Sampling Distribution of the sample mean ($\bar{X}$)

$$\boxed{X \simeq N(\mu, \sigma^2)}$$

توزيع

$$\bar{X} \simeq N(\mu_{\bar{x}}, \sigma^2_{\bar{x}})$$

$$\boxed{\bar{X} \simeq N\left(\mu, \frac{\sigma^2}{n}\right)}$$

$\because Z_x = \dfrac{x - \mu}{\sigma}$

$$\boxed{\therefore Z_{\bar{x}} = \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}$$

Example:→

Employee Salary of the Bank is Normal Distribution with Mean 520 and Variance ~~1225~~ 1225

(1) Find the sample Dist of the sample mean of four employees.

Salary → $X \sim N(520, 1225)$.

(1) $n = 4$

$$\bar{X} \sim N\left(520, \frac{1225}{4}\right)$$

$$\bar{X} \sim N(520, 306.25)$$

$$\boxed{\text{observe:} \rightarrow \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

③ find the propability of the sample Mean that is greater than 500

$P(\bar{X} > 500)$

$$Z_{\bar{x}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \qquad \frac{500 - 520}{35/2}$$

$$= -1.14$$

$P(Z_{\bar{x}} > -1.14) = 0.8729$

$P(Z_{\bar{x}} < 1.14) = 0.8729$

Example [2] :→

$X \sim N(15, (24)^2) \qquad n = 16$

① $P(\bar{X} \leq 6)$ ② $P(\sum x_i < 32)$

$\cancel{X}$ $\mu = 15$

$\sigma = 24$

$$Z = \frac{6.5 - 15}{24/4} = -1.41$$

$P(\cancel{Z} \leq -1.41)$

$P(\cancel{Z} \geq 1.41) =$

② $P(\sum x_i < 32)$ بالقسمة ÷ n

$= P\left(\frac{\sum x_i}{n} < \frac{32}{16}\right)$

$= P(\bar{x} < 2)$

$, \quad Z_{\bar{x}} = \frac{2 - 15}{24/4} = -2.16$

$P(Z_{\bar{x}} > 2.16)$

Ⅱ Central Limit Theorem (CLT) :→

μ , σ^2 , n

Proprities :→

μ , σ^2 , n

شرط

$\bar{x} \overset{app}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$, if $n \geq 30$

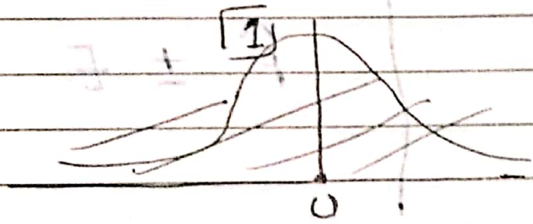
[3] σ known $\quad \bar{X} \quad$ σ unknown

$n \ngtr 30$

n is small

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \qquad t = \frac{\bar{X} - \mu}{s / \sqrt{n}}$$

t - distrubtion :→



degree of freedom (d.f) = n - 1

Confidence Intervals

Parameter

Point estimate $\pm$ E

$\mu \rightarrow \bar{x}$

$P \rightarrow \hat{P}$

Population — Sample

$\boxed{\text{erorr}}$

Confidence Interval

$\boxed{C.I}$

e.g $\rightarrow$

$\bar{x} \pm E$

$\hat{P} \pm E$

[1] Confidence Intervals for μ :$\rightarrow$

steps :$\rightarrow$

(1) Confidence level percent $\Longrightarrow \boxed{1-\alpha}$

$\boxed{\alpha} \longrightarrow \boxed{\frac{\alpha}{2}}$

(2) $\boxed{d.f = n-1}$

(3) $Z_{\alpha/2}$ | $t_{\alpha/2}$

اجيب منين جدول Z, t

if σ known (or) $n \geq 30$ | σ unknown, $n \ngtr 30$

$\rightarrow$ Concept

$P(Z > Z_{\alpha/2}) = \alpha/2$

$P(Z < Z_{\alpha/2}) = 1 - \alpha/2$

if $\alpha/2 = 0.1$

d.f $= n-1$

d.f	t.. 1
3	$\square \rightarrow \boxed{t_{\alpha/2}}$

4 law of conf. C.I for μ ?

