Paper 4

Statistics & Research Methodology

- BK SAVITRI, Pandav Bhawan, Mt. Abu
Email-bksavitrimadhuban@gmail.com
Mo. No.: 09414331060

BRAHMA KUMARIS

ANNAMALAI UNIVERSITY

Unit 1

- **Meaning of Statistics , functions , usefulness**
- **Frequency Distribution, Measures of Central Tendency-Mean , Median, Mode**

Unit 2

- Measures of Dispersion-Range, Quartile Deviation,MeanDeviation, Standard Deviation,
- Measures of Skewness-Types of Curves,Kurtosis,
- Correlation & Regression Analysis
- Correlation-Linear & Non-linear Correlation,Multiple,simple,Partial correlation
- Methods of studying Correlation-scatter Diagram,Karl Pearson Coefficient of correlation'Rank correlation
- Regression-Simple & Multiple, Linear & Non-linear regression,Multiple Linear Regression Equation
- Path Analysis

BRAHMA KUMARIS

ANNAMALAI UNIVERSITY

Unit 3 Problem Formulation & Hypothesis

- Characteristics Of Research
- Identification of Problem
- Common Errors in selecting & formulating a research Problem
- Research Design & Types-Research Methods
- The Questionnaire-Forms,Characteristics,
- Analysis & Interpretation of Questionnaire Responses-Advantages,Limitations
- Hypothesis-sources,Characteristics,
- Types,Testing of Hypotheis
- Difficulties in the formulation of Hypotheis

Unit 4 : Data collection, Analysis & Interpretation

- Sampling Theory- Basis of Sampling,Importance,Advantages/Disadvantages
- Charactristics of Good Sample
- Census Method, Sampling Method
- Techniques of Data Collection-Methods,
- Observation Technique,SurveyMethod,Documentary/Historical Method,The Experimental Method
- Observation Method-Structured & Unstructured observation,Participant & Non-participant Observation
- Survey Method-Interview Technique,Questionnaires-types,errors in use,pre-testing & checking schedules
- Documentary Method-Types: Life History,Diaries,Letters,Memories,public documents,social survey
- Experimental Method-

Unit 5 : Structure of Research Report

- Report Writing-style,content
- Diagrammatic & Graphic Representaion of Data
- TypesofDiagrams-OneDimensional/Bar,
- Two-dimentional diagrams, eg.,rectangles squares,circles
- Three dimensional diagrams,eg.,cubes,cylinders & spheres
- Pictograms & cartograms
- Pie Diagrams
- Graphs of Frequency Distribution-histograms,Frequency Polygon,Smoothed Frequency Curve,Ogives/cumulative Frequency Curves
- The Preliminary Section & Text, Context

Unit 1 Statistics

- **Origin**-Latin ---Status
  Italian—Statista
  German-Statistik = STATE
- Statist = Expert in ruling the State (Shakespear in Hamlet & Milton in Paradise regained-1st time word used in England)
- Definition-
- 'It is the science of dealing with numerical data; it encompasses all the necessary operations from the initial planning & assembling of data to the final presentation of conclusions. More specifically, it involves collecting statistical data, classifying them, analyzing,& interpreting them & drawing from them whatever conclusions are valid.'

  --- Ya-lun chou, Applied Business Economic Statistics

'Statistics is the science which deals with the methods of collecting, classifying, presenting, comparing & interpreting numerical data collected to throw some light on any sphere of inquiry

-Seligman

Importance of Statistics

- translate complex facts
- enriches experience
- lends precision
- gives knowledge on

