Why sample variance is n-1 ?

Population

$$\mu = \sum_{i=1}^{N} \frac{x_i}{N}$$

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

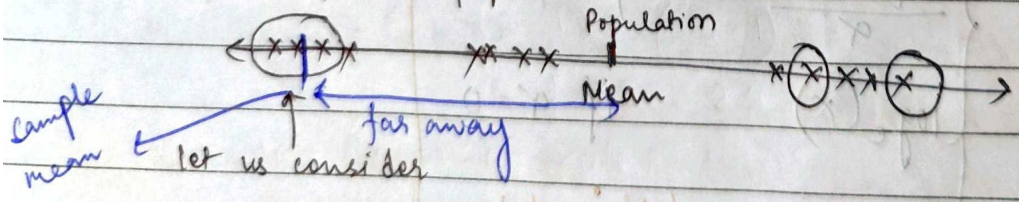
Sample

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

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

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n} \rightarrow \text{It is a biased estimation}$$

let us consides, from a population data we have some sample points

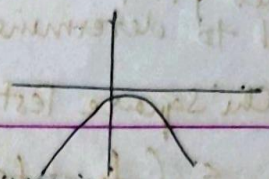


we have ~~see~~ selected this data points, then my sample mean will also be located there

In order to overcome the biased estimate, we divide by $(n-1)$

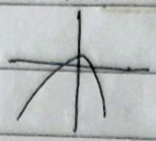
i) when divide by n :-

→ ~~estimate~~ $\bar{x} << \mu$



Sample variance ~~mean~~ << Population ~~mean~~ variance

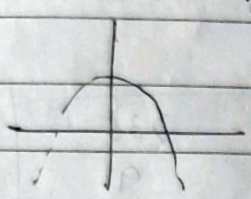
ii) when divide by $(n-1)$:-



$\bar{x} \approx \mu$

iii) when divide by $(n-2)$:- $n-3$

$\bar{x} >> \mu$. $\bar{x} >>> \mu$.



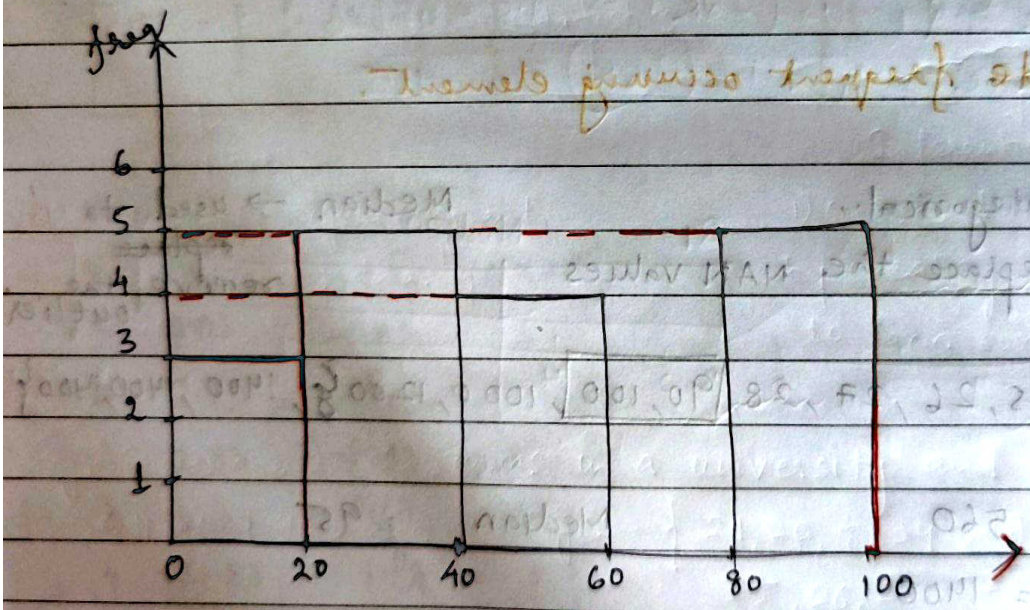
~~When~~ When dividing by $(n-1)$, the sample variance ~~is~~ tends to approach to the true population variance..

Assignment - 1 (19-06-2022)

1) {10, 13, 18, 22, 27, 32, 38, 40, 45, 51, 56, 57, 88, 90, 92, 94, 99}

bins = 5

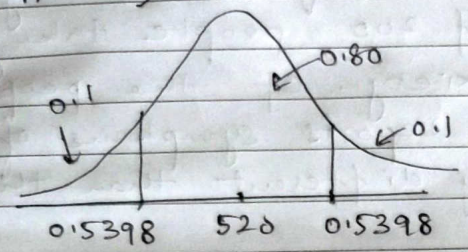
bin size = 20



Assignment 2.

In a quant test of the CAT exam, the population standard deviation is known to be 100. A sample of 25 tests taken has a mean of 520. Construct an 80% C.I about the mean.

$n = 25$, $\bar{x} = 520$, $\sigma = 100$, $\alpha = 0.2$, C.I = 80%.