$\bar{X} \pm E \Rightarrow (\bar{X}-E, \bar{X}+E)$

if ~~α~~ σ known

$\therefore \bar{X} \pm Z_{\alpha/2} * \frac{\sigma}{\sqrt{2}} \to E$

standard divation

if σ unknown

$\therefore \bar{X} \pm t_{\alpha/2} * \frac{S}{\sqrt{n}} \to E$

Proprities :→

1 Central of C.I = $\bar{X}$

$$\frac{(\bar{X} - \cancel{Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{2}}}) + (\bar{X} - \cancel{Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{2}}})}{2}$$

$= \frac{\cancel{2}\bar{X}}{\cancel{2}} = \boxed{\bar{X}}$

[2] Error, Margin Error, Bound Error

$$\boxed{E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}} \rightarrow \text{Concept}$$

$\bar{x} - \bar{x} - E = \boxed{-E}$, $\bar{x} - \bar{x} + E = \boxed{E}$

$\bar{x}-E$ $\qquad\qquad$ $\bar{x}$ $\qquad\qquad$ $\bar{x}+E$

[3] length of C.I

$$L = 2E$$

or $\rightarrow$ $\boxed{(\bar{x} + E) - (\bar{x} - E)}$

الحد الصغير − الحد الكبير

$\bar{x}$ − الحد الكبير

[4] Sample size [n]

$$\boxed{n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{E}\right)^2}$$

تقرب لأقرب
عدد صحيح

باقي حل الشيت

[2] C.I for the Pop. Propotion (P)

[1] $(1-\alpha)$ = Confidence level

$Z_{\alpha/2}$

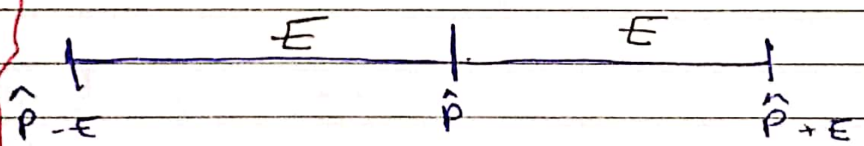
[2] C.I for P :→ $\hat{P} \pm E$

$$\boxed{\hat{P} \pm Z_{\alpha/2} * \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}}$$

observe :→

$$\hat{P} = \frac{X}{n} = \bar{X}$$

$$E = Z_{\alpha/2} * \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$



$$\hat{P}+E-\hat{P}=E$$

[3] length = $2E$

الحد الكبير - الحد الصغير

[4] $n = \left(\frac{Z_{\alpha/2}}{E}\right) * \hat{P}(1-\hat{P})$

Testing hypothesis

[1] For the PoP. mean (μ)

[1] Null-hyp $\longrightarrow$ $H_0 : \mu = \mu_0$

[2] Alterantive-hyp $\longrightarrow$ $H_a = H_1 : \mu > \mu_0$ (Right tailed)

$\longrightarrow$ $H_1 : \mu < \mu_0$ (left tailed test)

$\longrightarrow$ $H_1 : \mu \neq \mu_0$ (two tailed test)

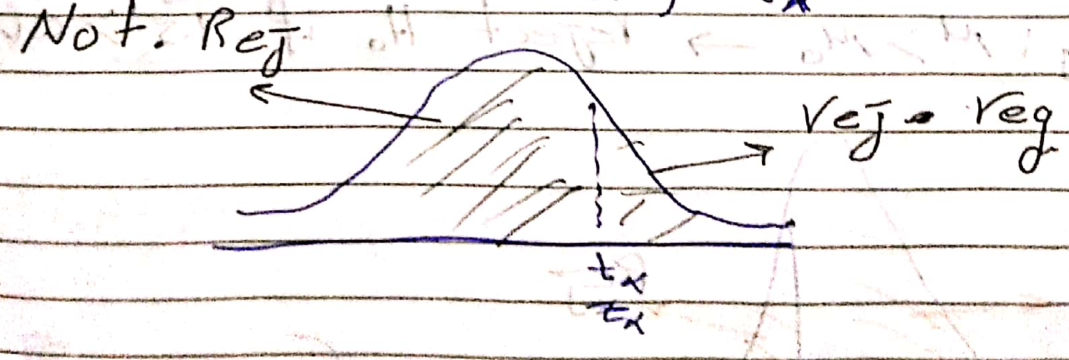
Test statistics $\longrightarrow$ (T.s) σ known $\longrightarrow$ $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{2}}$

