Section 3.5 & 12.2

Section 3.5: Measure of Association between two variables

Measure of Association →
- Covariance
- Correlation coefficient
- Regression

1 Covariance:

for sample : $S_{xy} = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$

for Population : $\sigma_{xy} = \dfrac{\sum (x_i - \mu_x)(y_i - \mu_y)}{N}$

* Given two variables x, y

x	x_1	x_2	- - - - - -	x_n
y	y_1	y_2		y_n

→ Covariance measures the type of the linear relationship between x & y.

Covariance > 0 : positive linear relation
Covariance < 0 : Negative linear relation
Covariance ≈ 0 : No relation

2 Correlation Coefficient

for Sample : $r_{xy} = \frac{S_{xy}}{S_x \, S_y}$

S_{xy}: Sample covariance
S_x : Sample std of x
S_y : Sample std of y

for population : $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \, \sigma_y}$

/rho /segma

σ_{xy} : pop. covariance
σ_x : pop std of x
σ_y : pop std of y

Notes. (1) $-1 \leq r_{xy} \leq 1$

(2) r_{xy} measures the type and strength of the relation between x and y.

$r_{xy} = 1$ perfect positive linear relationship

$r_{xy} = -1$ perfect negative linear relationship

$r_{xy} = 0 \longrightarrow$ No relation

$r_{xy} \approx 0 \longrightarrow$ weak relation

How to find r_{xy} and S_{xy}

* We will use the $\boxed{\text{R.E.G MODE}}$ to find r_{xy}
* then use the formula : $S_{xy} = r_{xy} S_x S_y$ to find S_{xy}

Exp: The following table summarizes the # of absences (X) and the grade (y) of sample of students in B.Z.U

X_i	6	8	3	1	4	5
y_i	65	40	81	94	76	69

Find
(1) The sample Mean of X : $\bar{X}$
(2) The sample Mean of y : $\bar{y}$
(3) The sample St.d of X : S_x
(4) The sample St.d of y : S_y
(5) The sample Variance of x : S_x^2
(6) The sample ~~Variation~~ Variance of y : S_y^2
(7) The sample Correlation Coefficient r_{xy} (Interpret)
(8) The sample Covariance S_{xy} (Interpret)

Using calculator

Step ①: $\boxed{\text{Shift}}$ $\boxed{\text{Mode}}$ $\boxed{3}$ $\boxed{=}$ $\boxed{=}$

Step ② $\boxed{\text{Mode}}$ $\underset{}{\overset{\text{REG}}{\boxed{3}}}$ $\overset{\text{Lin}}{\boxed{1}}$

Step ③ Entry values (x_i, y_i)

6 $\boxed{,}$ 65 $\boxed{M+}$

8 $\boxed{,}$ 40 $\boxed{M+}$

3 $\boxed{,}$ 81 $\boxed{M+}$

1 $\boxed{,}$ 94 $\boxed{M+}$

4 $\boxed{,}$ 76 $\boxed{M+}$

5 $\boxed{,}$ 69 $\boxed{M+}$

Step ④ Statistics:

$\boxed{\text{Shift}}$ $\boxed{2}$ $\boxed{1}$ $\boxed{=}$ $\longrightarrow$ $\boxed{\bar{X} = 4.5}$

$\boxed{\text{Shift}}$ $\boxed{2}$ $\boxed{3}$ $\boxed{=}$ $\longrightarrow$ $\boxed{S_x = 2.43}$

$\boxed{\text{Ans}}$ $\boxed{x^2}$ $\boxed{=}$ $\longrightarrow$ $\boxed{S_x^2 = 5.9}$

for y variable

$\boxed{\text{Shift}}$ $\boxed{2}$ $\boxed{\rightarrow}$ $\boxed{1}$ $\boxed{=}$ $\longrightarrow$ $\boxed{\bar{y} = 70.83}$

$\boxed{\text{Shift}}$ $\boxed{2}$ $\boxed{\rightarrow}$ $\boxed{3}$ $\boxed{=}$ $\longrightarrow$ $\boxed{S_y = 18.19}$

$\boxed{\text{Ans}}$ $\boxed{x^2}$ $\boxed{=}$ $\longrightarrow$ $\boxed{S_y^2 = 330.97}$

To find the correlation r_{xy}

$\boxed{\text{Shift}}$ $\boxed{2}$ $\boxed{\rightarrow}$ $\boxed{\rightarrow}$ $\boxed{3}$ $\boxed{=}$ $\longrightarrow$ $\boxed{r_{xy} = -0.98}$

