Intercept:

$$Y = a + bx + \varepsilon$$

(i) Define Regression?

The average hight of children tends to step back or to regress toward the average height of all men. This tendency toward the average height of all men was called a regression.

(ii) Define dependent variable?

The word regression is used in a quite different sense. It investigates the dependence of one variable, conventionally called the dependent variable.

The word regression is used in a quite different Sense. It investigates the dependence of one variable, Conventionally called the dependent variable.

(iii) Define independent Variable?

On one or more other variables, called independent variables, and Provides an equation to be used for estimating or Predicting the average value of the dependent variable from the known values of the independent variable.

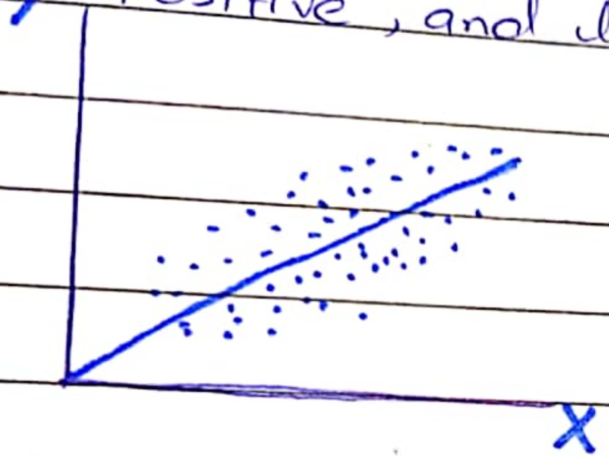
(iv) Regression Relation:

The relation between the expected value of the dependent variable and the independent variable is called regression relation.

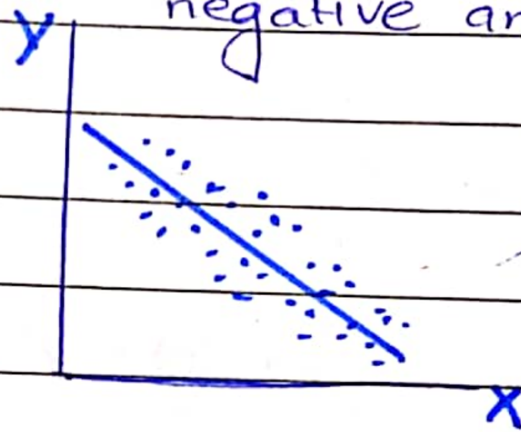
(v) Simple - variable:

When, we study the dependence of a variable on a single independent variable, it is called a simple or two - variable regression.

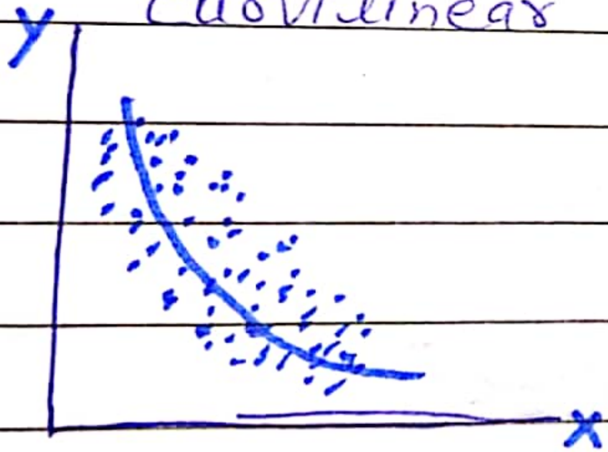
Positive, and linear



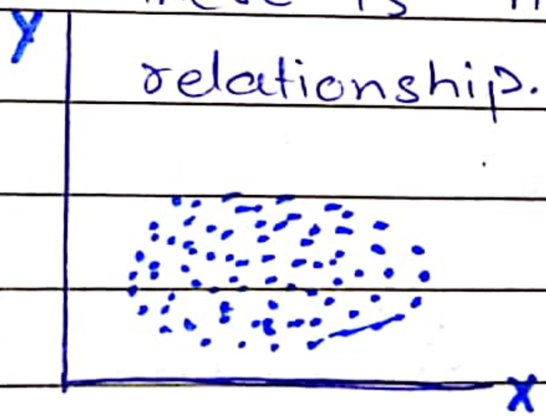
negative and linear



Curvilinear



there is no relationship.



Simple linear Regression Model:

regression.

able

dependent

l

: Simple Linear Regression Model:

Simple linear regression is a regression model that estimates the relationship between one independent variable and One dependent variable using a Straight line. both variable should be quantitative.

