Question :- Why did we divide by n-1 for sample variance?

→ The variance is the average squared deviation from the population mean

population variance

↓

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

sample variance

↓

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

→ As per the population and sample defination we should provide conclusion for population data by taking sample data

→ Which means population mean and sample mean are equal

i.e $\bar{x} \approx \mu$

→ and population variance & sample variance also equal

i.e $\sigma^2 \approx s^2$

Eg:-

$$x = \{ \overbrace{1, 2}^{\uparrow \text{ sample}}, 3, 4, 5, 6, 7 \dots \}$$

↓

lets consider mean μ will be middle of population data

→ I will take 1, 2, 3 as sample data, mean will be somewhat nearer to population data

→ If we take variance for sample data is $\sum_{i=1}^{n} \frac{(x-\bar{x})^2}{n}$ as then it will be very far from population data

i.e $\bar{x} <<< \mu$ which means underestimating the population variance

$s^2 <<< \sigma^2$

2022.6.27 20:48

If we take variance of sample data as $\sum_{i=1}^{n} \frac{(x-\bar{x})^2}{n-1}$ gape between sample data variance and population variance will decrease and will be nearer, so that's why we are considering n-1 for sample variance.

This is also called Bessel correction.

2022.6.27 20:48