1) Why sample variance is n-1?

Population

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

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

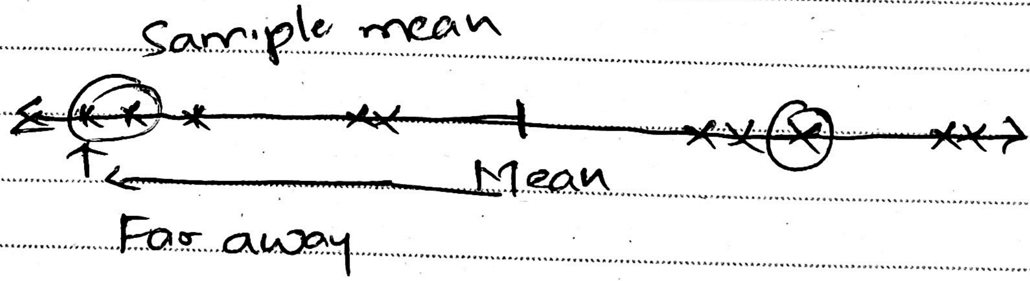
Sample

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

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

To get unbaised estimation we use n-1 at the base.

Lets consider, from a population data we have some sample points

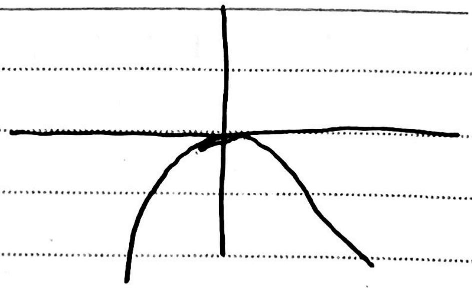


Let us consider, we have selected this data points, then my sample mean will also be located there.

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

(i) If we devide by n :-

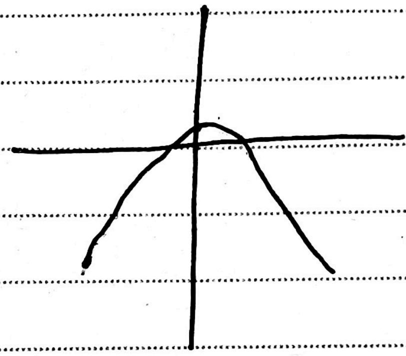
$\bar{x} << \mu$



Sample variance << population variance

(ii) If we devide by $n-1$

$\bar{x} \simeq \mu$



If we devide by $n-2$, $n-3$ or so on

$\bar{x} >> \mu$, $\bar{x} >>> \mu$ will be increasing

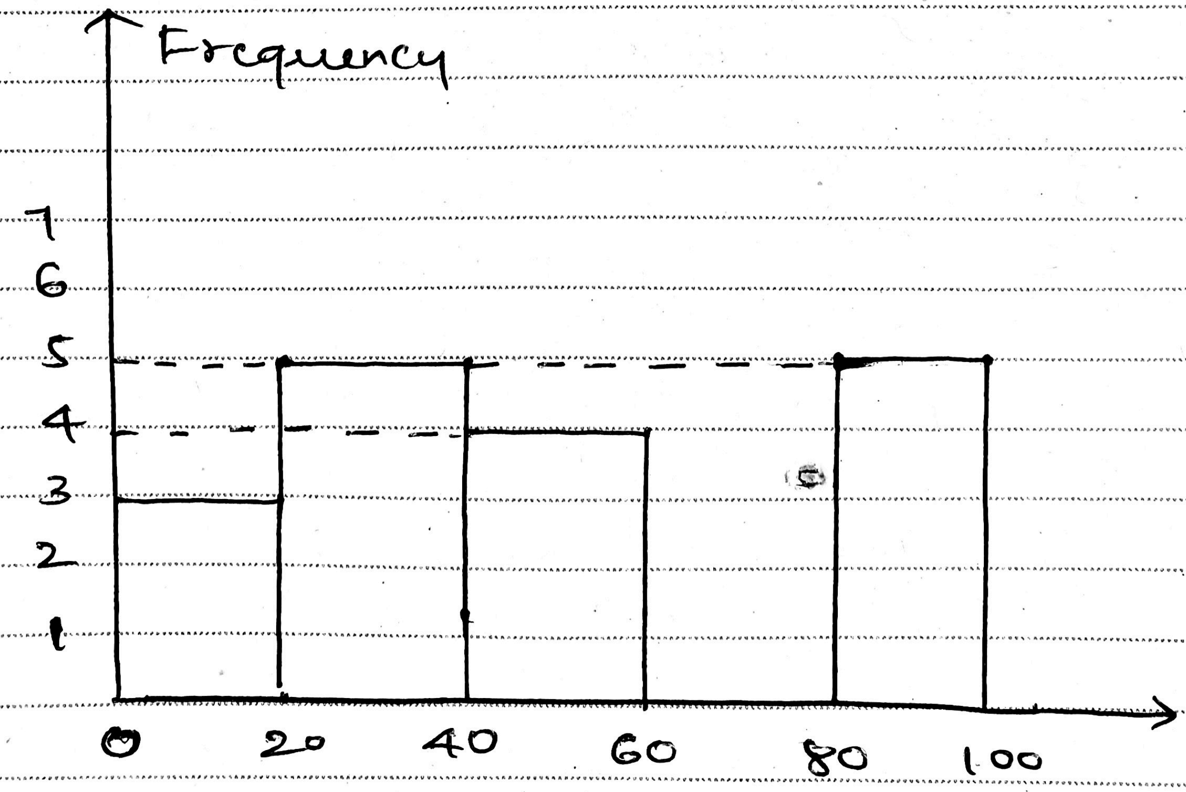
- When we divide by $n-1$ the sample variance tend to appear to the true population variance.

