right)}{\left(\frac{25}{10}\right)}$$

$$\frac{53}{10} \times \frac{10}{25} = \frac{53}{25}$$

Q. The mean of 40 observations was 160. It was detected on rechecking that the value of 165 was wrongly copied as 125 for computation of mean. Find the correct mean.

Median :

⇒ Median of a group of observations is the value of the variable which divides the group into two equal parts.

eg: [24, 36, 46, 17, 18, 25, 35]

Arrange in increasing/decreasing order.

17, 18, 24 | 25 | 35, 36, 46

3 obs. 3 obs.

median

Marks: of 5 students out of 50.

5, 9, 11, 10, 50

Mean marks = $\frac{85}{5}$ = 17

$\frac{5+9+11+10}{4} = \frac{35}{4}$ = ~8.8

5, 9, 10, 11, 50

0, 0, 45, 46, 47

Mean = $\frac{45+46+47}{5}$ = 27.6

median = 45

6, 15, 120, 50, 100, 80, 10, 20.

17.54, 89.76

[6, 10, 15, 20, 50, 80, 100, 120]

$$\frac{20+50}{2} = 35$$

35

$$= \frac{17.54 + 89.76}{2}$$

When No. of observations are even, then find the mean of middle two values to get median.

[Median value may or may not be part of dataset.]

Q. Find median : 19, 25, 59, 48, 35, 31, 30, 32, 51. If 25 is replaced by 52, find new median.

Q. [The median of the observations 11, 12, 14, 18, $x+2$, $x+4$, 30, 32, 35, 41 arranged in ascending order is 24. Find the value of x.]

23 25

$$\frac{x+2+x+4}{2} = 24$$

median $\Rightarrow$ $\frac{(x+2)+(x+4)}{2} = 24$

$$\frac{2x+6}{2} = 24$$

$$x = 21$$

expression

$2x+6$ [expanded] $\rightarrow (2)(x+3)$ factor form

$$\frac{2x+6}{2} = \frac{2x}{2} + \frac{6}{2}$$

$$= x+3$$

End of the chapter