Kunthavai Naachiyaar Govt. Arts College (W) Autonomous,
Thanjavur.

B.Sc. Degree - Geography Major

Allied Course - Applied Statistics - II

Sub Code :- 18KAGIAS2

Scanned by TapScanner

Unit - I

Association of Attributes :-

The Statistic deals with quantitative data only. Characteristics possessed by an individual item may be classified into 1. Numerical and 2. descriptive. The characteristics which are capable being measured quantitatively are termed as statistics of variables (numerical classification) for instance, height, weight, wages, length, income, expenditure etc. They are not capable of quantitative measurement are termed as statistics of attributes. The observer may find presence or absence or some attributes. There are certain phenomena like blindness, dumbness, deafness, literacy, sickness etc. Which cannot be measured directly. In such cases the presence or absence of the attributes may be studied, for example, when we count the number of units produced by a factory. Thus stations of variables is leared on numerical characters and the statistics of attributes is leared on the descriptive characters.

Classification :-

An observation of the population into two classes on the basis of an attribute, say literacy can be studied, when the whole population is applied to one attribute; the two - classes of attribute is present and other in which the attribute is absent. For instance, when we study the literacy of a village, then the population of the village is divided into two classes - one class of people who are literate and the other class who are not literate.

Scanned by TapScanner

If two attribute are studied their combination can be represented by the combination of the letters representing the two attributes. Thus, If blindness is represented by A and deafness by, B then AB would represent blindness and deafness; AB would represent blindness and absence of deafness. aB would represent absence of blindness and presence of deafness; and ab would represent absence of blindness and absence of deafness.

Correlation and Association :-

Correlation and association can be measured between any two sets of related phenomena, which are capable of being directly measured. But associated of attributes is measured by the degree of relationship of two phenomena, where they are not directly measurable, but studied by the presence or absence of particular attribute. In statistics two attributes are regarded as associated only if they appear together in greater number of cases than are expected if they are independent.

Significance of Study of Correlation :-

* Correlation is very useful to economics to study the relationship between variables, like price and quantity demanded. TO the businessmen, it helps to estimate cost, sales, price and other related variables.

Scanned by TapScanner

* The relation between variables can be verified and tested for significance, with the help of the correlation analysis. The effect of correlation is to reduce the range of uncertainty of our prediction.

* The coefficient of correlation is relative measure and we can compare the relationship between variables which are expressed in different units.

* Sampling error can also be calculated.

* Correlation is the basis for the concept of regression and ratio of variation.

Types of correlation :-

Correlation is classified into many types, but the important are

1. positive and negative
2. simple and multiple
3. partial and total
4. Linear and non-linear

1. positive and negative correlation :-

positive and negative correlation depend upon the of change of the variables. If two variables tend to move together in the same direction an increase in the value of one variable is accompanied

Scanned by TapScanner

by an increase in the value of other variable, the Corelation is Called positive or direct corelation Height and weight, rainfall and yield of crops, price and supply are example of positive Corelation.

2. Simple and multiple :-

When we study only two variables, the relationship is described a simple Corelation, example quantity of money and price level, demand and price etc. In a multiple Corelation we study more than two variables. Ex relationship of price, demand, Supply of Commodity.

3. partial and total :-

The study of two variables excluding some other variables is called partial Corelation. For example we study price and demand, eliminating the supply side. In the Corelation, all the facts are taken into account.

4. Linear and non-linear :-

If the ratio of change between two variables is uniform, then linear Corelation between them.

The ratio of change between the variables is the same. If we plot these on straight line.

Scanned by TapScanner

Nine Squares table :-

	A	α	To
B	(AB)	(αB)	(B)
β	(Aβ)	($\alpha\beta$)	(β)
To	(A)	(α)	(N)

N = (B) + (β)

N = (AB) + (αB) + (Aβ) + ($\alpha\beta$)

A = (AB) + (Aβ)

α = (αB) + ($\alpha\beta$)

B = (AB) + (αB)

β = (Aβ) + ($\alpha\beta$)

AB = B - αB

αB = B - AB

Aβ = β - $\alpha\beta$

$\alpha\beta$ = β - Aβ

Yule's Co-efficient of association :-

The above mentioned will give us rough idea about their association but the degree of association cannot be found out. Prof. yule has suggested a formula to measure the association falls between ±1. If Q = 0, no association if Q = +1, there is perfect positive association and if Q = -1, there is perfect negative association.

