Subject :                                        Date:

Central tendency:

* Mean $= \frac{\sum_{i=1}^{n} x_i}{n}$  * Ungrouped Data

Median → $n$ → odd $= X_{(\frac{n+1}{2})}$

Median → $n$ → even $= \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}$

Mode → Value with most frequency.

* Mean  * Grouped Data.

① Midpoint $= \frac{\text{Start} + \text{end}}{2}$

② Multiply Midpoint by the frequency $(f.x)$.

③ Mean $= \sum_{i=1}^{n} (f.x)$

Note: Mean > Median > Mode → Right Skewed.

Mean < Median < Mode → Left Skewed.

Mean = Median = Mode → Symmetric

Range (R) = largest value − Smallest value.

Interquartile Range (IQR)

① find $Q_2$ = Mean of data.

② find $Q_1$ = Mean of data befor $Q_2$

③ find $Q_3$ = Mean of data after $Q_2$

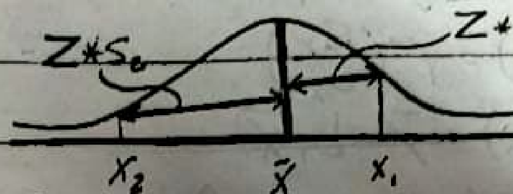
Subject:  Date:

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

Standard deviation $(S) = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n-1}}$

Z-Score $(Z) = \frac{x - \bar{x}}{S}$

$x = \bar{x} + Z * S$

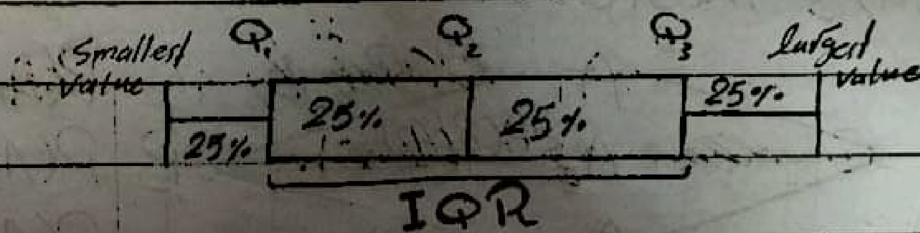



→ Positive → score above The mean

→ Zero → score is mean score.

→ Negative → score below The mean.

Box Plot:




Correlation (r):

$r = \frac{\sum_{i=1}^{n} Z_{x_i} * Z_{y_i}}{n-1}$

Sign
- +ve → Values increase or decrease Together
- −ve → When value decrease as other increase.

Value
- → Weak → 0.1, 0.2, 0.3
- → Moderate → 0.4, 0.5, 0.6
- → Strong → 0.7, 0.8, 0.9
- → Perfect → 1

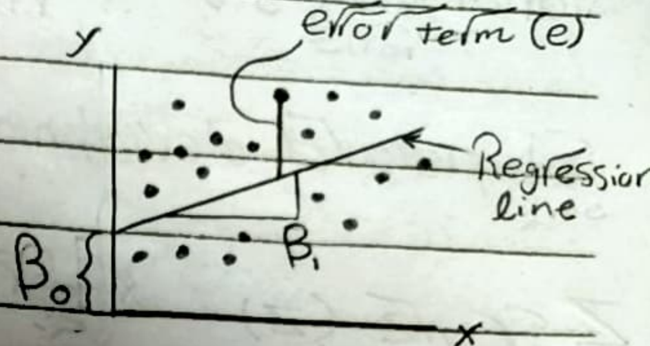
* Note: $0.6898 \approx 0.7$

$0.96978 \not\approx 1$

Subject:

Date:

## Simple Linear Regression:




$y = \beta_0 + \beta_1 X + e$

$E(y) = \hat{y} = b_0 + b_1 x$

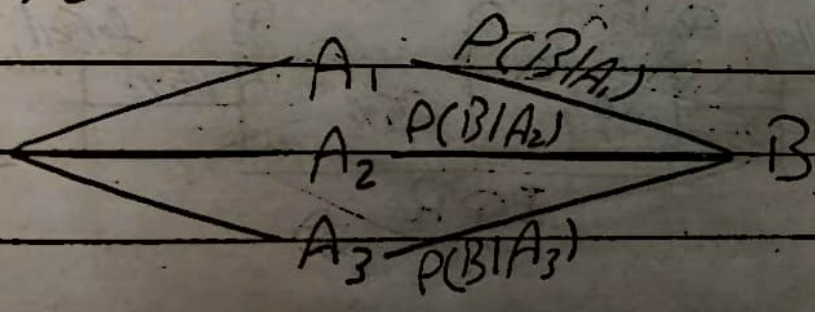
$b_1 = E(\beta_1) = r\left(\frac{S_y}{S_x}\right)$

$b_0 = E(\beta_0) = \bar{y} - b_1 \bar{x}$

$\text{Error }(e) = |y - \hat{y}|$

* Cofficient of determinition $(r^2) = \left(\frac{\sum_{i=1}^{n} Z_{x_i} * Z_{y_i}}{n-1}\right)^2$

## Bay's Theorem:-




$P(A_1 \cap B) = P(B|A_1)\,P(A_1)$

$P(A_2 \cap B) = P(B|A_2)\,P(A_2)$

$P(A_3 \cap B) = P(B|A_3)\,P(A_3)$

$P(B) = \text{Sum}$

* Note $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Subject: Date:

* Random Variable:

A Set of Possible values from Random experiment.