  NI, production, consumption pattern, population, natural resources

Scope

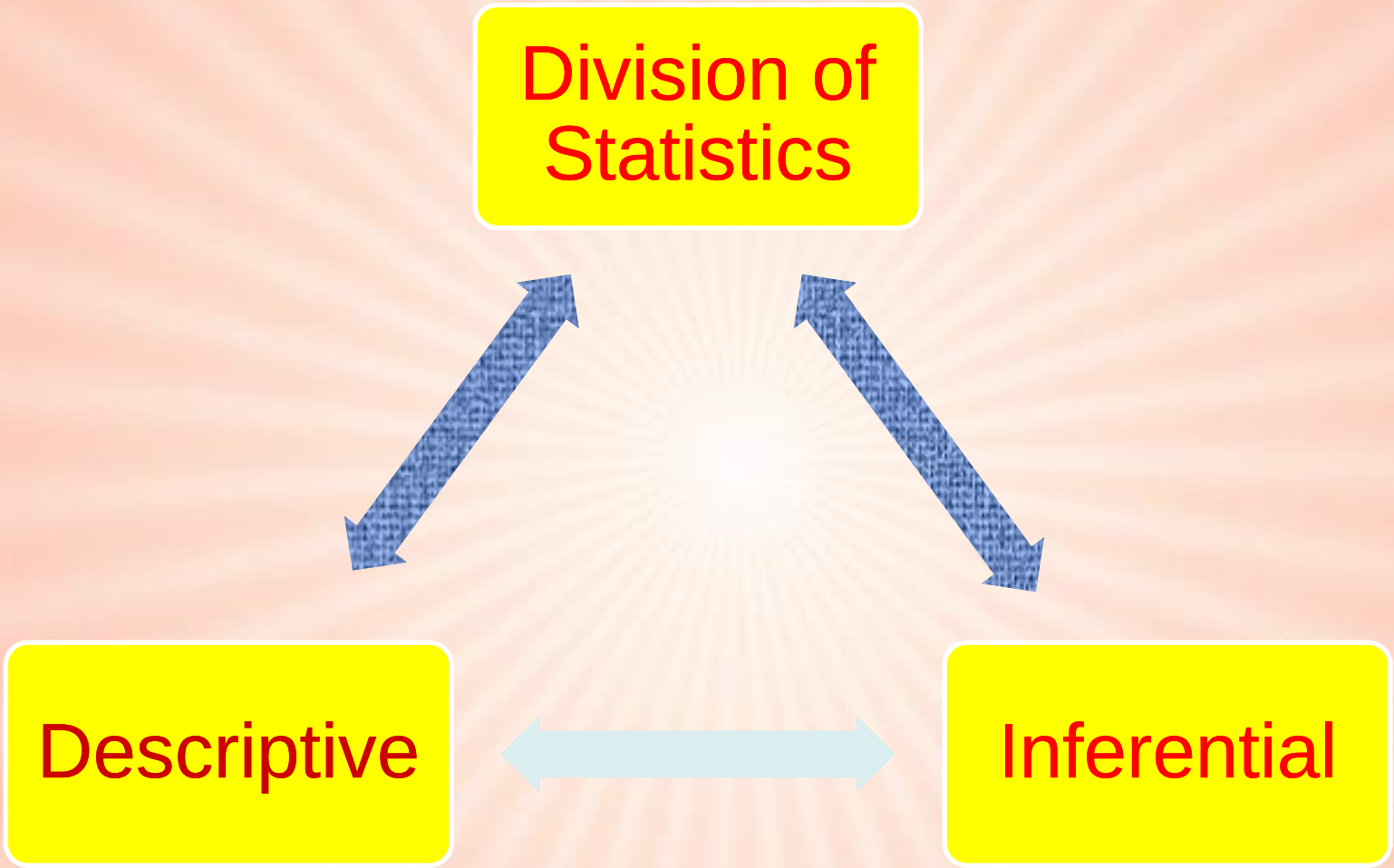
- vast, expanding
- method for collecting data
- as a mean of sound technique

  for handling & analysis & drawing valid inferences

contd……

- tools of all science for
  - research
  - intelligent judgment
  - recognized discipline
  - trade,industries,commerce
  - biology, botany, astronomy, physics, chemistry,education, medicine,geography,psycology sociology

Division of Statistics



Descriptive

- Data Processing to summarize &describe important features of the data

eg. Mean, ME, SD, Correlation, Coefficient , Frequency Distribution.