σ unknown $\longrightarrow$ $t = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$

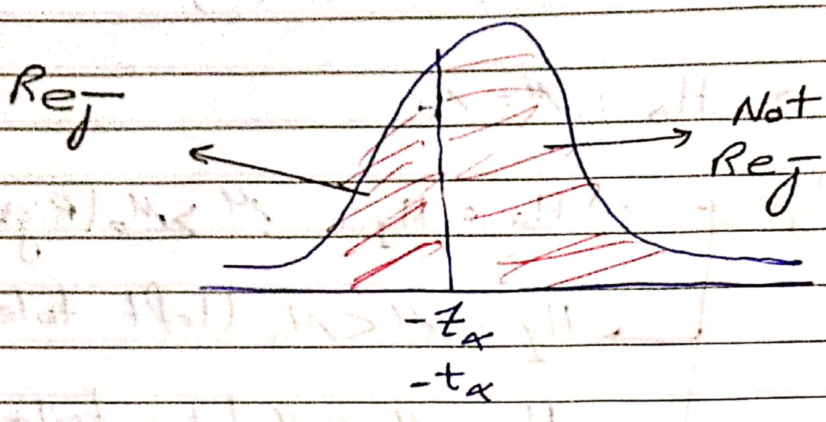
Critical Value $\longrightarrow$ (C.V) : $\longrightarrow$

[1] $H_1 : \mu > \mu_0$

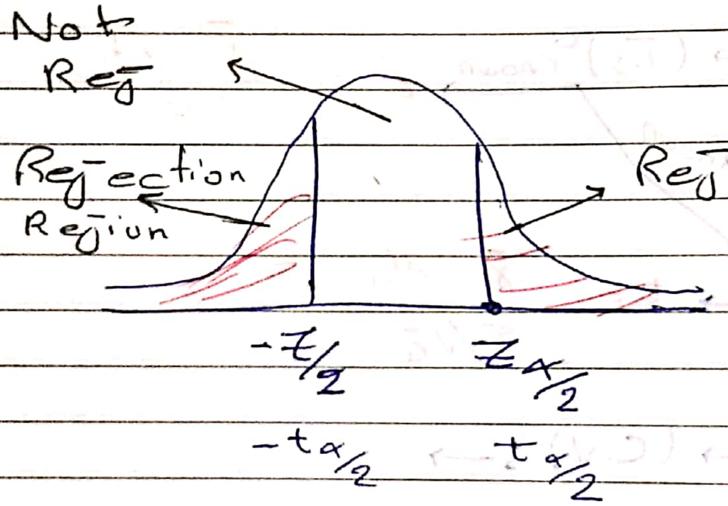
C.V : z_α , t_α .