Date: / /

Create a Histogram

2) {10, 13, 18, 22, 27, 32, 38, 40, 45, 51, 56, 57, 88, 90, 92, 94, 99}

bins = 5 bin size = 20



3) In a quant test of 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% CI about the mean.

$n = 25$, $\bar{x} = 520$, $\sigma = 100$, $\alpha = 0.02$

$CI = 80\%$

CI = Point of estimate, ± Margin error

$$CI = \bar{x} \pm t_{\alpha/2} \left(s/\sqrt{n} \right)$$

~~CI = 520 + t0.~~

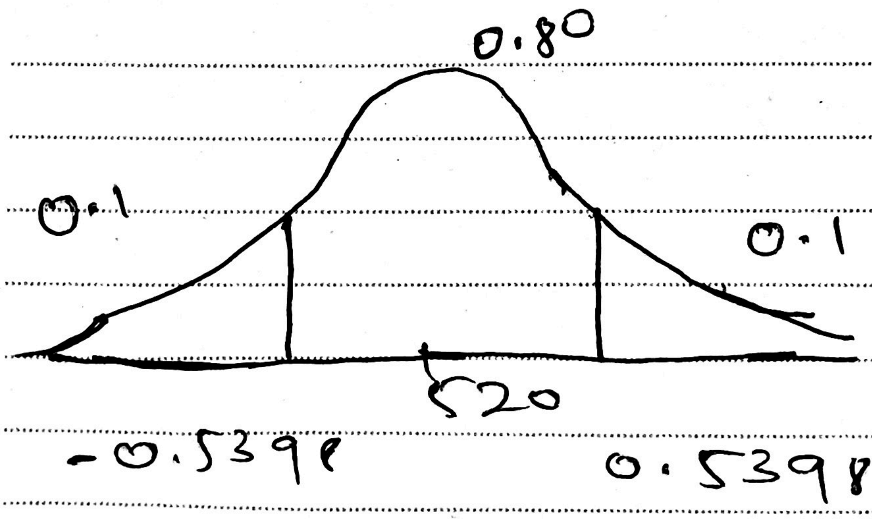
$$CI = 520 \pm t_{0.1} \times \frac{100}{\sqrt{25}}$$

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

$$CI = 520 \pm 1.318 \times \frac{100}{\sqrt{25}}$$

$$CI = 520 + 1.318 \times \frac{100}{5} \Rightarrow 493.64$$

$$CI = 520 - 1.318 \times \frac{100}{5} \Rightarrow 546.36$$



$$CI = [\,493.64 \longleftrightarrow 546.36\,]$$

$H_0 = \{ \mu = 520 \}$

$H_1 = \{ \mu \neq 520 \}$

Accept Null Hypothesis.

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}

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

$$= \frac{99}{100} \times 21 \Rightarrow 20.79$$ → So consider nearest index i.e 20

$\Rightarrow \underline{12}$

5. A car company belives that the percentage of residents in city ABC, that owns a vechile 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 vechile.

a) State the Null & Alternate Hypothesis.

