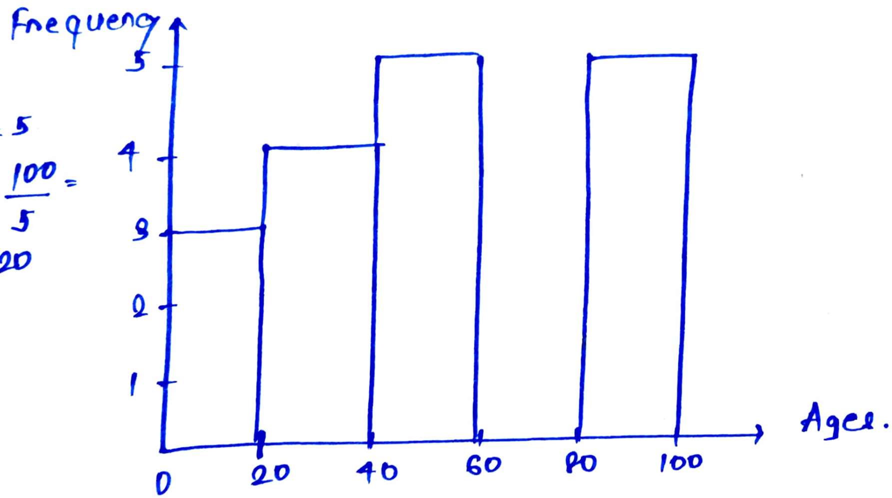
ASSIGNMENT-1

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

Create a histogram

Given: Bins = 5

∴ Bin size = $\frac{100}{5}$ = 20



Bins	Frequency
0-19	3.
20-39	4
40-59	5.
60-~~7~~79	0
80-99	5.

A sample of 25 test takers has a mean of 520.

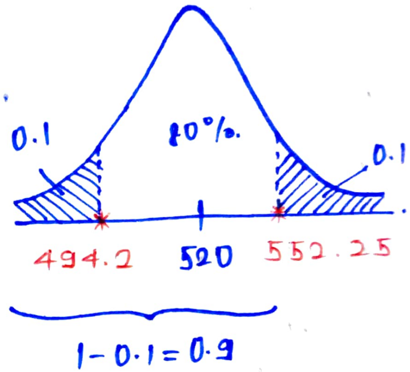
Construct an 80% CI about the mean.

Soln) Given :- $\sigma = 100$, $n = 25$, $\bar{x} = 520$

Significance value, $\alpha = 1 - CI$

$= 1 - 0.8$

$= 0.2$



$$Z_{\alpha/2} = Z_{0.2/2} = Z_{0.1} = 1.29$$

Lower fence $= \bar{x} - Z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$

$$= 520 - 1.29 \times \frac{\cancel{100}\ 20}{\sqrt{25} = \cancel{5}}$$

$= 494.2$

Higher fence $= \bar{x} + Z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$

$= 520 + 1.29 \times 25$

$= 552.25$

Q) 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 and alternate hypothesis.

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

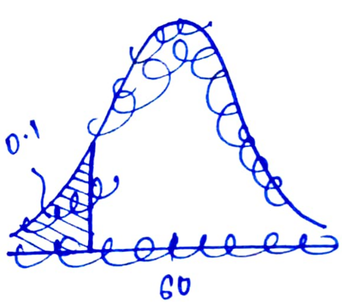
Soln) i) Null Hypothesis, H0 : $P_0 \leq 60\%$

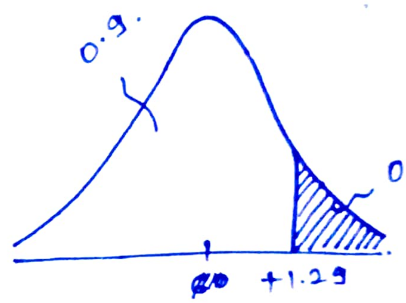
Alternate Hypothesis H1: $P_0 > 60\%$ {One-Tailed Test}