[2] $H_1 : \mu < \mu_0 \longrightarrow$ C.V : $-Z_\alpha$, $-t_\alpha$

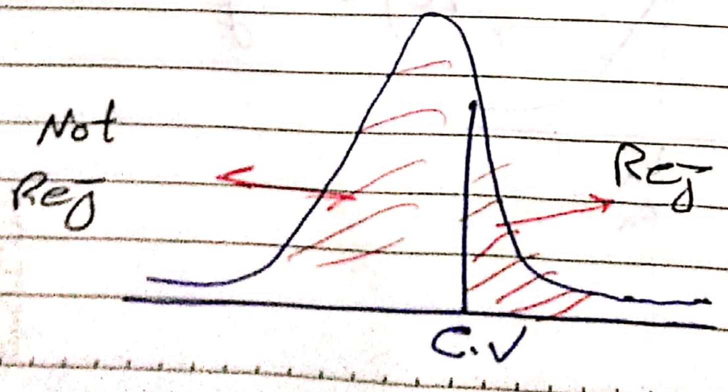


[3] $H_1 : \mu \neq \mu_0 \longrightarrow \pm Z_{\alpha/2}$, $\pm t_{\alpha/2}$

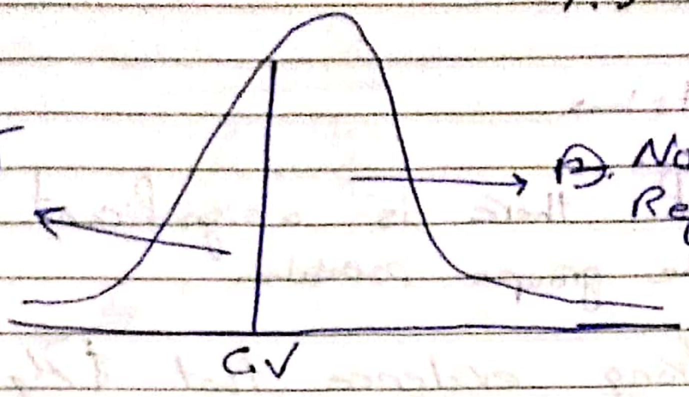


Decision :$\longrightarrow$

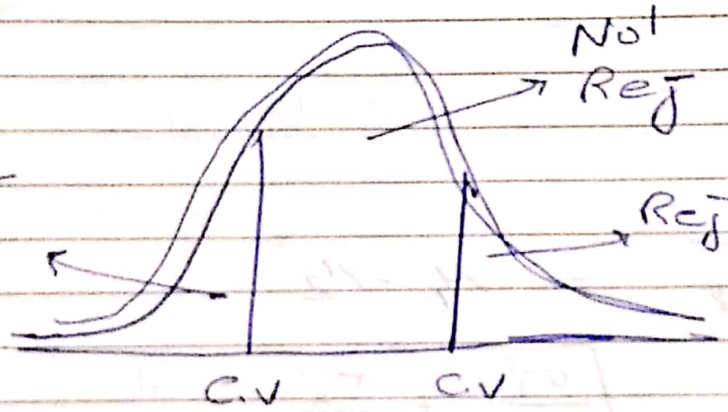
[1] $H_1 : \mu > \mu_0 \rightarrow$ Reject H_0 if T.S > C.V



[2] $H_1 : \mu < \mu_0 :\longrightarrow$ Rej H_0 if
$T.S < C.V$



[3] $H_1 : \mu \neq \mu_0 :\longrightarrow$ Rej H_0 if $T.S > Z_{\alpha/2}, t_{\alpha/2}$
(or) $T.S < Z_{\alpha/2}, t_{\alpha/2}$



* Two Types of Errors :→

[1] α (Type I error) = α → significance level
$(1-\alpha)$ → confidence level

[2] Type II error :
$(1-\alpha)$

Find Yourself to Be what you want

Inference for Two population Mean (μ)

Common Points :→

+ Testing if there is a significant difference between the groups population

(Is there strong evidence that $\mu_1 \neq \mu_2$)

+ Estimating $\mu_1 - \mu_2$ with a confidence interval
$\bar{X}_1 - \bar{X}_2$

Observe :→

[1] $\bar{X}_1 - \bar{X}_2 = \mu_1 - \mu_2$

[2] $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

[3] $\bar{X}_1 - \bar{X}_2$ is normal Distrubution

[4] $(1-\alpha)$ confidence Interval for $\mu_1 - \mu_2$ is

$$\bar{X}_1 - \bar{X}_2 \pm Z_{\alpha/2} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

standard divation

Margin of Error

If we very ofen want to test $H_0: \mu_1 = \mu_2$

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

+ if H_0 is true, this z test statistic has the standard normal distribution

Problems :→

[1] σ_1 and σ_2 are almost never Known

[2] Replacing σ_1 and σ_2 with s_1 and s_2 does not result in a statistic with a t distribution