C.I = Point Estimate ± Margin Error.

$$C.I = \bar{x} \pm t_{\alpha/2}\left(S/\sqrt{n}\right)$$

$$C.I = 520 + t_{0.1} \times \frac{100}{\sqrt{25}}$$

$df \Rightarrow n-1$
$= 25-1$
$= 24.$

$$C.I = 520 \pm 1.318 \times \frac{100}{\sqrt{25}}$$

$$C.I = 520 - 1.318 \times \frac{100}{\sqrt{25}}$$

$C.I = 493.64$ //

$$C.I = 520 + 1.318 \times \frac{100}{\sqrt{25}}$$

$C.I = 546.36.$

$C.I = [493.64 \longleftrightarrow 546.36]$.

$H_0 = \{\mu = 520\}$
$H_1 = \{\mu \neq 520\}$ → Accept Null Hypothesis.

Assignment 4.

What is the value of 99 percentile

→ {2, 2, 3, 4, 5, 5, 5, 6, 7, 8, 8, 8, 8, 8, 9, 9, 10, 11, 11, 12}

$$\text{Value} = \frac{\text{Percentile} \times (n+1)}{100}$$

$$= \frac{99}{100} \times 21 = 20.79 \rightarrow \text{not present}$$

↓ 20th Index

12 // Ans.

Assignment

(1) A car company believes that the percentage of residents in city ABC that owns a vehicle is 60% or less. A sales manager disagrees with this. He conducts a hypothesis testing surveying 250 residents and found that 170 responded yes. to owning a vehicle.

a) State the Null & Alternate Hypothesis

b) At 10% significance level, is there enough evidence to support the idea that vehicle ownership in city ABC is 60%

i) Null Hypothesis : $H_0 : p_0 = 60\%$

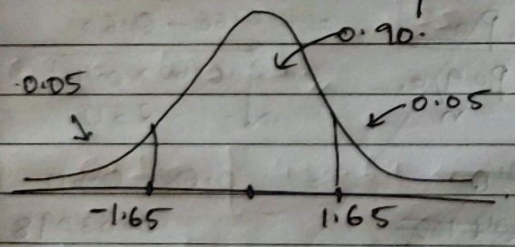
$H_1 : p_0 \neq 60\%$

$n = 250, \quad x = 170.$

$$\hat{p} = \frac{x}{n} = \frac{170}{250} = 0.68$$

$$q_0 = 1 - p_0 = 1 - 0.60 = 0.40.$$

$\alpha = 0.10, \quad C.I = 90\% = 0.90.$



$$\text{Z-test} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

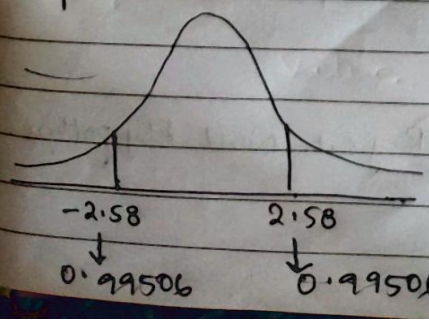
$$= \frac{0.68 - 0.60}{\sqrt{\frac{0.60 \times 0.40}{250}}} \qquad = \frac{0.68 - 060}{0.03098}$$

$$\frac{> 0.08}{0.03098} = 2.58.$$

Z-test = 2.58.

$2.58 > 1.65$ Reject Null Hypothesis.

p-value



$1 - 0.99506 = 0.00494$

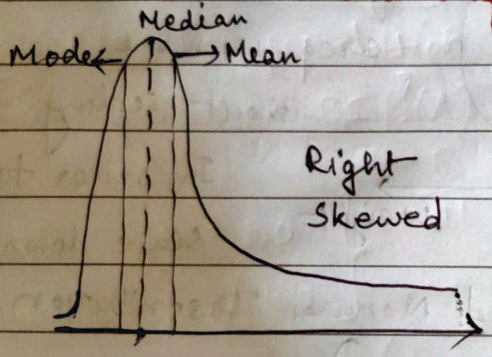
p-value $= 0.00494 + 0.00494 = 0.0098$

p-value $< \alpha$

$0.0098 < 0.10$ → Reject Null Hypothesis

Assignment 2 - 28-06-2022.

What is the relationship between the below 2 distribution



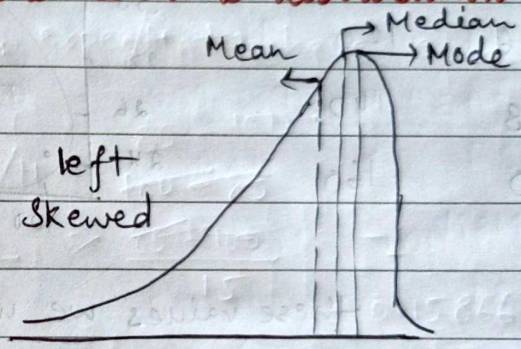
Positive skew

Eg: i) Wealth distribution

ii) Length of comments in you tube channel.

Relationship:

Mean > Median > Mode.



Negative skew.

Eg: i) life span of human being

Relationship:

Mode > Median > Mean