(a) Discrete: not Interval.

(b) Countinuous: Interval.

* Probability distribution [$f(x)$]:

| (P.M.F) | (P.D.F) |
|---|---|
| Probability mass function | Probability density function |
| Gives The Probability That a discrete random variable. | Gives The Probability that a Countinuous random Variable |
| $f(x)$ is P.M.F if: | $f(x)$ is P.D.F if: |
| (1) $0 \leq f(x) \leq 1$ | (1) $0 \leq \int_A f(x)\,dx \leq 1$ |
| (2) $\sum_{x_1}^{n} f(x) = 1$ | (2) $\int_R f(x)\,dx = 1$ |

Binomial distribution:

To study number of successes in a sequence of $n$ independent expirments.

$n \rightarrow$ total number of trails.

$p \rightarrow$ Probability of success.

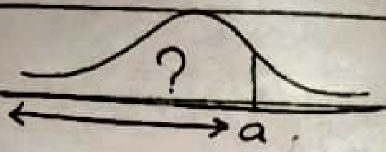
$q \rightarrow$ Probability of failure.

$$f(x) = P(x=r) = \binom{n}{r} p^r q^{n-r}$$

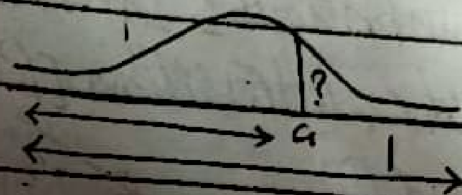
# Normal distribution:

$E(X) = \mu$ , $Var(X) = \sigma^2$ , $Z = \frac{X-\mu}{\sigma}$

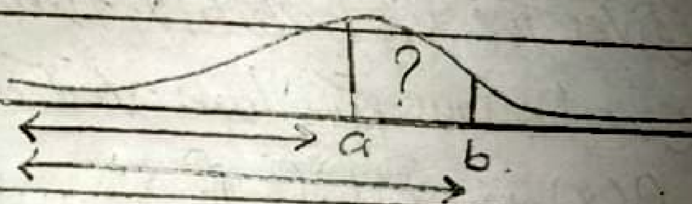
* $P(Z<a)$ = Z Table




* $P(Z>a) = 1 - P(Z<a)$



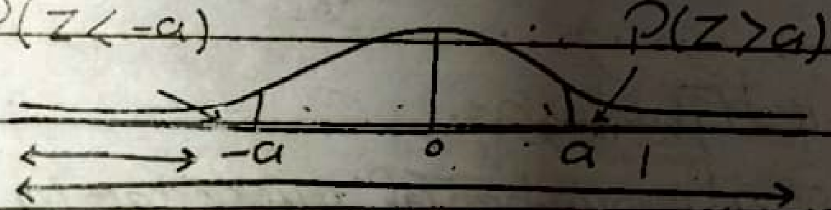

* $P(a<Z<b) = P(Z<b) - P(Z<a)$




* $P(Z<-a) = P(Z>a) = 1 - P(Z<a)$

$= P(Z<-a) + P(Z<a) = 1$

$P(Z>a) = P(Z<-a)$




* Note: $P(Z<a) = 1$ if $a > 3.49$

$P(Z<a) = 0$ if $a < -3.4$

# Central Limit Theorem:

Mean $(\mu) = \frac{\sum_{i=1}^{n} x_i}{n}$

$Z = \frac{X-\mu}{\sigma}$

Variance $(\sigma^2) = \frac{\sum_{i=0}^{n} (x_i - \bar{x})^2}{n-1}$

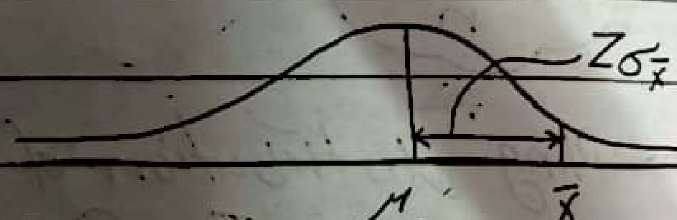
Subject: Date:

## Sampling distribution of sample mean:

$$\mu_{\bar{x}} = E(\bar{x}) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{\cancel{n}}\cancel{n}\mu \quad \therefore \boxed{\mu_{\bar{x}} = \mu}$$

$$\sigma_{\bar{x}}^2 = var(\bar{x}) = var\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n^2}var\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} var(x_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^{\cancel{2}}}\cancel{n}\sigma^2 \quad \therefore \boxed{\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}} \quad \boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

$$\therefore Z_{\bar{x}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$




## Sampling distribution of difference betweet Two means:

$$\mu_{\bar{x}_1 - \bar{x}_2} = E(\bar{x}_1 - \bar{x}_2) = E(\bar{x}_1) - E(\bar{x}_2) = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$$

$$\sigma_{\bar{x}_1 - \bar{x}_2}^2 = var(\bar{x}_1 - \bar{x}_2) = var(\bar{x}_1) + var(\bar{x}_2) = \sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_1^2}{n_2}$$

$$\therefore \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$Z_{\bar{x}_1 - \bar{x}_2} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Subject: ______ Date: ______

## Sampling distribution of sample Proportion:

$$\mu_{\hat{p}} = E(\hat{p}) = E\left(\frac{x}{n}\right) = \frac{1}{n}E(x) = \frac{1}{n}(nP) = P \qquad \boxed{\therefore \mu_{\hat{p}} = P}$$

$$\sigma_{\hat{p}}^2 = Var(\hat{p}) = Var\left(\frac{x}{n}\right) = \frac{1}{n^2}Var(x) = \frac{1}{n^2}(nPq) \qquad \therefore \boxed{\sigma_{\hat{p}}^2 = \frac{Pq}{n}}$$

$$\therefore \boxed{\sigma_{\hat{p}} = \sqrt{\frac{Pq}{n}}}$$

$$Z = \frac{\hat{P} - P}{\sqrt{\frac{Pq}{n}}}$$

## Sampling distribution of difference between Two Proportions:

$$\mu_{\hat{P}_1 - \hat{P}_2} = E(\hat{P}_1 - \hat{P}_2) = E(\hat{P}_1) - E(\hat{P}_2) = \mu_{\hat{P}_1} - \mu_{\hat{P}_2} = P_1 - P_2$$

$$\boxed{\mu_{\hat{P}_1 - \hat{P}_2} = P_1 - P_2}$$

$$\sigma_{\hat{P}_1 - \hat{P}_2}^2 = Var(\hat{P}_1 - \hat{P}_2) = Var(\hat{P}_1) + Var(\hat{P}_2) = \sigma_{\hat{P}_1}^2 + \sigma_{\hat{P}_2}^2 = \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}$$

$$\therefore \boxed{\sigma_{\hat{P}_1 - \hat{P}_2}^2 = \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}$$

$$\boxed{\sigma_{\hat{P}} = \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}$$

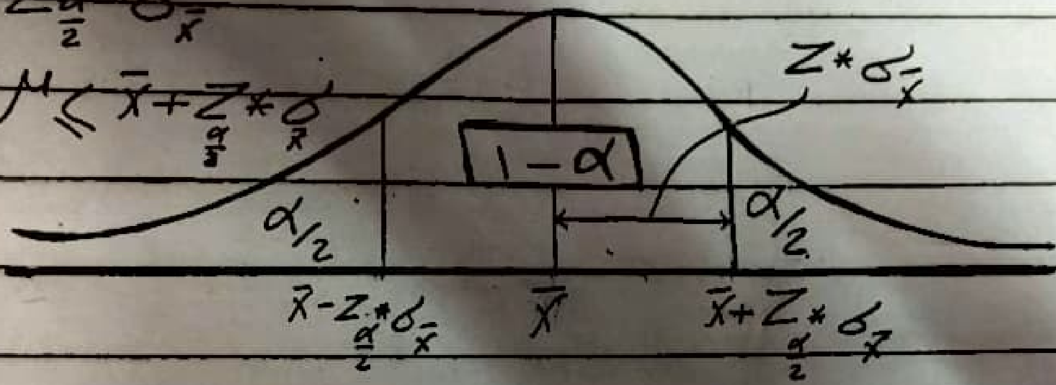
$$Z_{\hat{P}_1 - \hat{P}_2} = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}$$

Subject: Date:

# Confidence Intervals:

$CI = \bar{x} \pm Z_{\frac{\alpha}{2}} * \sigma_{\bar{x}}$

$\bar{x} - Z_{\frac{\alpha}{2}} * \sigma_{\bar{x}} \leqslant \mu \leqslant \bar{x} + Z_{\frac{\alpha}{2}} * \sigma_{\bar{x}}$




1. Find number of observations ($n$) and Calculate sample mean $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
2. Deside $CI = (1-\alpha)\%$
3. Population standerd deviation ($\sigma$)

if ($\sigma$) Known
use Z table
$CI = \bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

if ($\sigma$) unknown
(S) Known

if $n \geqslant 30$
use Z-table
$CI = \bar{x} \pm Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$

if $n \leqslant 30$
use t-table
$CI = \bar{x} \pm t_{\frac{\alpha}{2}, V} \frac{S}{\sqrt{n}}$
$V = n-1$