Semester: V Subject: Statistics for AIDS Academic Year: 2023-2024

Estimates of variability:

A variance is the amount by which something changes or is different from something else.

Example:-

Consider 2 dataset $A = \{4,6\}$ $B = \{1,9\}$.

$A = \{4,6\}$

Lets find the mean.

$$\text{Mean } \bar{x} = \frac{4+6}{2} = \frac{10}{2}$$

$$\boxed{\bar{x} = 5}$$

$B = \{1,9\}$

$$\bar{x} = \frac{1+9}{2}$$

$$\boxed{\bar{x} = 5}$$

In both the cases the Mean is same, but the data distribution is different.

$\{4,6\} \rightarrow$ The data is near the mean. It is not that spread.

$\{1,9\} \rightarrow$ The data is spread far from mean.

This dispersion is known as variance.

The formula to calculate population variance is,

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

Lets calculate the variance in both the case.

$$\sigma^2 = \frac{(4-5)^2 + (6-5)^2}{2}$$

$$= \frac{1+1}{2} \boxed{= 1}.$$

$$\sigma^2 = \frac{(1-5)^2 + (9-5)^2}{2}$$

$$= \frac{(-4)^2 + (4)^2}{2}$$

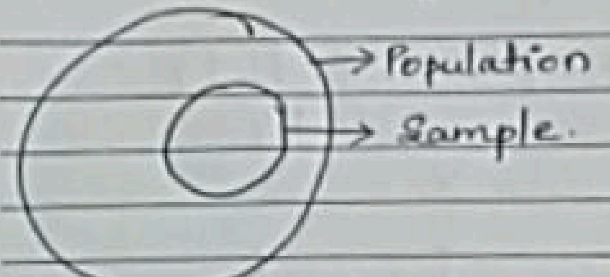
$$= \frac{16+16}{2} = \frac{32}{2} = \boxed{16}.$$

The variance has huge difference.

Semester : 4. Subject : Statistics for AIDS Academic Year: 2023-2024

There are two formulas to calculate variance.

* Population variance
* Sample variance.

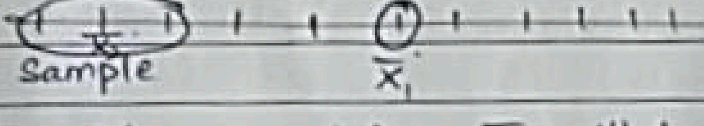


$$\boxed{\sigma^2 = \frac{\sum_{i=1}^{n} (x-\bar{x})^2}{N}}$$

Sample variance

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

* Compare N and n-1, n is greater than n-1
* n-1 is smaller.
* If denominator is smaller then σ^2 is greater.
* We need the value of $\sigma^2 \approx \sigma^2$ (ie) approximately it should be equal. This can be understood by a graph.



If we consider population $\bar{x}_1$ will be mean, whereas if we consider sample $\bar{x}_2$ will be mean. The range is very different between them. To make them approximately equal we have different denominators.

Standard Deviation:-

A quantity expressing by how much the members of a group differ from the mean value for the group.

Example:-

Consider 2 data sets

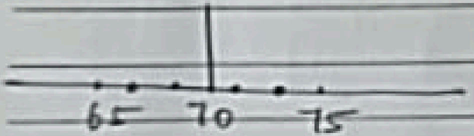
History Test

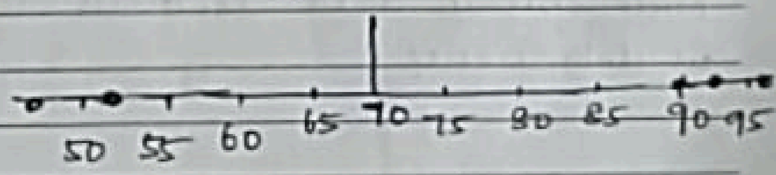
Name	Score	Abs(S-A)
A	75	5
B	72	2
C	68	2
D	65	5
E	67	3
F	73	2
		3.16

Maths Test

Name	Score	Abs
A	93	23
B	96	26
C	43	27
D	47	23
E	51	19
F	90	20
		23

Average = $\overline{x}$ = 70 Average = $\overline{x}$ = 70





Mean is same, but plotting in graph show wide spread of data.

3.16 is the mean absolute deviation for first data set & 23 for the second data set. Here again it is wide spread.

Semester : 4 Subject : Statistics for AIDS Academic Year: 2023 - 2024

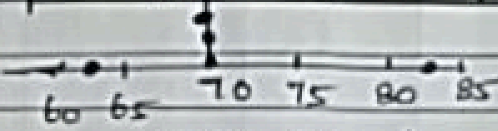
Histogram

Consider a case like this:

Name	Score (S)	Abs	$(S-A)^2$
A	75	5	25
B	72	2	4
C	68	2	4
D	65	5	25
E	67	3	9
F	73	3	9
		3.33	12.66 (Avg)
			3.55 $\sqrt{Avg}$

Name	Score	Abs	$(S-A)^2$
A	83	13	169
B	70	0	0
C	70	0	0
D	63	7	49
E	70	0	0
F	70	0	0
		3.33	36.33 Avg
			6.02 = $\sqrt{Avg}$

$$SD\ \sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$



Wide spread.

Data distribution is wide.

The higher the number the data is more spread. The standard deviation gives you clear result of wide spread data distribution. It is widely used in datascience.