Correlation :-

In many kinds of biological data the relationship b/n 2 variables is of ~~very~~ great importance. In such cases, the magnitude of 1 of the variables changes as the " of second " " changes. This is known as correlation

The term was given by Auguste Bravais. & it's graphical form was given by Francis Galton.

Correlation can be defined as the study of relationship b/n 2 or more variables.

Types:

- Linear
  - +ve
  - −ve.
- Non-Linear or curvilinear
  - +ve
  - −ve
- Zero correlation

Methods to study correlation

1. Graphical Method
   (i) Scatter Diagram
   (ii) Simple graph

2. Mathematical Method
   (i) Karl Pearson Correlation Coefficient
   (ii) Spearman's Rank Correlation Coefficient

1.) **Positive or Direct Correlation.**
A +ve correlation implies that for an ↑se in the value of 1 variable the ~~value~~ other variable also ↑ses or a ↓se in the value of 1 " results in a corresponding ↓se in the value of other variable.

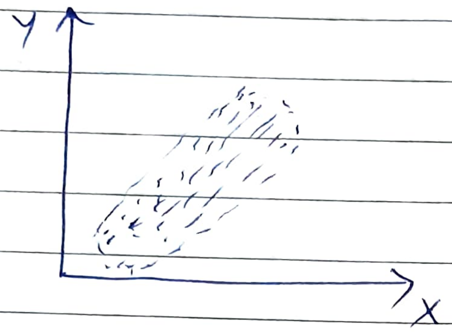
If A & B are 2 variables thus
$A \propto B$.

2.) **Negative or Inverse Correlation**
It indicates that ↑se in the value of 1 variable is associated by a ↓se in the value of other variable or vice-versa.

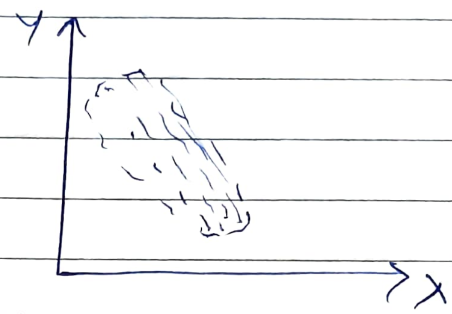
Here $A \propto \frac{1}{B}$.

3.) Zero Correlation
In this case variation in 1 variable has no relation with the variation in the other the 2 variables have no correlation & it is called as zero correlation. It denotes that there is no linear association b/w the magnitudes of the 2 variables i.e. a change in " of 1 variable does not result in the change of magnitude of other.

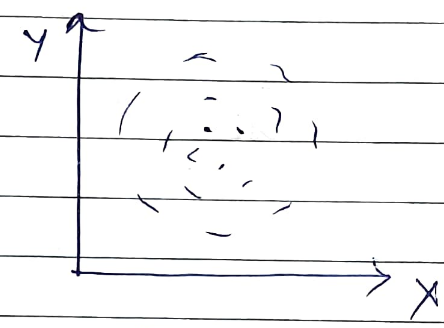
Visually correlation can be shown in the form of scatter diagram where each point represents the items.



+ve



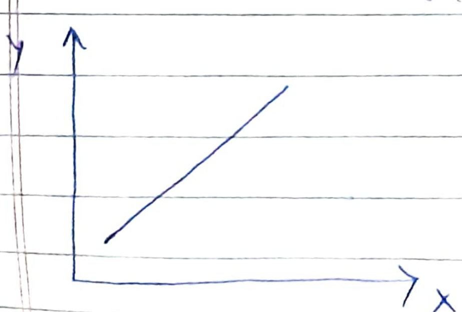
- ve



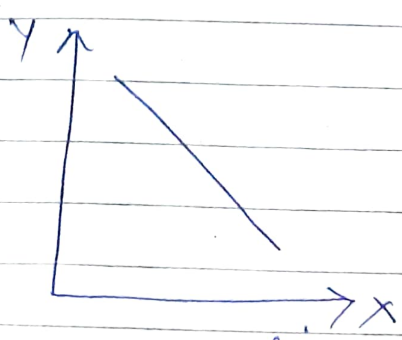
Zero.

Linear & Non Linear Correlation :-

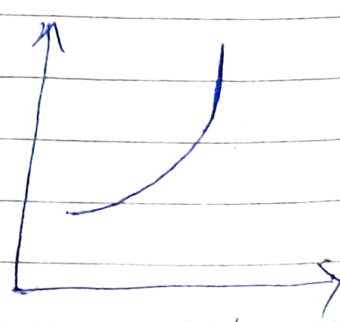
If the changes b/n 2 variables are constant it will give a linear line on the graph but when the changes in the variable values are not constant it will give a non linear or curvilinear graph.