Scanned by TapScanner

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)} \quad //$$

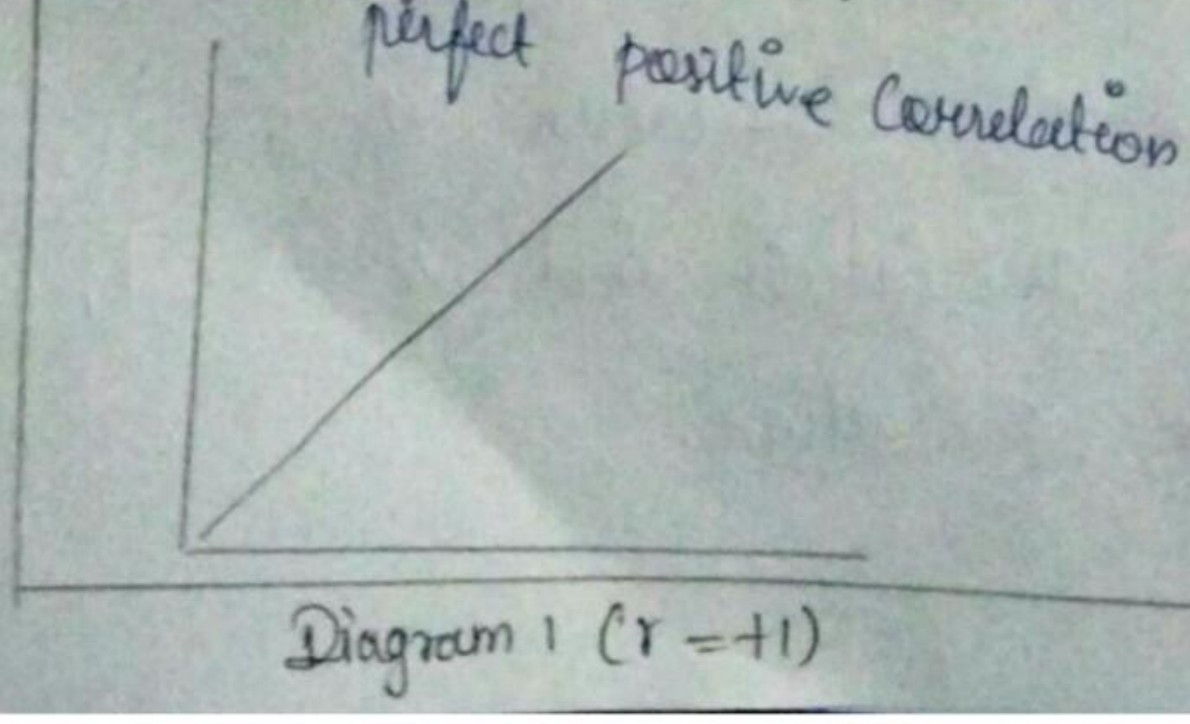
Unit - II

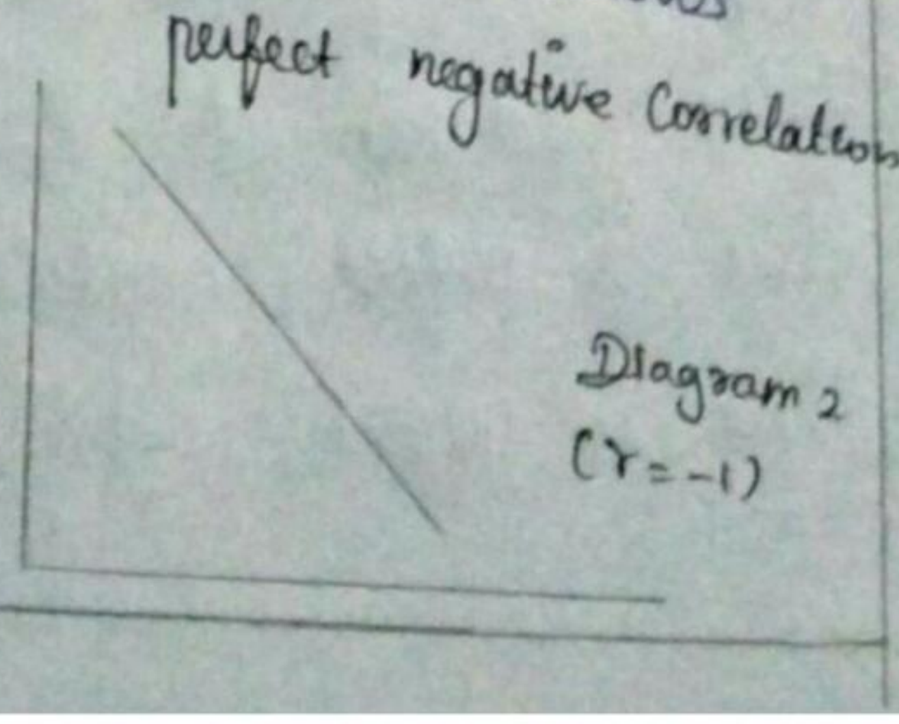
Simple Correlation :-

Simple linear Correlation is a measure of the degree to which two variable vary together, or a measure of the intensity of the association between two variables. Correlation often is abused. You need to show that one variable actually is affecting another variable. A simple Correlation coefficient can range from 1 to 1. However, maxi; or mini) values of some simple Correlation cannot reach unity.