Given: $n = 250$, $x = 170$, $\hat{P} = \frac{x}{n} = \frac{170}{250} = 0.68$, $q_0 = 1 - P_0 = 1 - 60\% = 0.4$

ii) $\alpha = 10\%$,

iii) Decision ~~Rule~~ Boundary





iv) Z-test with proportion

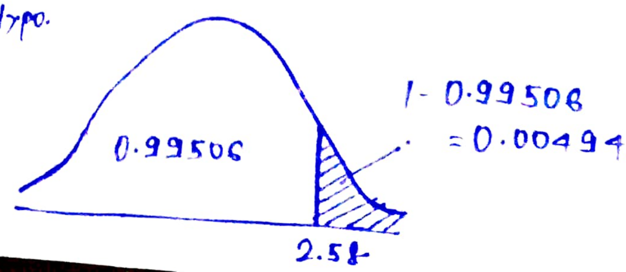
$$Z = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0 q_0}{n}}} = \frac{0.68 - 0.60}{\sqrt{\frac{0.6 \times 0.4}{250}}} = 2.58.$$

v) Conclusions :

$\because Z = 2.58 > 1.29$ $\therefore$ Reject the Null Hypothesis {Vehicle owner in ABC city is > 60%}

Using p-value

p-value = 0.00494 < α (0.1) => Reject Null Hypo.



Que4). 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

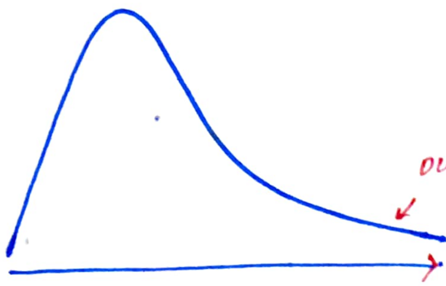
$$\text{Index - Value} = \frac{\text{Percentile}}{100} * (n+1)$$

$$= \frac{99}{100} * (20+1)$$

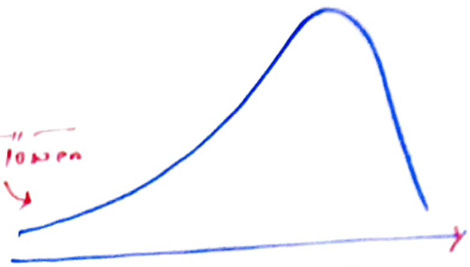
$$= 20.79$$

$\therefore$ Value of 99 percentile is 12

(a) Relationship between mean, median, mode in the two dist. ASSIGNMENT II



outliers will be tnt in higher side — " — lower — " —



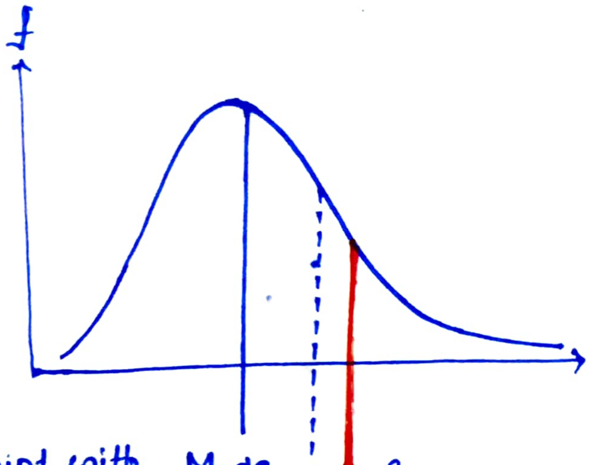
Rt skewed

Left skewed.

Ex:- Life span

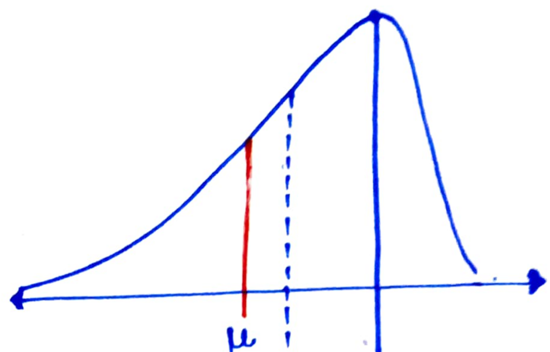
Mean > Median > Mode

Mode > Median > Mean.