Inferential Statistics

Methods based on

totaling of observation = population

on the basis of part of totality = sample

study problems - estimation

- tests of hypothesis

BRAHMA KUMARIS

ANNAMALAI UNIVERSITY

Limitations

- Only aggregate of facts
- not related with individuals
- Not studying qualitative phenomenon
- laws are true only on an average
- liable to be misused

Therefore data must be uniform, homogeneous

Functions

- presents facts in a definite form
- condenses mass of figures
- facilitates comparison
- studies relationship
- enlarges experiences
- formulation of policies, hypothesis
- helps in forecasting
- measures uncertainty

Classification of Data into Different Data Series

- Individual series – eg., 3,4,5,6,7…
- Discrete Frequency distribution – data in full number, its values are exact , eg inclusive class interval 15-19, 20-29, gap of 1 between two class
- Continuous frequency distribution-when values are in large no., classes in continuous eg., 65-70, 70-75,…exclusive type of class

Frequency Distribution

- Grouped Frequency Distribution
- Meaning-When the range of values of variable is large-0 to 100, then data are classified into classes, then recording the no of observations into each group
- Types of classes-Inclusive,exclusive,open end class
- Cumulative Frequency Distribution

BRAHMA KUMARIS

ANNAMALAI UNIVERSITY

EXAMPLE – FREQUENCY DISTRIBUTION

एक कक्षा में छात्रों के गणित के विषय में परीक्षा के प्रतिशत अंकों को दर्शाया गया है। इस हेतु 100 छात्रों के अंकों का निदर्शन (observation) निम्न रूप से दर्शाया गया है:

81	85	62	71	70	81	86	67	96	51	63	71	75
69	48	34	87	86	73	75	42	91	58	93	52	82
90	95	82	72	53	38	77	93	85	47	70	68	57
71	96	40	70	92	68	88	58	51	90	74	52	63
96	77	83	76	48	92	81	83	92	73	84	78	
78	72	60	84	78	60	43	70	83	64	96	93	
55	73	58	40	88	96	72	53	87	92	73	77	
63	58	71	80	38	63	56	76	82	61	76	63	

FREQUENCY TABLE

सारणी (Table-1)

गणित के 100 छात्रों के प्रतिशत अंक

प्रतिशत अंक चर/Variable	मिलान रेखाएं-Tally-bar	छात्रों की संख्या आवृत्ति/Frequency
30-40	𝍸	4
41-50	𝍸	5
51-60	𝍸 𝍸 𝍸	15
61-70	𝍸 𝍸 𝍸 |	16
71-80	𝍸 𝍸 𝍸 𝍸 𝍸	24
81-90	𝍸 𝍸 𝍸 𝍸 |	21
91-100	𝍸 𝍸 𝍸	15
कुल		100

EXAMPLE – LESS THAN & MORE THAN CUMULATIVE FREQUENCY

Width of the $\frac{73.5 - 45.1}{6} = \frac{28.4}{6} = 4.7 = 5$

∴Class-intervals

∴Let us consider the class intervals as 45-50, 50-55 etc. The frequency table will be as follows:

Weight (in kg.)	Tally Marks	No. of workers (frequency)	Percentage frequency)	Relative (frequency)
45-50	|||| |	6	15.0 %	0.150
50-55	|||| ||	7	17.5 %	0.175
55-60	|||| ||||	10	25.0 %	0.250
60-65	|||| |||| ||	12	30.0 %	0.300
65-70	|||	3	7.5 %	0.075
70-75	||	2	5.0 %	0.050
	Total	40	100.0%	1.000