Two options :→

[1] The Pooled-Variance t Procedure

[2] The welch (unPooled) t Procedure

other options

+ Mann-whitney U

+ Bootstrap methods and permutation tests

Linear Regression

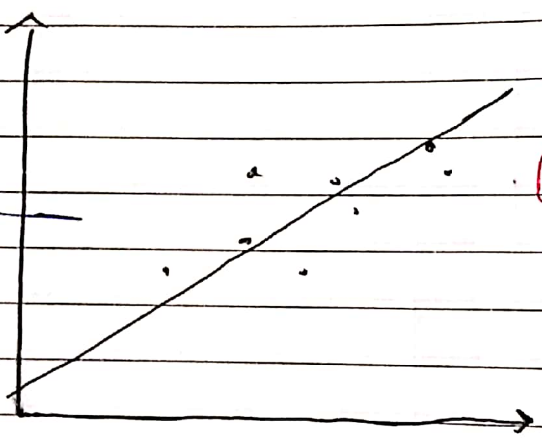
المعادلة العامة ⟵

$y = mx + b$

slope (of m)

$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

n data pairs



weak relationship

least - squares line

$y = \hat{a}x + \hat{b}$

slope (of $\hat{a}$)

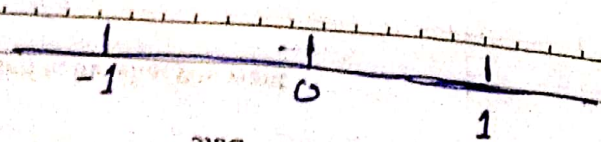
y- intercept (of $\hat{b}$)

observe :⟶

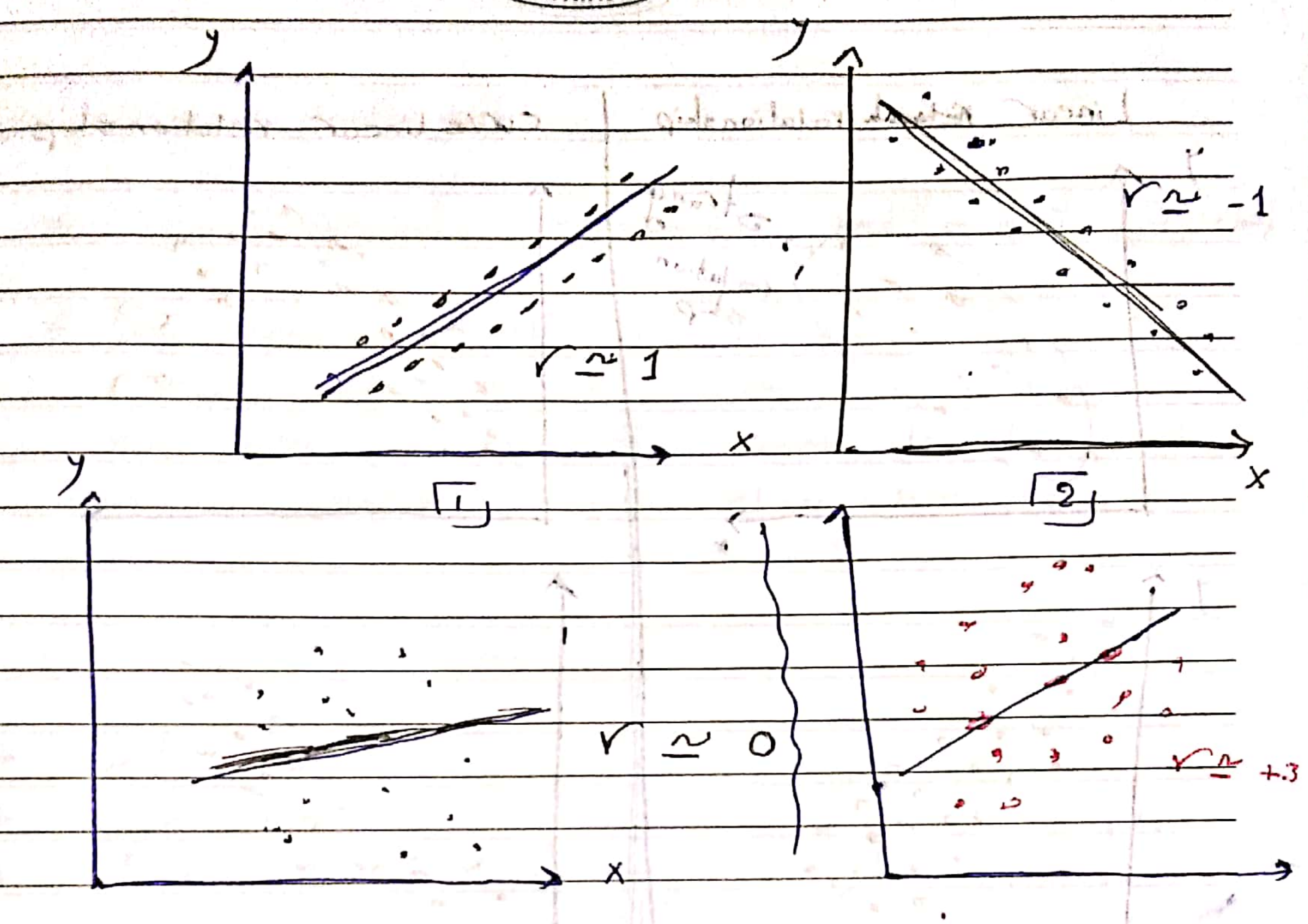
x is independent Value
y is dependent Value

