U3A - Maths

Term 2 / Week 6 28-05-21

Standard Deviation

Review

- Last week we discussed various "measures of dispersion" — which is a measure of spread of given data.

- Mean Deviation $= \frac{\Sigma |x-\bar{x}|}{N}$

(not suitable for rigorous maths due to use of "modulus" or "Absolute" value)

- Variance $(s^2) = \frac{\Sigma_1 (x-\bar{x})^2}{N}$

ie, Mean of "square of deviations"
(Suitable for rigorous maths!)

-2-

- To obtain a value which has the same "dimension" as the given raw data, we defined,

Standard Deviation $(S) = \sqrt{\frac{\Sigma_1 (x-\bar{x})^2}{N}}$

ie, $S = \sqrt{(S^2)}$!!

- Also, we have Arithmetic Mean $(\bar{x}) = \frac{\Sigma_1 x}{N}$

- Both Mean $(\bar{x})$ and Standard Deviation (S) are extensively used for (practical) statistical analysis.

- Hence, it is important to have clear under-standing of both!

-3-

Standard Deviation

- Can also be defined as the "typical distance" from the mean.
- For Example, for: $\bar{x} = 8$ & $S = 2$ we have

$(\bar{x} - S) = 8 - 2 = 6$

$(\bar{x} + S) = 8 + 2 = 10$

Typically, majority (68%) of data values lie between 3 & 7

- In fact, in practice 99.7% of the data values lie within 3 std. deviations

For our example, it is

$8 - 3 \times 2 = 2$ } 99.7% of data value
$8 + 3 \times 2 = 14$ } lie bet. 2 and 24!

-4-

- A more efficient equation for the calculation of Std. dev. can be derived as below:

We have

$$S = \sqrt{\frac{\Sigma_1 (x-\bar{x})^2}{N}} = \sqrt{\frac{\Sigma_1 (x^2 - 2x\bar{x} + \bar{x}^2)}{N}}$$

$$= \sqrt{\frac{\Sigma x^2}{N} - \frac{\Sigma_1 2x\bar{x}}{N} + \frac{\Sigma_1 \bar{x}^2}{N}}$$

we have $\bar{x} = \frac{\Sigma_1 x}{N}$ & $\sum_1^N \bar{x}^2 = N.\bar{x}^2$

$$= \sqrt{\frac{\Sigma_1 x^2}{N} - 2\bar{x}\underbrace{\left(\frac{\Sigma_1 x}{N}\right)}_{\bar{x}} + \frac{\cancel{N}.\bar{x}^2}{\cancel{N}}}$$

$$= \sqrt{\frac{\Sigma x^2}{N} - 2\bar{x}^2 + \bar{x}^2} = \boxed{\sqrt{\frac{\Sigma \bar{x}^2}{N} - \bar{x}^2}}$$

-5-

- we have, Std. dev.

$$S = \sqrt{\left(\frac{\Sigma x^2}{N}\right) - \bar{x}^2}$$

This is a simpler formula for hand calculations.

- Most std. dev. Calculations are done using "statistic calculators" !!
- You can download & instal "Scientific Calculator 995" developed by 'C20 studio"
- Identical to "Casio fx 995" and it is "Ad free"

-6-

- We have two alternative equations for standard deviation, which can be very confusing!

- Std. dev (σ) for population $= \sqrt{\dfrac{\Sigma (x-\mu)^2}{n}}$

where μ - population Mean
n - No. of values in population

- Std. dev for samples $S_N = \sqrt{\dfrac{\Sigma_1 (x-\bar{x})^2}{N}}$

$$S_{N-1} = \sqrt{\frac{\Sigma_1 (x-\bar{x})^2}{(N-1)}}$$

$\bar{x}$ - sample Mean
N - No. of values in Sample data

-7-

- Why (N-1) is used for Sample data (Good question!)

Explanation No. 1

- sum of deviations $\left(\sum_1^N (x-\bar{x})\right)$ adds up to zero, hence there are only (N-1) degrees of freedom!

- This is an escapist explanation! Even population has $\sum_1^n (x-\mu) = 0$, so we should use $(n-1)$ for population also!

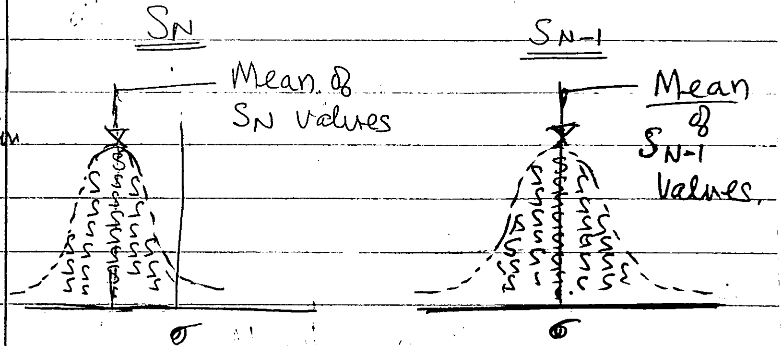
Explanation No. 2

- It is the "Bessel Correction factor", which helps to

-8-

obtain a more accurate estimate of "Population Std. Dev." from sample data.

- The distribution of S_N & S_{N-1} value for various samples from a population are as below.



σ - Population Std. Dev.

- Corrected Std Dev (S_{N-1}) $= \sqrt{\left(\dfrac{\Sigma (x-\bar{x})^2}{N}\right)\left(\dfrac{N}{N-1}\right)}$ Bessel Correction Factor

$$S_{N-1} = S_N \cdot \left(\sqrt{\frac{N}{(N-1)}}\right)$$