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

(i) Null Hypothesis: $H_1 : P_1 \neq 60\%$

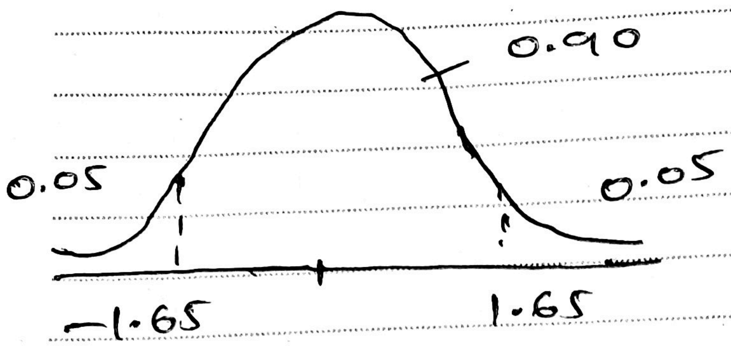
$H_0 : P_0 = 60\%$

$n = 250, \quad x = 170$

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

$$Q_0 = 1 - P_0 = 1 - 0.60 = 0.40$$

$\alpha = 0.10, \quad CI = 90\% = 0.90$



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

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

$$\Rightarrow \frac{0.08}{0.03098} = 2.58$$

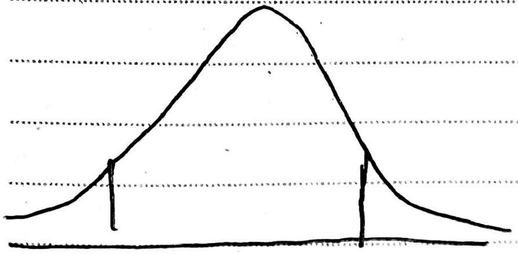
$2.58 > 1.65$ Reject null Hypothesis

p-value

$1 - 0.99506 = 0.00494$

p value $= 0.00494 + 0.00494$

$= 0.0098.$



-2.58 2.58

$\downarrow$ $\downarrow$

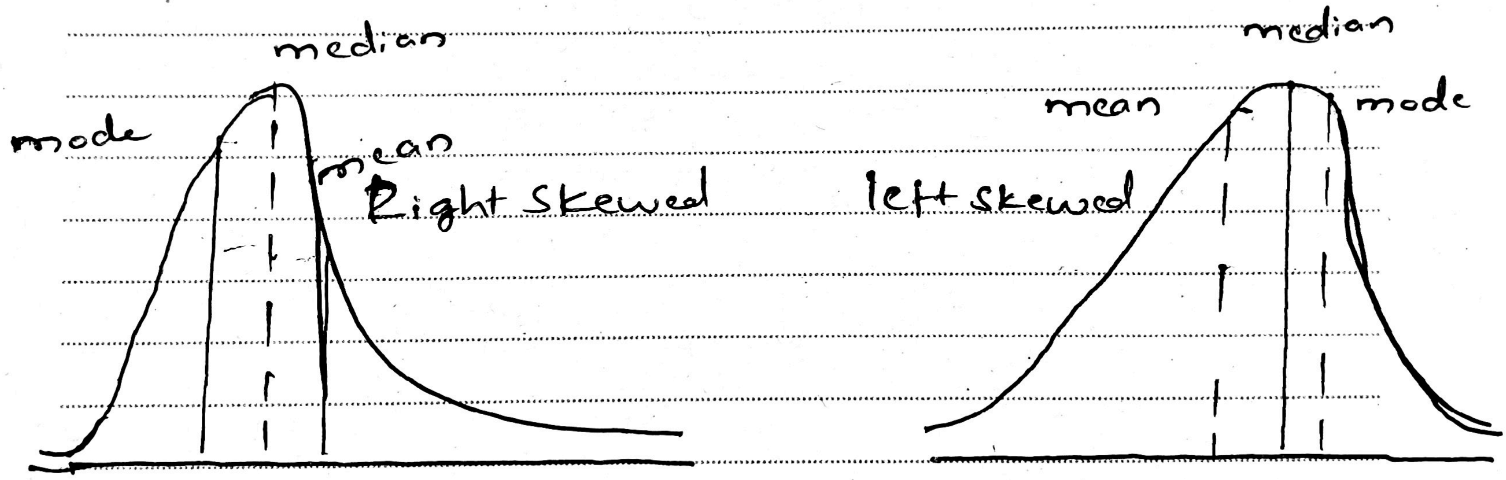
0.99506   0.99506

Pvalue $< \alpha$

$\boxed{0.0998 < 0.10}$

Reject null Hypothesis.

6.) What is the relation between the below 2 distribution.



positive skew   Negative skew

(i) mean > Median > mode   mode > median > mean

Ex: wealth distribution   : lifespan of human