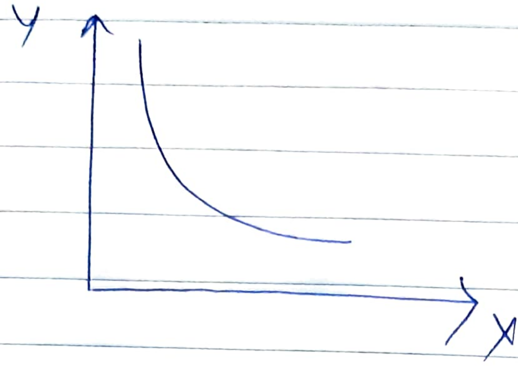
+ve linear



- ve linear



Curvilinear +ve



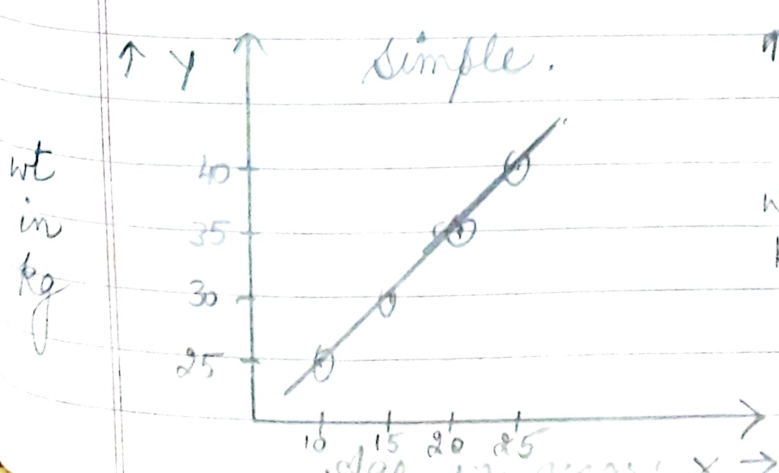
curvi linear. -ve.

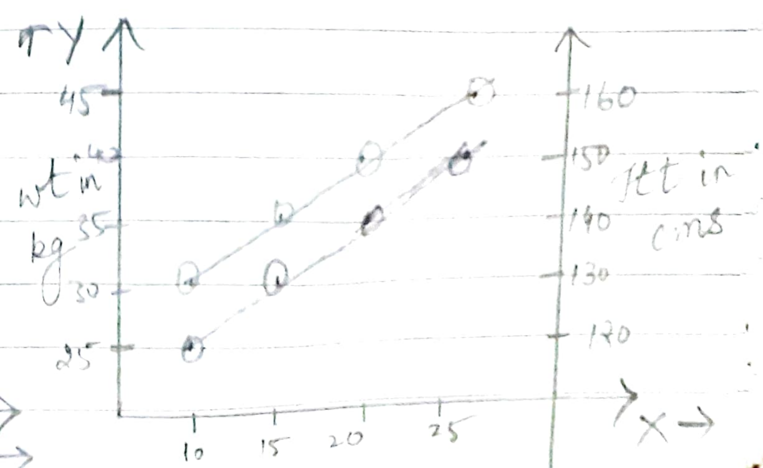
18/11/19

Simple, Multiple & Partial correlation :-

Bivariate.

Variable	Simple		Multiple			Partial		
Name of variable.	Age X	wt Y	Age X	wt Y	Ht Z	Age X	wt Y	Ht Z
Variable values.	10	25	10	25	130	20	32	145
	15	30	15	30	140	20	34	148
	20	35	20	35	150	20	35	150
	25	40.	25	40.	160.	20	38.	156.





This division is based on the no of inter dependent variables.

1. Simple
when the no of variables are 2 having 1 independent & 1 dependent variable.
It is a bivariate distribution

2. Multiple.
If we measure more than 2 variables 1 of w̄ is independent, It is a multiple correlation. It is called partial if the independent variable is constant
It is a multi variate distribut^n as it has many variables

Correlation coefficient (r)

Karl Pearson.

$$r = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma (x-\bar{x})^2 \, \Sigma (y-\bar{y})^2}}$$

The scatter diagram is a convenient way of displaying the results but it does not give the idea of degree of correlation. To measure this " coefficient of " was given by Karl Pearson

It helps to find out the linear relationship b/n 2 variables.

