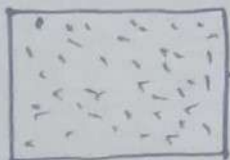
Venkatesh. K.H Date 05/11/2019

Measures of

1. Central Tendency 2. Asymmetry
3. $\boxed{\text{Variability}}$ $\longrightarrow$ Coefficient of variation

$\downarrow$ Variance $\downarrow$ Standard deviation

Different formulas for Sample and Population

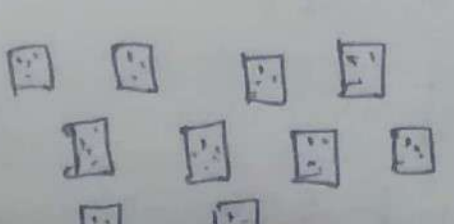


$\longrightarrow$ population data

when you have whole population Each data point is node. 100% sure of the measures you are calculating

$\longrightarrow$ Sample data

When you have Sample data. Sample statistics are an approximation of the population parameter.



$\longrightarrow$ 10 different Sample data

10 different Samples give 10 different measures

A population data set contains all members of a specified group.

Ex:- 11 members cricket team come all of them for fielding. 11 → population

A sample data set contains a part, or a subset

Ex:- Amongst, 11 members cricket team come only 2 of them for "Batting".

2 → Sample out of 11

11 → Cricket Team (population).

2 → Sample data.

Variance

Measures the dispersion of a set of data points around their mean value.

population variance denoted by

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Sample variance denoted by

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

Variance

population
1
2
3
4
5

$$\text{Mean} = \frac{1+2+3+4+5}{5}$$

$$\boxed{\text{Mean} = 3}$$

Then we apply the formula of population variance

observation → Mean

$$\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N} = \frac{(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5+3)^2}{5}$$

population size

$$= \frac{4+1+0+1+4}{5}$$

$$= \frac{\cancel{10}\,2}{\cancel{5}}$$

$$\boxed{\text{Population Variance} = 2}$$

Sample variance formula is used when our set of observations is a sample drawn from a bigger population

$$\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1 \to 5-1} = \frac{(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2}{4}$$

Sample variance = 2.50 → we had sample, but did not know the population. ∴ there is more uncertainty.

Mean = 3.00

Population variance = 2.00 → we had all the data and we calculated the variance.

Standard Deviation

is a measure of how spread out numbers are.

Denoted by — σ (sigma)

population standard deviation $\sigma = \sqrt{\sigma^2}$

Sample standard deviation $S = \sqrt{S^2}$

Standard deviation is the most common measure of variability for a SINGLE DATASET

The population SD : $\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2}$

The sample SD. $s = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$