Table - 3

TABLE – MORE THAN & LESS THAN CUMULATIVE FREQUENCY

Weight class (in kg.)	No. of workers (frequency)	Weight (in kg.)	No. of workers whose weight is less than	No. of workers with weight greater than or equal to x (Cumulative frequency of greater than or equal to type)
		45	0	6 + 34 = 40
45-50	6	50	0 + 6 = 6	7 + 27 = 34
50-55	7	55	6 + 7 = 13	10 + 17 = 27
55-60	10	60	13 + 10 = 23	12 + 5 = 17
60-65	12	65	23 + 12 = 35	3 + 2 = 5
65-70	3	70	35 + 3 = 38	2 + 0 = 2
70-75	2	75	38 + 2 = 40	0

Table – 4

Here, the number of workers with weight less than a given value x is called cumulative frequency of less than type for weight (x). Thus 0, 6, 13, 23, 35, 38, and 40 are less than type cumulative frequencies for the weight (in kg) 45, 50, 55, 60, 65, 70 and 75 respectively.

Various Measures of Central Tendency

- Arithmetic Mean-Methods of Computation
- Geometric Mean
- Harmonic Mean
- Weighted Arithmetic Mean
- Median
- Mode
- Comparison of Mean, Median & Mode
- Frequency Distribution Curve

CALCULATION – ARITHMETIC MEAN

Here, to find out the average income of all families, first add incomes of all families together, and then divide it by no. of families.

$$\text{A.M.} = \frac{1600 + 1500 + 1400 + 1525 + 1625 + 1630}{6}$$

$$= 1{,}547 \text{ (Rs.)}$$

So, on an average one family rerise Rs. 1547 - 00 monthly income.

CALCULATION – GEOMETRIC MEAN

Example : Suppose three figures are given as viz. - 3, 25 and 45, find out its G.M., here N = 3

$$\text{Geometric Mean} = \sqrt[3]{3 \times 25 \times 45}$$

$$= \sqrt{3375}$$

CALCULATION – HARMONIC MEAN

Example : Suppose if one car goes at 60 km. for every hour, and return back at the speed of 40 km. for every hour, then to find out its average speed, although its distance is of equal, here we can use the H.M. to find out its average speed.

As per above formula - N is number of observations.

In above example the speed of car is 60 km. & 40 km. Therefore, its average speed would be :

$$\frac{2}{\left(\frac{1}{60}+\frac{1}{40}\right)} = \frac{2}{\frac{5}{120}} = \frac{2 \times 120}{5} = 48 \text{ km. on an average}$$

CALCULATION – WEIGHED ARITHMETIC MEAN

उदाहरण : तीन त्रों के A,B,C निम्न रूप में 4 विषयों में Mark दिए गए हैं। विषय के ार (Weights) ी बताए गए हैं। निर्णय कीिए कि इन तीन त्रों में से कौन-सा त्र सर्वोत्तम है।

Table - 2

Subject	Marks			Weight	WA	W B	WC
	A	B	C	W			
P	28	35	30	4	112	140	120
Q	30	25	35	3	90	75	105
R	40	20	30	2	80	40	60
S	20	15	20	1	20	15	20
Total				**10**	**302**	**270**	**305**

$$\text{Weighted Average of A} = \quad WA = \frac{302}{10} = 30.2$$

$$\text{Weighted Average of B} = \frac{\Sigma WB}{\Sigma W} = \frac{270}{10} = 27$$

$$\text{Weighted Average of C} = \frac{\Sigma WC}{\Sigma W} = \frac{305}{10} = 30.5$$

C सर्वोत्तम त्र है।

CALCULATION – MEDIAN

other hand when N is even, the mean of the $\frac{N}{2}^{th}$ and $\left[\frac{N+1}{2}+1\right]^{th}$ is the values (that is $\frac{N+1}{2}$ the value) will be taken as median of the variable.