In the formula ; x & y are the variables
&
r = Karl Pearson's coefficient of Correlation.

Properties of Correlation Coefficient.

1. Value of r always ranges from $(-1 \text{ to } +1)$
$-1 \geq r \leq +1$

when $r = +1$ it is a perfect +ve correlation & all points on the graph lie in a straight line.

when $r = 0$ it shows no correlation b/n the 2 variables & all the pts onl the graph are scattered irregularly.

when $r = -1$ it is a perfect -ve correlatn & all the pts on the graph lie in a straight line.

2. r is a measure of closeness of fit.

3. It is also a measure of relationship b/n 2 variables in a statistical sense.

Date
Page

Degree of correlation :-

	Degree	Positive	Negative.
1.	Perfect	+1	−1
2.	High level	+0.75 to +1	−0.75 to −1
3.	Moderate "	+0.25 to +0.75	−0.25 to −0.75
4.	Low level.	+0.00 to +0.25	0 to −0.25.
5.	Zero	0	0.

Importance of correlation.

1. Correlation studies the relationship b/n 2 variables x & y. Such a relationship is called statistical relationship. There are diff methods to discover the relationship b/n 2 variables from a bivariate distribution

i) Correlation Analysis.
when Both variables in the bivariate data are quantitative then we use correlation analysis method to describe whether there is any statistical relationship b/n the 2 variables. It is done with the help of scatter diagram or line graph.

ii) Correlation coefficient.
the " analysis method just describes the relationship b/n 2 variables & doesn't give a measure of correlation b/n those 2 variables

Thus by using correlation coefficient the relationship is calculated which will express the degree of relationship.

So correlation can be defined as a tendency towards inter relation b/n the variables & the coefficient of correlation is a measure of such tendency that is it shows the degree to which the 2 variables are inter-related.

Spearmann's Rank Correlation Coefficient (R or ρ)

$$R \text{ or } \rho = 1 - \frac{6 \Sigma D^2}{n(n^2-1)}$$

The coefficient of Rank Correlation is based on various values in a data which can't be measured quantitatively but qualitative assessment is possible Eg:- Honesty, Beauty, intelligence etc.
In this case the best individual is given rank 1. and so on

D^2 = square of diff of corresponding ranks

n = No of pairs of observation.

Steps :-
1) Assign ranks.
2. Calculate $D = X_R - Y_R$
3. Calculate D^2
4. Calculate R.

Eg. Ten students got the following % of marks in Maths & Physics

Maths X : 8, 36, 98, 25, 75, 82, 92, 62, 65, 35

Physics Y : 84, 51, 91, 60, 68, 62, 86, 58, 35, 49

Find the coefficient of rank correlation.

X	Y	X_i = Rank in X	Y_i = Rank in Y	$D = X_i - Y_i$	D^2
8	84	10	3	7	49
36	51	7	8	-1	1
98	91	1	1	0	0
25	60	9	6	3	9
75	68	4	4	0	0
82	62	3	5	-2	4
92	86	2	2	0	0
62	58	6	7	-1	1
65	35	5	10	-5	25
35	49	8	9	-1	1
				$\Sigma D = 0$	$\Sigma D^2 = 90$

$$R = 1 - \frac{6 \Sigma D^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 90}{10(10^2-1)}$$

$$= 1 - \frac{6 \times 90}{10(100-1)}$$

$$= 1 - \frac{6}{11} = 1 - 0.545$$

$$= 0.455$$

Merits of Scatter diagram.

1. Easy to understand / Readily comprehensible
2. enables us to form a rough idea of the nature of relationship b/n 2 variables
3. Provides info about what kind of relationship b/n 2 variables
Helps to obtain an approx inaate estimatn line

Demerits.

1 Not suitable for larger no of observatn
2 does not provide exact measure of the extent of relationship.

Regression :-

It means stepping back towards the avg value. It is the measure of avg relationship b/n 2 or more variables in terms of units of data. (M.M. Blaire)
The term was given by Francis Galton.

Regression means estimation of unknown value of 1 variable from the known value " other "
It is an imp statistical tool & is used to study the relationship b/n 2 or more variables & estimating the unknown value.

Uses :-

1. Regression finds cause & effect relationship.
2. used for related variables like sales & purchase, volume & Pressure etc.
3. Useful in agriculture & economics.
4. It finds the best estimate of line which means an avg line is drawn that gives the best estimation.