(Point with highest frequency) Mode

μ {The mean value will be shifted to right because the outliers is tnt on right side}



μ

Mode

Median.

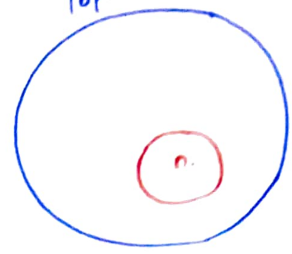
Median is the middle value of the dataset when sorted. ∴ We need to find a value for which half of the values are above that value and 1/2 of the values are below

In other words, at what value do we have equal ~~value~~ area above and below that value.

Assignment-2

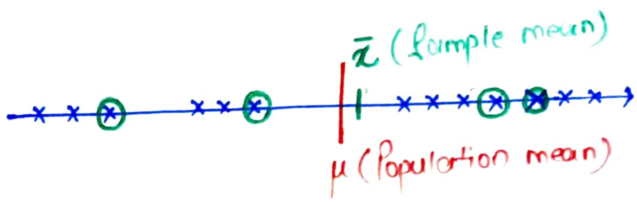
Q) Why sample ~~variation~~ variance is divided by (n-1)

Population = N

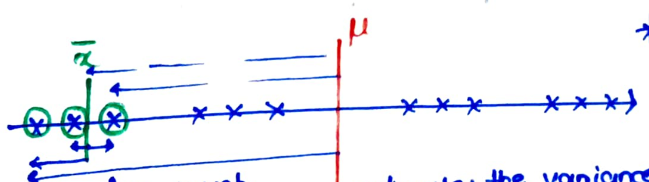


	Population (N)	Sample (n)
Mean	$\mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$	$\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$
Variance	$\sigma^2 = \dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$	$s^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$ } Unbiased estimation.

$n = 3$ $\qquad$ $N = 13$



$\bar{x} \approx \mu$. In this case the sample variance and population variance maybe similar.



* The distance of the data points to the sample mean will always be shorter, compared to the dist. to the population mean. ∴ When we estimate the variance, based on estimated mean, we will underestimate the population variance

This will give much lower estimate than the true variance from the actual population mean. We are underestimating the true population variance.

When divided by $(n-1)$, $\sigma^2 \approx s^2$

$\bar{x} = 166$, $\mu = 170$, $\sigma^2 = 49$

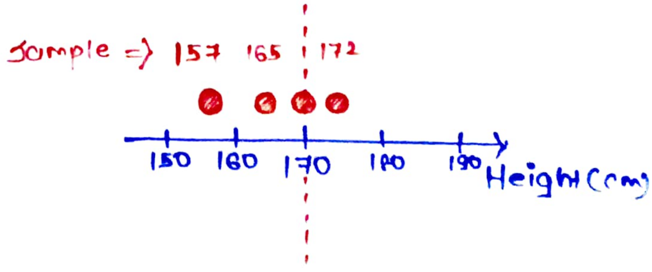
i) $s_1^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = 44.7 \approx 49$

ii) $s_2^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = 33.5$

iii) $s_3^2 = \dfrac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} = 49.5 \approx 49$

$\mu = 170$

$\sigma^2 = 49 \therefore \sigma = 7$

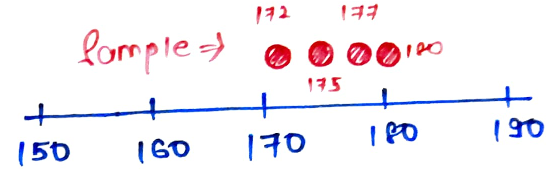


- $\bar{x} = 176$, $\sigma^2 = 49$

$s_1^2 = 11.3$

$s_2^2 = 8.5$

$s_3^2 = 44.5$



- If we repeat the ~~previus~~ previous two processes 10,000 times and cal. the mean.

$$\text{Mean of the } s_1^2 = \frac{44.7 + 11.3 + \dots + 71.3}{10000} = 48.9 \approx 49$$

—" —" — $s_2^2 = 36.71 << 49$

—" —" — $s_3^2 = 49.01 \approx 49$.