Correlation

Cofficient of correlation :⟶ $-1 \leq r \leq 1$



Examples of Approximate r Values

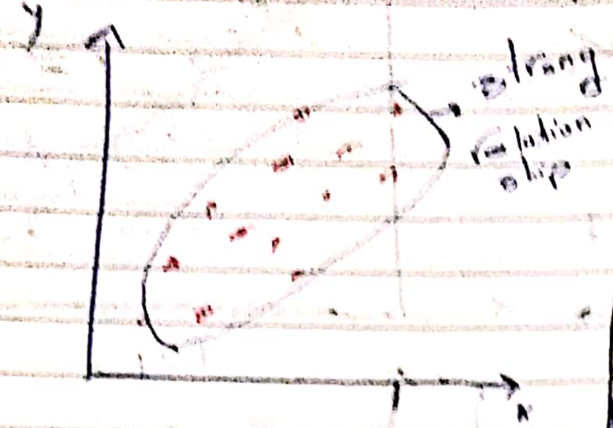


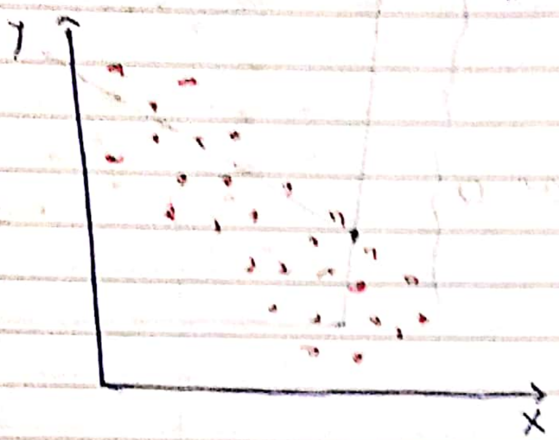
observe ↦

Correlation ↦ analysis is used to measure strength of the association (linear ~~slope~~ relationship) between $\boxed{\text{Two Variables}}$