Regression equations :-

In " there is 1 dependent & 1 independent variable w̄ are generally written as (X on Y) or (Y on X)

Algebric expression of the regression line w̄ shows the relationship b/n the independent & dependent variable in a mathematical form is called regression equation.

{X on Y ; $X = a + by$ }

{Y on X ; $Y = a + bx$ }

The linear equation when Y is a dependent variable & X is independent is shown as

$$Y = a + bx$$

where a is intercept ; b is slope or regression coefficient

The regression cofficient b for (Y on X) is shown by b_{yx}

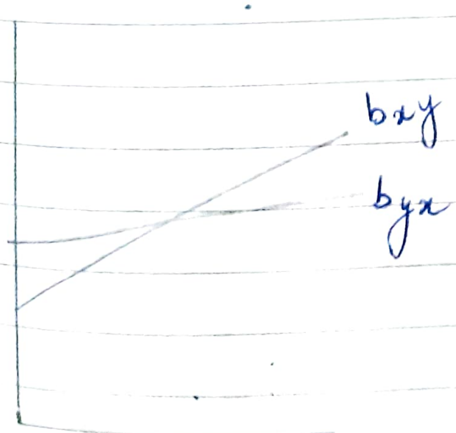
The regression cofficeent b for (X on Y) is b_{xy}

$$b_{yx} = \frac{\text{Convariance of}(X,Y)}{\sigma_x^2}$$

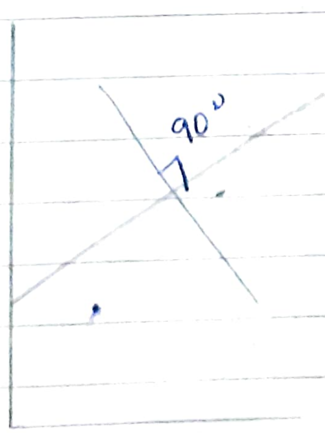
$$b_{yx} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

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

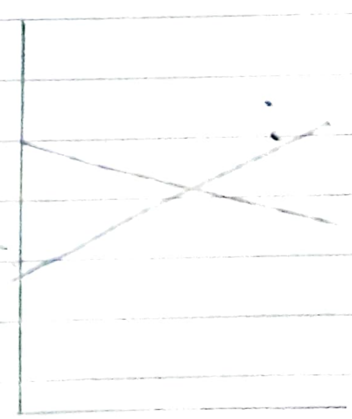
The 2 regression equations pass through the pt $\bar{x}$ & $\bar{y}$. from the regression lines we can say about correlation.



(1)



(2)



(3)

If the regression lines are very close it shows high degree of correlation.

If the 2 regression lines are ⊥ to each other then correlation is zero.

If the lines are broad it shows low degree of correlation

Properties of Regression coefficient.

1. The 2 regression coefficient b_{xy} & b_{yx} should have the same sign
2. If both are -ve then r is also -ve.
3. " " " +ve " " " " +ve

4. r is the geometric mean of ~~the~~ both the regression coefficient.

5. If 1 " " is greater than 1 the other will be always less than 1.

6. Avg of 2 regression coefficient is greater than or equal to the r

$$\frac{1}{2}\left(b_{xy} + b_{yx}\right) \geq r.$$

Regression	Correlation.
1. Regression means returning to the avg value & it expresses the avg relationship b/n 2 variables	1) It means relationship b/n 2 or more variables in such a way that changes in 1 bring about corresponding changes in other.
2. Regression analysis aims at establishing the functional relationship b/n 2 variables & then use this relationship to predict the value of dependent variable for any given value of the independent variable.	2.) Correlation coeffi. b/n 2 variables is the measure of direct^n & degree of linear relationship b/n them.
3. Regression coefficient are asymmetric i.e. $b_{xy} \neq b_{yx}$	r is symmetric & does not depend on independent or dependent variables
4. Regression analysis directly indicates cause & effect relationship b/n 2 variables	It may not show cause & effect relation ship.
5. Regression coefficients b_{xy} & b_{yx} are absolute measures representing	r is a relative measure of the linear relationship

	change in the value of 1 variable for a unit change in the value of other variable. On substituting the value of independent variable we can obtain the value of dependent variable & this value will be in the units of measurement of the variable.	b/n X & Y & is independent of units of measurement Value of r always lies b/n +1 & -1
6	There is no such thing like nonsense regression.	Their may be non sense correlat^{n} b/n 2 variables it is due to pure chance & has no practical relevance.