Explination : There is a Strong negative linear relationship between x & y

To find the covariance S_{xy}

$$S_{xy} = r_{xy} \cdot S_x S_y.$$
$$= (-0.98) \cdot (2.43)(18.19) = -43.32$$

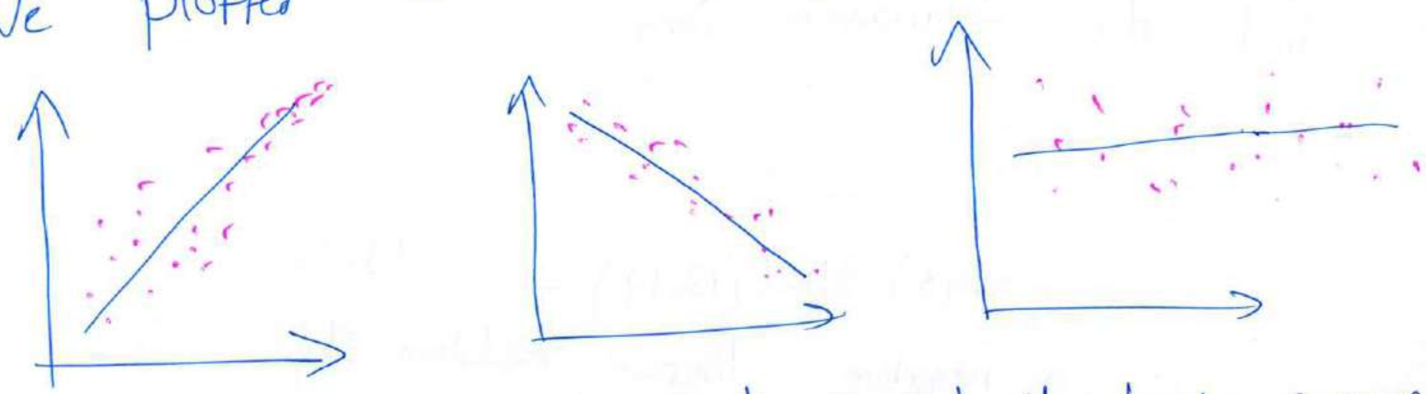
There is a negative linear relationship.

Section 12.2 linear regression

Recall:- Given two variable X and y

X	x_1	x_2	- - - - -	x_n
y	y_1	y_2	- - - - -	y_n

We plotted a scatter diagram & trendline



We can use a method called the least-square Method to find the equation of the best trendline which is called Estimated linear regression equation

$$\boxed{\hat{y} = A + \underset{\text{slope}}{B}\, x} \qquad \text{or } (\hat{y} = b_0 + b_1 x)$$

where A, B are constant (from calculator)

used to estimate (predict)

$\hat{y} = A + BX$ (calculator)

$\hat{y} = b_0 + b_1 X$ (Least squares Method)

$$B = \text{Slope} = b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$A = y\text{-intercept} = b_0 = \bar{y} - b_1 \bar{x}$$

ملاحظة:- لا يتم استخدام القوانين

← نستخدم REG. Mode Calculator

Exp: Given the following sample of x and y

x	10	18	15	20	4
y	171	200	180	240	80

(1) Find the estimated linear regression equation

(2) Estimate y when $x = 19$

(3) Find the correlation coefficient and Interpret your answer

(4) Find the covariance

Solution:

Step ① $\boxed{\text{Shift}}$ $\boxed{\text{Mode}}$ $\boxed{3}$ $\boxed{=}$ $\boxed{=}$

Step ② $\boxed{\text{Mode}}$ $\boxed{3}$ $\boxed{1}$

Step ③ 10 $\boxed{,}$ 171 $\boxed{M+}$

⋮

$\boxed{4}$ $\boxed{9}$ $\boxed{80}$ $\boxed{M+}$

(1) A : $\boxed{\text{Shift}}$ $\boxed{2}$ $\boxed{\rightarrow}$ $\boxed{\rightarrow}$ $\boxed{1}$ $\boxed{=}$ $\Rightarrow$ $\boxed{A = 57.30}$

B: $\boxed{\text{Shift}}$ $\boxed{2}$ $\boxed{\rightarrow}$ $\boxed{\rightarrow}$ $\boxed{2}$ $\boxed{=}$ $\Rightarrow$ $\boxed{B = 8.72}$

$$\boxed{\hat{Y} = 57.3 + 8.72X}$$

(2) $\hat{y}(19) = 57.3 + 8.72(19) = 222.98$

(3) $r_{xy} = ??$ $\boxed{\text{Shift}}$ $\boxed{2}$ $\boxed{\rightarrow}$ $\boxed{\rightarrow}$ $\boxed{3}$ $\boxed{=}$ $\Rightarrow$ $\boxed{r_{xy} = 0.96}$

Interpret your answer: Strong ~~the~~ positive linear relationship

(4) Covariance $S_{xy} = ??$

– First $\boxed{S_x = 6.47}$ $\boxed{S_y = 58.98}$ $\boxed{r_{xy} = 0.96}$

$S_{xy} = r_{xy} S_x S_y = 366.34$ positive linear relationship