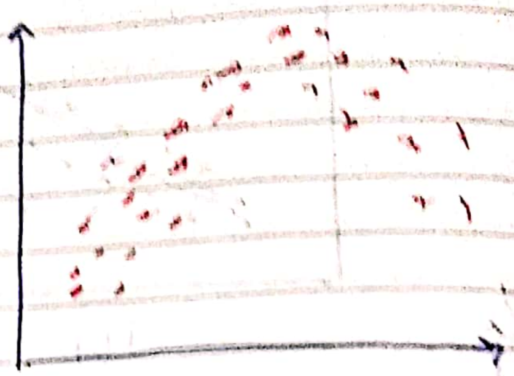
scatter plot Examples

Linear relationship

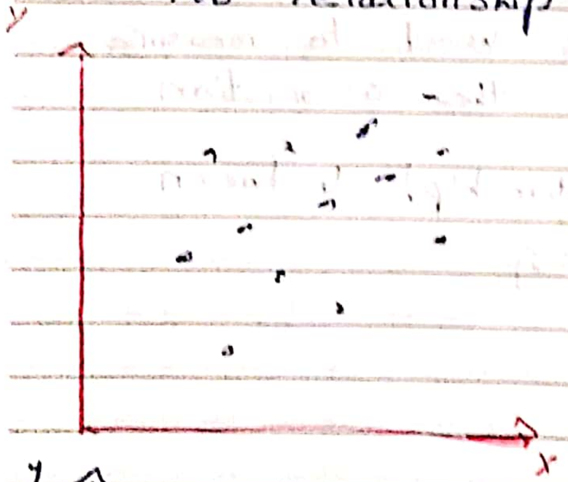


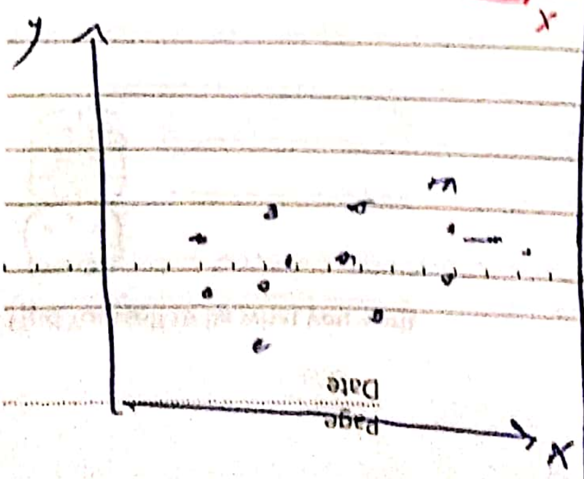


Curve linear relationship



No relationship





Correlation Cofficient :→

- The population correlation coeffieient ρ (rho)
- The sample correlation coeffieient r is an estimate of ρ is used to measure the strength of the linear relationship in the ~~example~~ sample observation

propritics of ρ and r

- unit free
- range between -1 and 1
- The closer to -1, The stronger negative linear relation.
- " " " 1, " " positive " "
- " " " 0, The weaker The linear relationship

Calculating the correlation coefficient

$\boxed{1}$ sample correlation coefficient

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\left[\sum (x-\bar{x})^2\right]\left[\sum (y-\bar{y})^2\right]}}$$

or The algebraic equivalent

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left[n(\sum x^2) - (\sum x)^2\right]\left[n(\sum y^2) - (\sum y)^2\right]}}$$

$r \longrightarrow$ sample correlation cofficient
$n \longrightarrow$ sample size
$x \longrightarrow$ independent Variable
$y \longrightarrow$ dependent Variable

Significance Test for correlation

- **Hypotheses**

$H_0 : \rho = 0$ (no correlation)

$H_1 : \rho \neq 0$ (correlation exist)

$\rho > 0$ Positive relationship
$\rho < 0$ negative relationship

- **Test statistic**

$$t = \frac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$$

example screenshots

Regression Analysis :→

Simple linear Regression Model

- only one independent variable X
- Relationship between X, y is discribed by a linear function → $y = \hat{a}X + \hat{b}$
  or $y = a + Xb$
- Changes in y are assumed to be caused by changes in X

The Population regression model

exist — linear component — Random error component

$$y = \underbrace{\beta_0 + \beta_1 X}_{\text{linear Component}} + \boxed{E} \rightarrow \text{Random Error term, or residual}$$

y → independent

β_0 → y intercept

β_1 → slope

X → independent

Estemation or Predicated Regeression Model

$$\boxed{\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_0 \hat{X}} \longrightarrow \text{Estimated}$$

$$\boxed{\hat{y}_i = b_0 + b_1 X}$$

if observe :→

(if) b_1 = zero

∴ $y = \boxed{b_0} \longrightarrow$ Constant

not exist linear relationship

if $b_1 = -1 \longrightarrow$ ~~positive~~ negetive relationship

if $b_1 = +1 \longrightarrow$ positive relationship

The Least squares Equation

The formulas for b_1 and b_0 are :→

[1]

$$b_1 = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2}$$

algebraic equivalent :→

$$b_1 = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

[2]

$$b_0 = \bar{y} - b_1 \bar{x}$$

interpretation of the slope and the Intercept

- b_0 is estimated average value of y when the value of x is zero (x, [illegible])

- b_1 is the estimated change in the average value of y as default of a one-unit change in x