Calculation of Median of the Raw Data:

The formula to find the median of the raw data is given by :

Me = The value of $\frac{N+1}{2}^{th}$ the item in a series of ascending order.

Example: Suppose in given set of data there are following observations.

5, 7, 6, 1, 8, 10, 12, 4, 3

Arranging them in an ascending order :

1, 3, 4, 5, 6, 7, 8, 10, 12

Here middle value is 6. Therefore Median is also 6. Here half of the values are greater than 6 and half are less than 6.

CALCULATION – MODE

Ex. : Calculate modal life.

Life in hours	No. of bulbs
1000 – 1100	40
1100 – 1200	80
1200 – 1300	100
1300 – 1400	60
1400 – 1500	60
1500 – 1600	50

Solution :

Here the maximum frequency is 100, corresponding to the class interval 1200 – 1300.

∴ The modal class is 1200 – 1300

$$Z = l_1 + \frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)} (l_2 - l_1)$$

Here, $l_1 = 1200$, $l_2 = 1300$, $f_1 = 100$, $f_0 = 80$, $f_2 = 60$

$$\therefore \quad = 1200 + \frac{100 - 80}{(100 - 80) + (100 - 60)} (1300 - 1200)$$

$$= 1200 + \frac{20}{20 + 40} (100)$$

$$= 1200 + \frac{100}{3}$$

$$= 1200 + 33.33 = 1233.33$$

Modal life is 1233.33 hours.

Unit 2

Measures of Dispersion,Skewness,Kurtosis, Correlation,Regression

- Measure of Dispersion-Absolute/Relative
- Range
- Quartile Deviation
- Mean Deviation
- Standard Deviation

BRAHMA KUMARIS

ANNAMALAI UNIVERSITY

Unit 2-Measures of Dispersion(MD)-1

Need-

Measures of Central tendency(MCT) explains only typical representative figure to the whole set of its values

In real situation,of those sets of observations whose central tendency are same but they may differ individually from each other-eg., graph shows A,B & C curves have the same Mean but have different variability from one another

Mean, Mode, Median tell us only part of the characterisics of data

Measures of Dispersion tell us more about it Spread & Variability

BRAHMA KUMARIS

ANNAMALAI UNIVERSITY

Measures of Dispersion-2

- Meaning-It is a measure of the extent to which the individual set of data are expressed in different units,
- Eg., inches of heights of students vs centimeteres of heights of another set of students
- Two types of Measure of Dispersion-
- Absolute Measure measures dispersion of one set of data
- Relative Measure measures ratio of a measure of absolute dispersion to an arithmetic mean of a particular fixed value-a coefficient, for comparing the variability of the distributions

Range

- Range-difference between the Largest & Smallest values of the Variable/set of data
- Range=L-S
- Coefficient of Range = L-S

  L+S

Quartile Deviation/Semi Inter-quartile Range(1)

Values which divide the data set into a number of equal parts are called the Quartiles

Some important partition values are Median,quartiles,deciles,percentiles

$$QD = \frac{Q3 - Q1}{2}$$

Relative Measure of QD is the Coefficient of Quartile Deviation

$$CQD = \frac{Q3-Q1}{Q3+Q1}$$

BRAHMA KUMARIS

ANNAMALAI UNIVERSITY

Mean Deviation(2)

- It is the A.M. of the numerical deviations of the individual values of the data from the Measures of Mean or Median
- MD = formula
- CQD = formula

Standard Deviation(3)

- Suggested by Karl Pearson-1893
- It is the positive square root of the arithmetic mean of the squared deviations of the measurements/observations of a set from their arithmetic mean-denoted as small sigma/called as root mean squared deviation
- The square ot the standard deviation is known as variance
- Formula
- Coefficient of Variation-CV

Measures of Skewness

- For those Distributions which differ widely in their Nature & Composition, their Shape & Size differ from one another, although they have same Mean
- Graphs-
- Normal curve
- -ve skewed curve
- +ve Skewed curve

