Statistics : Science of planning studies & experiment, obtaining data, organizing, summarizing, presenting, analyzing, interpreting data & drawing conclusions from there.

Data : collection of observations such as measurements, genders, survey response etc.

Population: elements / individuals / items whose characteristics are being studied

Sample : A portion of population selected for study

Arithmetic mean

$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$, n = population

Median

- data sorted ascendingly
- value right in the middle.

if even n, median = $\frac{x_{\frac{n}{2}} + x_{(\frac{n}{2}+1)}}{2}$

if odd n, ceil $\left(\frac{n}{2}\right)$.

Mode

- most common occurrence.

Bimodal : 2 values repeat the most 22, 22, 23, 23, 4, 5, 6.

Multi modal : >2 values repeat 22, 22, 22, 25, 25, 25, 26, 26, 26, 1, 7, 5

No mode : No repeat 1, 4, 7, 8

Range

- Highest value - lowest value.

Outlier

- A datapoint that differs significantly from other observations

Causes:

- Variability in measurement
- result of experimental error.
- indication of novel data.

Variance

Dispersion: - Extent in which data values differ from their average.

- How spread out each data is from the mean : $x_i - \bar{x}$

eg: A = 2, 3, 5, 6.

$\bar{x} = \frac{2+3+5+6}{4} = 4$

- variance is 1 way to measure dispersion
- Range is another way to measure dispersion (only 2 datapoints used, not descriptive of entire data)

population variance $\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$ ← to convert -ve values to +ve.

sample variance $s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$ ← empirically found that $n-1$ closer approximation to σ^2

Standard deviation

- Another way to calculate dispersion.

population SD : $\sigma = \sqrt{\sigma^2}$

sample SD : $s = \sqrt{s^2}$