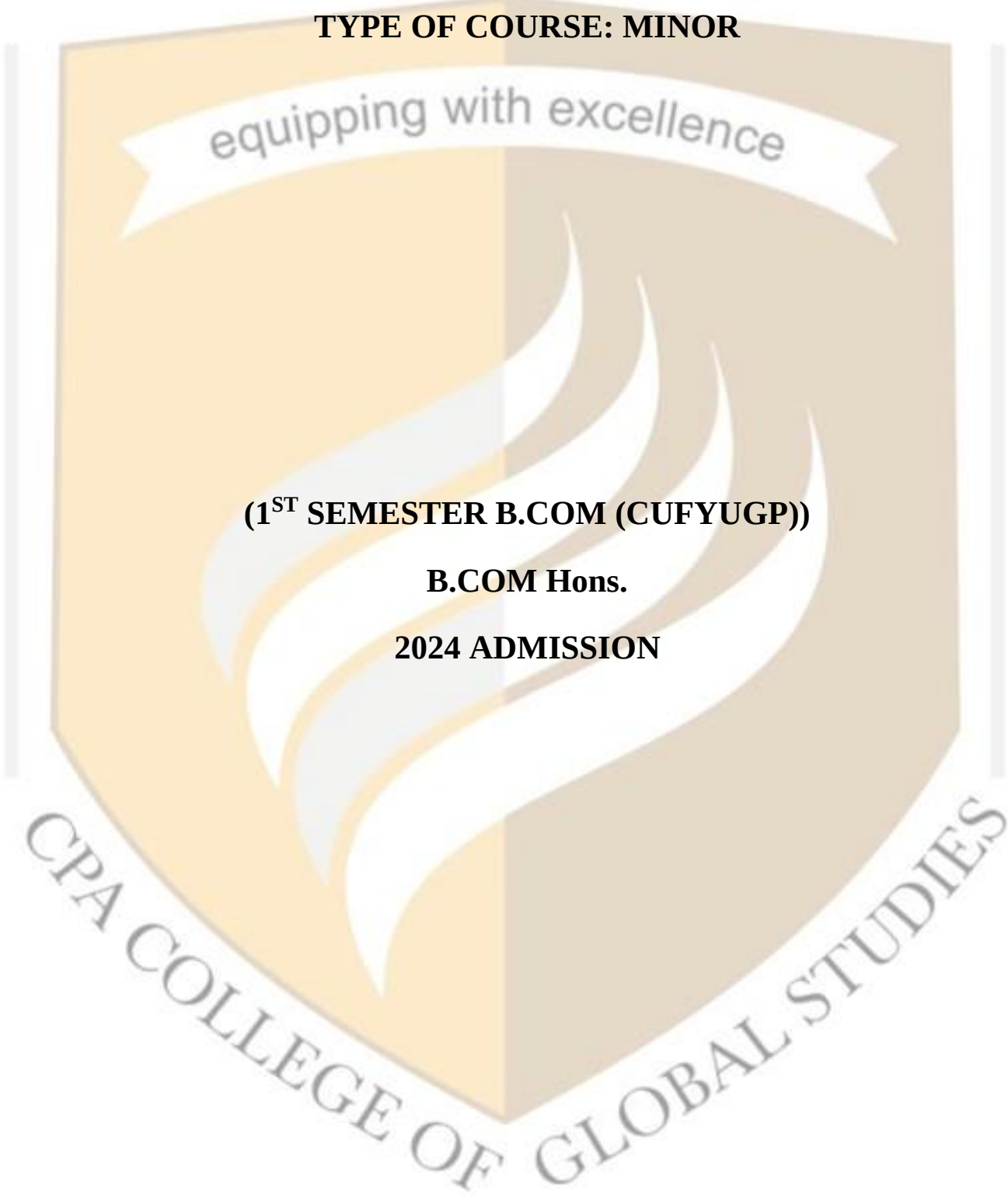
COURSE CODE: COM1MN109
ESSENTIAL STATISTIC FOR BUSINESS ANALYTICS
TYPE OF COURSE: MINOR



(1ST SEMESTER B.COM (CUFYUGP))

B.COM Hons.

2024 ADMISSION

Prepared by
Mohammed Jabir cv
Department of commerce (co-operation)
CPA college of global studies

COURSE CODE: COM1MN109
ESSENTIAL STATISTIC FOR BUSINESS ANALYTICS
TYPE OF COURSE: MINOR

Syllabus

Module 1

1.Introduction to sampling and sampling theory
2. Sampling distribution-using probabilistic sampling techniques
3. Estimating sampling errors and confidence intervals-sampling error and non-sampling error
4. Central limit theorem
5. Sampling techniques

Module 2

1.Introduction to hypothesis testing-null hypothesis-alternate hypothesis
2 testing hypothesis for larger samples- test for single proportion-test for different of proportions
3.testing hypothesis for small samples -T-test
4.calculating analysis of variance- two-way factorial ANOVA-multivariate analysis of variance
5.performing chi-square test -testing the goodness of fit

Module 3

1. Correlation -meaning & types - spearman's rank correlation, Karl Pearson's co efficient of correlation
2. Analysis of Bi variate data
3. Regression - meaning -regression lines- properties of regression lines and regression coefficient

Module 4

1.Objectives and uses of time series analysis
2.componants of time series
3.measuring trend using free hand curve method and method of semi averages
4.measurement of trend by moving averages method
5.least squares method

Module 1
Introduction to Sampling

1. Introduction to Sampling and Sampling Theory

- Definition of Sampling: Selecting a subset of individuals or observations from a larger population to represent the characteristics of the population.
- Example: A survey of 1000 voters to predict the outcome of an election.
- Purpose of Sampling: To make inferences about the population based on the sample data.
- Example: Using a sample of students to estimate the average GPA of the entire student body.