Unit 2 Measures of Kurtosis

- Kurtosis describes charactristics of Frequency Distribution
- It refers to the degree of the Peakedness or flatness of the top of the distribution-in relation to a symmetrical distribution
- Diagram

Correlation-1

- Its usefulness & Meaning -

  - units of two variables are different,eg height & age

  -change in value of one variable affects correspondingly change in value of another variable either in the same direction (+ve)or in opposite direction(-ve), then two Variables are said to be correlated,eg.,rainfall & yield of crop are positively correlated, but price & demand are negatively correlated

If change is in same direction , in same proportion=relation is perfect positively correlated

If change is in opposite direction=the variables are said to be perfect negatively correlated

BRAHMA KUMARIS

ANNAMALAI UNIVERSITY

Correlation-2

- Definition-
- “If two or more quantities vary in sympathizing so that movements in the one tend to be accompanied by corresponding movements in the others,then they are said to be correlated”

-L R Cornor

“when the relationship is of a quantitative nature, the appropriate statistical tool for discovering & measuring the relationship and expressing it in a brief formation is known as correlation” – Croxton & Cowden

BRAHMA KUMARIS

ANNAMALAI UNIVERSITY

Correlation-3

- Linear & Non-linear Correlation
- If for a unit change in one variable , if there is a CONSTANT CHANGE in the other Variable over the complete range of Values, eg.,mathamatically can be written as Y = 5 X + 2
- X : 1 2 3 4 5
- Y : 7 12 17 22 27
- This relation when traced in the graph,it gives a straight line with slope
- Such correlation is found in physical as well as absolute sciences
- Non-linear/curvilinear relationship-when a unit change in one variable, does not bring change in another variable at a constant rate but brings a change at a Fluctuating Rate-eg in economics & social sciences we get non-linear curve (no straight line)

BRAHMA KUMARIS

ANNAMALAI UNIVERSITY

Simple, Multiple, Partial Correlation - 4

- Simple Correlation - two variables are studied. Gives an idea of Degree & Direction of the relationship
- Multiple Correlation - three or more variables are studied.Coorelation coefficient measures combined relation between a dependent and a series of independent variables-height of the son is dependent variable, & of father –mother is independent variable

Partial Correlation – More than two variables,but effect of two variables influencing each other is studied, efffect of the rest of the other variables is kept constant.eg.,effect of the height of the mother is kept constant in studies.

Contd…..

It does not tell us about the Cause & Effect relationship between the Variables

If the Variables, in fact, have cause & effect relationship it implies Correlation. But the converse is not true, i.e. even a high degree of correlation between the two variables need not imply a cause & effect relation between them. It establises only co-variation or joint variation.

The high degree of correlation between the variables may be due the following causes ;-

-BOTH THE VARIABLES MAY BE MUTUALLY INFLUENCING EACH OTHER,SO THAT NEITHER CAN BE DESIGNATED AS THE CAUSE AND THE OTHER AS THE EFFECT

-BOTH THE COORELATION VARIABLES MAY BE INFLUENCE BY ONE OR MORE OTHER VARIABLES OR EXTERNAL FACTORS

-THE COORELATION MAY BE DUE TO PURE CHANCE

Methods of studying Simple Linear Correlation - 5

- Scatter/Dot Diagram Method
- Karl Pearson”s Coefficient of Coorelation
- Spearman’s Rank Coorelation Coefficient

Scatter Diagram-1

- If height is measured in centimeters, weight in kg. of variables X & Y, then in studying the correlation between them following points should be borne in mind
- Poor or low coorelation-when greater the SCATTER of the plotted points on the graph, the lesser would be the relationship between the two variables.
- The more closely the points come to a straight line, the higher the degree of relationship

Scatter Diagram-2

- If points reveal trend=variables correlated,viceversa
- A band/strip from left – bottm to upper right side top,figure-1=positive correlation,values move in the same direction.
- If band shows downward trend-figure2=negative correlation