Scatter diagram Method :-

This is the simplest method of finding out whether there relationship present between two variables by plotting the values on a chat, known as scatter diagram. X variables are on the horizontal axis and Y variables on the vertical axis. This will shows the type of Correlation.

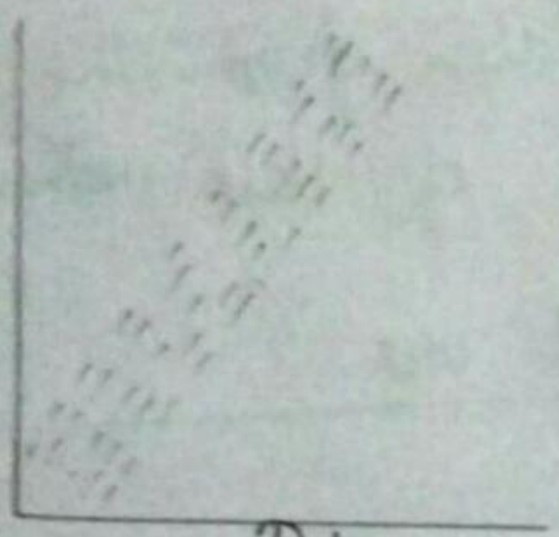




Scanned by TapScanner

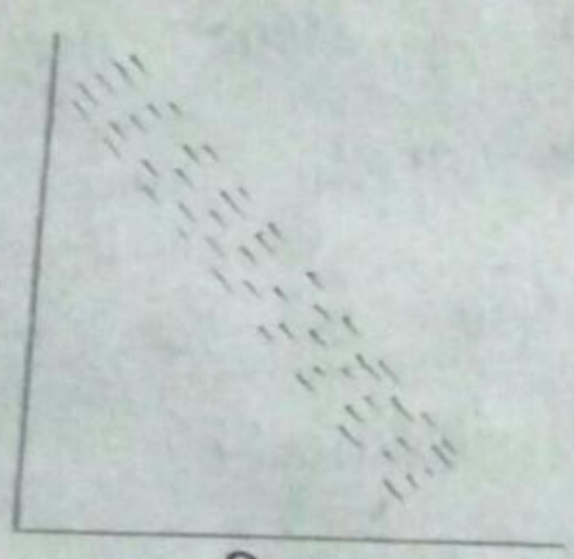
If the plotted point form a straight line running from the lower left-hand corner to the upper right hand corner, the perfect positive correlation ($r = +1$). on the other hand, the point of straight line, having a falling trend from the upper left-hand corner to lower right hand corner, is negative or inverse correlation.

If the plotted points fall in narrow band, point from left hand to upper right hand corner. If the plotted points fall in narrow band from the the upper left-hand corner to the lower right hand corner.



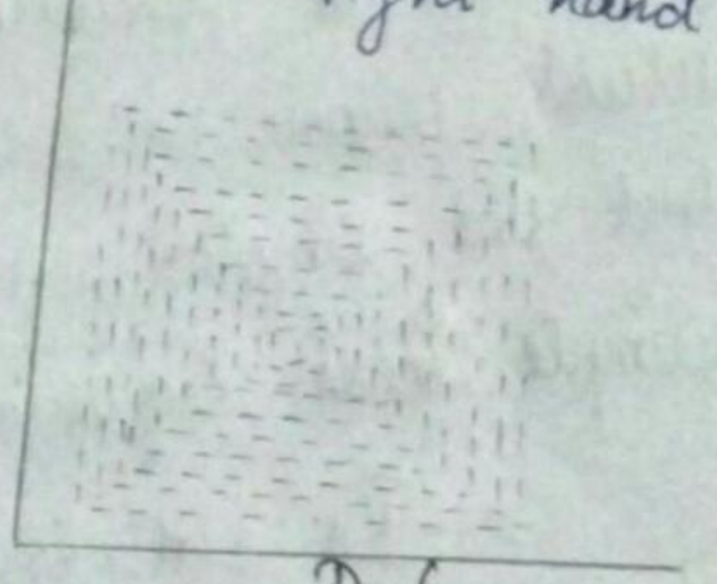
D-4

High degree of positive correlation



D-5

High degree of negative correlation



D-6

No correlation ($r = 0$)