Formula:

Population slope coefficient

random Error term

$$Y = \underbrace{\beta_0 + \beta_1 X_i}_{\text{Linear Component}} + \varepsilon_i$$

dependent variable

Population Y intercept

independent variable

Linear Component

Mon Tue Wed Thu Fri Sat Date: __/__/20__

:Define Errors:

Errors are defined as the differences between observed values and the corresponding values predicted or estimated by the fitted model equation.

·Formula: :Intercept:

ε = Error = $y - \hat{y}$ fitted model

where

$$\hat{y} = a + bx$$

Mon Tue Wed Thu Fri Sat

Date: __/__/20

$$\frac{34677 - 30804}{11772 - 10404} = \frac{3873}{1368} = 2.831$$

$$a = \bar{y} - b\bar{x}$$

$$a = \frac{\Sigma xi}{n} - b\left(\frac{\Sigma xi}{n}\right)$$

$$\frac{302}{9} - (2.831)\left(\frac{102}{9}\right)$$

$$a = 33.5 - (2.831)$$

$$\hat{Y} = a + bx$$

$$\bar{Y} = \frac{\Sigma x}{n} : \qquad \bar{X} = \frac{\Sigma x_i}{n} .$$

Example: 10.2

In an experiment to measure the stiffness of a spring, the length of the spring under different loads was measured as follows.

X = Loads (lb)	3	5	6	9	10	12	15	20	22	28
Y = Length (in)	10	12	15	18	20	22	27	30	32	34

The estimated regression equation appropriate for Predicting the length, Y given the weight X, is

$$\hat{Y} = a_0 + b_{yx} X,$$

where $b_{yx} = \dfrac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{n\Sigma X^2 - (\Sigma X)^2}$

$$= \frac{(10)(3467) - (130)(220)}{(10)(2288) - (130)^2}$$

$$= \frac{6070}{5980} = \boxed{1.02}$$

$$a_0 = \bar{Y} - b_{yx}\bar{X}$$

$$= 22 - (1.02)(13)$$

$$= \boxed{8.74}$$

Y^2	XY
	30
	60
	90
	162
	200
	264
	405
	600
	704
	932
	3467

opropriate

e weight

Desired estimated regression equation is:

$$\hat{Y} = 8.74 + 1.02X$$

$$\hat{X} = a_1 + b_{xy} Y_1$$

where

$$b_{xy} = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{n\Sigma Y^2 - (\Sigma Y)^2}$$

$$= \frac{(10)(3467) - (130)(220)}{(10)(5486) - (220)^2}$$

$$= \frac{6070}{6460} = 0.94$$

$$a_1 = \bar{X} - b_{xy}\bar{Y}$$

$$= 13 - (0.94)(22)$$

$$= \boxed{-7.68}$$

Date: __/__/20

X	Y	X^2	Y^2	XY
3	10	9	100	30
5	12	25	144	60
6	15	36	225	90
9	18	81	324	162
10	20	100	400	200
12	22	144	484	264
15	27	225	729	405
20	30	400	900	600
22	32	484	1024	704
28	34	784	1156	932
Total: 130	220	2288	5486	3467

The estimated regression equation appropriate for predicting the length, Y given the weight X, is

$$\hat{Y} = a_0 + b_{yx} X,$$

$$a = 33.5 - (2.831)$$

$$\hat{Y} = a + bx$$

$$\hat{y} = 3.47 + 2.83x$$

$$\hat{y} = 438.702$$

:Find error:

$$\hat{Y} = (a + bx)$$

6	19	114	
8	23	184	64
10	28	280	100
12	36	432	144
13	41	533	169
15	44	660	225
16	45	720	256
17	50	850	289
102	302	3853	1308

$$b = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{9(3853) - (102)(302)}{9(1308) - (102)^2}$$

Date: / /20

Example: 10.1

Compute the least square regression equation of y on x for the following data. what is the regression coefficient and ....?

X	5,	6	8	10	12	13	15	16	17
Y	16,	19,	23,	28,	36,	41,	44,	45,	50

X	Y	XY	X^2
5	16	80	25
6	19	114	36
8	23	184	64
10	28	280	100
12	36	432	144
13	41	533	169
15	44	660	225
16	45	720	256
17	50	850	289

Mon Tue Wed Thu Fri Sat

34677 - 3

11772 - 1

$a = \bar{Y}$

$a = \Sigma Y$

$\frac{302}{9}$

Simple or two-variable regression.

(vi) Multiple Regression:

When the dependence of a variable on two or more than two independent variables is studies, it is called multiple regression.

(vii) Define Scatter Diagram?