Scatter Diagram-3

- Figure-3, if all the points lie on straight line from left-bottom towards right top=coorelation is perfect-positive
- Figure-4-points lie on a straight line from left top and coming down to right bottom=perfect negative
- Figure-5 if points are scattered in all direction,no strip of points=absence of linear relationship
- Figure-6if no corelation=zero correlation

Karl Pearson Coefficient of Correlation

- A mathematical device to measure the intensity of /magnitude of Linear relationship(1867)
- SD indicates the amount of variability of the values from their A.M. in the data set
- Covariance between the two variables - X & Y, measures the joint variation of the values of the two variables from the A.M.in the bivariate data
- COVARIANCE between x & Y= COV(X,Y)
- FORMULA r = Cov.(X,Y)/(S.D. of X) (S.D. of Y

Properties of the Correlation of Coefficient (CC)

- CC = r- is a pure number,
  independent of the units of measurement
  hence, comparisons between the correlation can be easily made
- C is independent of the choice of origin 7 scale of observations- formula,
- R (U,V) = r (X,Y)
- CC/r lies between -1&+1
- +1=perfect positive correlation
- If r = 0.9 & 0.8 = high degree of +ve correlation
- If r = 0.1 & 0.2 = low degree of + ve correlation
- If r = 0 = no/zero correlation
- If -1 < r <0 = negative correlation
- If r = - 0.9 & - 0.8 = high degree –ve correlation
- If r = - 0.1 & - 0.2 = low degree of –ve correlation
- If r = -1 = perfect –ve correlation

Rank Correlation of Charles Spearman

- When variables (in the form of Atributes) cannot be measured in quantitative measurement, but can be arranged in Serial Order-when we deal with Qualitative characteristics,eg intelligence,honesty indicates the rank in the group
- Example for calculation
- Formula

Linear Regression Analysis

- Regression – "return to the mean value",

Predicting/estimating the relationship between the two variables, advertising expenditure & sales

It deals with the derivation of an appropriate functional relationship between two variables

It is a mathematical measure expresing the average relationship between two or more variables

In general R = the estimation of the unknown value of an variable from the known value of the other variable

BRAHMA KUMARIS

ANNAMALAI UNIVERSITY

Definitions

- “ One of the most frequently used techniques in economics & business research, to find a relation between two or more variables that are related causaly is regression analysis”- Taro Yamane
- Types - Simple,Multiple
  - Linear,Non-linear Regression

Equations of Regression Lines OF y ON x, & x on y

- The mathematical equation of the regression curve is straight line, is called the regression equation, enables us to study the average change in the value of the dependent variable for any given value of the independent variables.
- Regression Coefficients

BRAHMA KUMARIS

ANNAMALAI UNIVERSITY

Some Results of Regression Equations

- The two regression lines coincide when r =+1,perfect correlation
- It intersect each other when the correlation

Coefficient r lies in between -1 and + 1,i.e.,

$-1 < r < +1$

The point of intersection of two regression lines is at x bar, y bar in the scatter diagram

Difference between Correlation & Regression Analysis

- C=relationship between two variables which vary in sympathy to one another
- R=returning back to the average relationship between the two variables
- CC=is a measure of the direction & degree of the linear relationship, without caring for which one is dependent/independent variable
- R=studies the functional relationship, predict/estimate the value of the dependent variable for any given value of independent variable.
- The regression coefficients are not symmetric in x & y

Remaining points…

Multiple Regression & Correlation

- In practise/social sciences behaviours of each variable can be described by number of factors-eg., yield of crop depends on factors such as rainfall, fertility, temperature, fertilizer etc.,
- Multiple Linear Regression Equation

Path Analysis

- In Plant Breeding, influence of all the independent variable together and also separately on dependent variable can be examined with the help of the Multiple Linear Regression Analysis