Karl pearson's Coefficient of Correlation :-

Karl pearson, a great biometrician and Statistician, agreesed a mathematical method for measuring the magnitude of linear relationship between two karl pearson's method is the most widely

Scanned by TapScanner

used method in practice and is known as pearsonian Coefficient of Corelation. It is denoted by the Symbol 'r'. the formula for Calculating pearsonian

(i) $r = \frac{\text{Covariance of } xy}{\sigma x \times \sigma y}$ (2) $r = \frac{\Sigma xy}{N \sigma_x \times \sigma_y}$, (3).

(3) $= \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$

$X = (x - \bar{x})$ $Y = (y - \bar{y})$

σx = Standard deviation of series x

σy = Standard deviation of series y

When the deviation of item are taken from the actual mean, we Can apply any one of these methods, but the simplest formula is the third one.

Simple Linear Regression Analysis :-

In statistics, simple linear regression is a linear regression model with a single explanatory variable.

$y = \alpha + \beta x$

β = Slope

α = y- intercept

y = Y- coordinates

x = x - Coordinates

Scanned by TapScanner

Simple linear regression is the model that estimates the relationship between one independent variable and one dependent variable using a straight line. Both variables should be quantitative.

Ex. 1.
The coefficient of rank correlation of marks obtained by 10 students in english and economics was found to be 0.5. It was later discovered that difference in ranks two subject obtained by one of student was wrongly taken as 3 instead of 7. Find the coefficient of rank correlation.

Solution :-

$$r = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

Sub the values in the formula

$$0.5 = 1 - \frac{6\Sigma D^2}{10(100-1)}$$

$$\frac{6\Sigma D^2}{990} = 1 - 0.5 = 0.5$$

$$6\Sigma D^2 = 0.5 \times 990$$

$$\Sigma D^2 = \frac{0.5 \times 990}{6}$$

$$\Sigma D^2 = 82.5$$

$$\Sigma D^2 = 82.5 - 3^2 + 7^2 = 122.5$$

Scanned by TapScanner

$$\text{Corrected } r = 1 - \frac{6 \times 122.5}{10 \times 99}$$

$$= 1 - \frac{735}{990}$$

$$= \frac{225}{990}$$

$$r = +0.2576$$

---

2. In a Co-educational institution out of 200 students, 150 were boys. They took an examination and it was found that 120 passed, 10 girls failed. Is there association between sex and success in the examination.

Solution:-

Let A denote boys and α denote girls.
Let B denote those who passed the examination and β denote those who failed.

We have given $N = 200$, $(A) = 150$, $(AB) = 120$, $(\alpha\beta) = 10$.

Other frequencies can be obtained from nine square table.

	A	α	Total
B	120	40	160
β	30	10	40
Total	150	50	200

Other frequecies Applying yule's method:-

Scanned by TapScanner

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$Q = \frac{(120 \times 10) - (30 \times 40)}{(120 \times 10) + (30 \times 40)}$$

$$= \frac{1200 - 1200}{1200 + 1200} = 0 \text{ //.}$$

3. For two attributes A and B, we have $(AB) = 8$; $(A) = 18$; $(\alpha\beta) = 5$; $x = 35$. Calculating the co-efficient of association.

Solution :-

	(A)	(α)	To
(B)	8 (AB)	12 (αB)	20 (B)
(β)	10 (Aβ)	5 (αβ)	15 (β)
To	18 (A)	17 (α)	35 N

$$Q = \frac{(AB)(\alpha\beta) - (\alpha B)(A\beta)}{(AB)(\alpha\beta) + (\alpha B)(A\beta)}$$

$$= \frac{8 \times 5 - 10 \times 12}{8 \times 5 + 10 \times 12}$$

$$= \frac{-80}{160}$$

$$= -0.5 \text{ //.}$$

Scanned by TapScanner

A. Find if A and B are independent, positively associated or negatively associated from the given below.

(A) = 470, (B) = 620, (AB) = 320, N = 1000

Sol:-

Attributes A and B shall be called the independent if:

$$(AB) = \frac{(A) \times (B)}{N}$$

$$\text{Expectation of } (AB) = \frac{470 \times 620}{1000} = 291.4$$

Since (AB) actual observation (320) is more than the expectation (291.4) attributes A and B are positively associated.