Scatter Diagram:

Scatter diagram are used to study and identify the possible relationship between the changes observed in two different sets of variable.

Least Square Estimates in Simple Linear Regression:

Date: __/__/20

Let there be a set of observation

$$\{(X_i, Y_i), i = 1, 2, \ldots, n\},$$

$$Y_i = \alpha + \beta X_i + \epsilon_i,$$

or in terms of sample data as

$$Y_i = a + bX_i + e_i,$$

where a and b are the least-squares estimates of α and β, e_i commonly called residual, is the deviation of the observed Y_i from its estimate provided by $\hat{Y}_i = a + bX_i$.

Date: __/__/20

Simple

ervation

},

ta as

uares

y

7

Date: __/__/20

Find Sum of residual:

$$e_i = \Sigma(y - \hat{y}) = 0$$

Find Sum of Square of residual:

$$e_i^2 = \Sigma(y - \hat{y})^2$$

$$b_{yx} = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$$

$$b_{xy} = \frac{n\Sigma xy - (\Sigma y)(\Sigma x)}{n\Sigma y^2 - (\Sigma y)^2}.$$

$$\bar{y} = \frac{\Sigma x}{n} : \qquad \bar{x} = \frac{\Sigma x_i}{n}.$$

Mon Tue Wed Thu Fri Sat

Date: __/__/20__

Positive Correlation:

Positive Correlation describes the relationship between two variables that change together. e.g "Height" and "Weight".

Negative / inverse Correlation:

An inverse Correlation describes the relationship between two variables which change in opposing directions.

which change in opposing directions.

Properties of the Least - Squares Regression Line.

(i) The least Squares regression line always goes through the Point $(\bar{X}, \bar{Y})$ the means of data.

(ii) The sum of the deviations of the Observed values of Y_i from the least Squares regression line is always equal to zero i.e $\Sigma(Y_i - \hat{Y}) = 0$.

(iii) The sum of the Squares of the deviations of the observed value from the least-Squares regression line is a minimum

i.e $\Sigma(Y_i - \hat{Y}_i)^2 =$ minimum

10	17	170	100
17	8	136	289
19	9	171	301

Total:

92	68	952	1496

$$b = \frac{(6)(952)-(92)(68)}{6(1496)-(92)^2}$$

$$b = -1.06$$

$$a = \bar{y} - b\bar{x}$$

$$= 11.33 - (1.06)(15.33)$$

$$a = 27.57.$$

$$\hat{y} = a + bx$$

$$\hat{y} = 27.57 + (-1.06)x$$

$$27.57 - 1.06(20) = 6.37$$

$$27.57 - 1.06(11) = 15.91$$

$$27.57 - 1.06(15) = 11.6$$

$$27.57 - 1.06(10) = 16.9$$

$$27.57 - 1.06(17) = 9.5$$

$$27.57 - 1.06($$

Scanned with CamScanner

Thu Fri Sat

Date: __/__/20

(b) $\Sigma n (Y - \hat{Y})$

$5 - 6.37 = -1.37$
$15 - 15.91 = -0.91$
$14 - 11.67 = 2.33$
$17 - 16.97 = 0.03$
$8 - 9.55 = -1.55$
$9 - 7.43 = 1.57$

Total = 0.1

(b) $e_i = (Y - \hat{Y})^2$

$(-1.37)^2 = 1.8769$
$(-0.91)^2 = 0.8281$
$(2.33)^2 = 5.4289$
$(0.03)^2 = 0.0009$
$(-1.55)^2 = 2.4025$
$(1.57)^2 = 2.4649$

Total: $\boxed{13.00}$

Standard Error —.

$Syn = \sqrt{\frac{\Sigma (Y - \hat{Y})^2}{n-2}}$

$Syn = \sqrt{\frac{13.00}{4}}$

$Syn = 1.80$

that
"Weight"
):
es
ariables
tions.

(iv) The least squares regression line obtained from a random sample is the line of best fit because a & b are the unbiased estimates of the parameters α and β.

:Questions:

10.6: Given the following:

X	Y	XY	X^2
20	5	100	400
11	15	165	121
15	14	210	225
10	17	170	100
17	8	136	289
19	9	171	301

$$= 13 - (0.94)(22)$$
$$= \boxed{-7.68}$$

(ii) Define Correlation?

Correlation is a statistical tool that helps to measure and analyse the degree of relationship between two variables.

Correlation analysis deal with the association between two or more variables.

the observed y_i from its estimate
Provided by $y_i = a + bx_i$.

Formula: 01

least - Squares estimate of a, as

$$a = \bar{y} - b\bar{x}.$$

Formula: 02

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