Scanned by TapScanner

Unit - III

Basic Sampling Methods:-

* Simple random sampling. In this case each individual is chosen entirely by chance and each member of the population has an equal chance or probability of being selected.

* There are five types of sampling : Random, Systematic, Convenience, Cluster and stratified. Sampling methods are the way to choose people from the population to be considered in a sample survey. Samples can be divided based on following criteria. probability Samples - In such samples, each other population element has a known probability or chance of being chosen for the sample.

* Finally, the best sampling method is always the one that could best answer our research question while also allowing for others to make use of our result (generalisability of results). When we cannot afford a random sampling method, we can always choose from the non-random sampling methods.

Scanned by TapScanner

The basic concept of sampling and a population is the group who is the main focus of a researcher's interest ; a sample is the group from whom the researcher actually collects data. Sampling involves selecting the observation that you will analyze. To conduct sampling a researcher starts by going where your participants are.

probability Sampling methods.

Simple random sampling. In a simple random sample, every member of the population has an equal chance of being selected. Systematic Sampling. The Systematic Sampling is similar to Simple random Sampling, but it is similar harder to conduct. Stratified Sampling. Cluster sampling.

There are two types of Sampling methods :-

* The probability Sampling involves random selection, allowing you to make statistical inferences about the whole group.

* Non-probability Sampling involves non-random selection based on convenience or other criteria, allowing you to easily collect initial data. Second, Stratified random Sampling will generally have more statistical precision than

Scanned by TapScanner

Simple random of sampling method is the statistically most efficient.

Simple random sampling :-

Simple random sampling is a type of probability sampling in which the researches randomly selects a subset of participants from a population. Each member of the population has an equal chance of being selected. Data is than collected from as large a percentage as possible of this random subset.

In Statistic, a simple random sample is subset of individuals chosen from a larger set. Each individual is choosen randomly and entirely by chance, such that each individual has the same probability.

The simple random sampling. In a simple random sample, every member of the population has an equal chance of being selected.

There are requires using randomly generated numbers to choose a sample.

Stratified random sampling and cluster random sampling and these systematic random sampling.

Scanned by TapScanner

Stratified random Sampling :-

In Statistics, Stratified Sampling is a method of sampling from a population which can be partitioned from a the partitioned into subpopulation. In Statistical Surveys, when subpopulation within an overall population vary, it could be advantageous to Sample each subpopulation independently.

Stratified random Sampling is method of Sampling that involves the division of a population into smaller sub-groups known as Strata. In Stratified random Sampling or the Stratification, the strata are formed based on members shared attributes or characteristics such as income or educational attainment.

A Stratified random Sampling involves dividing the entire population into homogeneous groups Called strata (plural of from Stratum). A random Sample from each stratum is taken in a number proportional to the stratum's size when Compared to the population. There subsets of the Strata are the pooled to form a random Sample.

Scanned by TapScanner

* Systematic Sampling is a Symmetrical process where the researcher chooses the sample after a specifically defined interval. Sampling like this leaves the researcher no room for bias regarding choosing the sample.

* There are probability Sampling method where researchers select members of the population at a regular interval - for example, by selecting every 15th person on list of population. If the population is random order, this can imitate the benefit of Simple random Sampling.

* Probability Sampling means that every member of the target population has a known chance of being included in the Sample.

There are the methods include Simple random Sampling, Systematic Sampling, Stratified Sampling and the Cluster Sampling.

(k) = N/n

There are calculated by dividing the population size by the desired sample size.

Scanned by TapScanner

Systematic Sampling :-

Systematic sampling is a type of probability sampling method in which sample members from a larger population are selected according to a random starting point but with a fixed, periodic interval. This interval called the sampling interval, is calculated by dividing the population size by the desire sample size.

$$k = \frac{N}{n}$$

k = Systematic sampling interval

N = population size

n = Sample Size

* Calculating the Sampling interval (the number of households in the population divided by the number of households needed for the sample)

* Select a random start between 1 and Sampling interval.

* Repeatedly add sampling interval to select subsequent households.

Scanned by TapScanner

Quota Sampling :-

Quota Sampling is a method for the Selecting Survey participants that is a non-probabilistic version of Stratified sampling.

There are defined as a non-probalesility Sampling method in which researchers create a Sample involving individuals that represent a population.

They decide and create samples can lee useful in collecting data. There samples can lee generalized to the entire population.

It means to take a very tailored sample that is proportion to some characteristic or trait of a population. For example, you could divide a population ley the State they live in, income or education level or sex... Care is taken to maintain the correct proportions representation of the population.

Scanned by TapScanner