2. Sampling Distribution - Using Probabilistic Sampling Techniques

- Definition of Sampling Distribution: The probability distribution of a sample statistic (e.g., sample mean, sample proportion).
- Example: The distribution of sample means from repeated samples of size 100 from a population with a mean of 50.
- Types of Sampling Distributions: Sampling distribution of the mean, sampling distribution of the proportion.
- Example: The sampling distribution of the proportion of successes in a sample of 100 trials, where the population proportion is 0.4.

3. Estimating Sampling Errors and Confidence Intervals

- Definition of Sampling Error: The difference between the sample statistic and the population parameter.
- Example: The difference between the sample mean and the population mean.
- Types of Errors: Sampling error (random error) and non-sampling error (systematic error).
- Example: A survey question that is worded in a way that leads to biased responses (non-sampling error).
- Confidence Intervals: A range of values within which the population parameter is likely to lie.
- Example: A 95% confidence interval for the population mean, based on a sample mean of 50 and a sample standard deviation of 10.

4. Central Limit Theorem (CLT)

- Statement of CLT: Given a large enough sample size, the sampling distribution of the sample mean will be approximately normally distributed, regardless of the population distribution.
- Example: The sampling distribution of the mean of a sample of 1000 observations from a population with a skewed distribution.
- Assumptions of CLT: Large sample size ($n \geq 30$), random sampling, independence of

observations.
- Example: A survey of 500 customers, where each customer is randomly selected and interviewed independently.

5. Sampling Techniques

- Probability Sampling Techniques:
- Simple Random Sampling (SRS): Each member of the population has an equal chance of being selected.
- Example: A lottery drawing, where each ticket has an equal chance of being selected.
- Stratified Random Sampling: The population is divided into subgroups, and a random sample is taken from each subgroup.
- Example: A survey of customers, where the population is divided into subgroups by age, income, and location.
- Non-Probability Sampling Techniques:
- Convenience Sampling: A sample is selected based on ease of access or convenience.
- Example: A survey of customers at a shopping mall, where the sample is selected based on who happens to be present.
- Quota Sampling: A sample is selected to match the characteristics of the population, but the selection is not random.
- Example: A survey of customers, where the sample is selected to match the age, income, and location characteristics of the population

Module 2
Hypothesis Testing

1. Introduction to Hypothesis Testing

Hypothesis testing is a statistical technique used to make inferences about a population based on a sample. It involves testing a null hypothesis (H0) against an alternative hypothesis (H1 or Ha).

Key Concepts:
- Null Hypothesis (H0): A statement of no effect or no difference.
- Alternate Hypothesis (H1 or Ha): A statement of an effect or difference.
- Type I Error: Rejecting the null hypothesis when it is true.
- Type II Error: Failing to reject the null hypothesis when it is false.

2. Testing Hypothesis for Larger Samples

When working with larger samples, we can use various tests to determine if there is a significant difference between the sample and the population.

Key Concepts:
- Test for Single Proportion: Used to determine if a population proportion is equal to a known value.
- Test for Difference of Proportions: Used to determine if there is a significant difference between two population proportions.
- Z-Score: A measure of how many standard deviations an observation is away from the mean.

3. Testing Hypothesis for Small Samples

When working with smaller samples, we use the t-test to determine if there is a significant difference between the sample and the population.

Key Concepts:
- T-Test: Used to determine if there is a significant difference between the means of two groups.
- T-Score: A measure of how many standard deviations an observation is away from the mean.
- Degrees of Freedom: The number of values in the final calculation of a statistic that are free to vary.

4. Calculating Analysis of Variance (ANOVA)

ANOVA is a statistical technique used to compare means among three or more groups.

Key Concepts:
- Two-Way Factorial ANOVA: Used to determine the effect of two independent variables on a continuous outcome variable.
- F-Score: A measure of the ratio of the variance between groups to the variance within groups.
- P-Value: The probability of observing the test results assuming that the null hypothesis is true.

5. Performing Chi-Square Test

The chi-square test is a statistical technique used to determine if there is a significant association between two categorical variables.

Key Concepts:
- Chi-Square Test: Used to determine if there is a significant association between two categorical variables.
- Chi-Square Statistic: A measure of the difference between the observed and expected frequencies.
- P-Value: The probability of observing the test results assuming that the null hypothesis is true.

ILLUSTRATION AND SOLUTION

1. Introduction to Hypothesis Testing
Null Hypothesis (H0)
A statement of no effect or no difference.

Alternate Hypothesis (H1 or Ha)
A statement of an effect or difference.

Example:

- H0: The average height of adults in a population is 175 cm.
- H1: The average height of adults in a population is not 175 cm.

2. Testing Hypothesis for Larger Samples
Test for Single Proportion
Used to determine if a population proportion is equal to a known value.

Example:

- A company claims that 80% of its customers are satisfied. A random sample of 100 customers shows that 75 are satisfied. Test the claim at a 5% significance level.

Solution:

- H0: $p = 0.8$
- H1: $p \neq 0.8$
- Sample proportion ($\hat{p}$) = 75/100 = 0.75
- Standard error (SE) = $\sqrt{(\hat{p}(1-\hat{p})/n)} = \sqrt{(0.75(1-0.75)/100)} = 0.043$
- Z-score = $(\hat{p} - p) / SE = (0.75 - 0.8) / 0.043 = -1.16$
- P-value = $2P(Z > |z\text{-score}|) = 2P(Z > 1.16) = 0.245$
- Since the p-value is greater than the significance level (0.05), we fail to reject the null hypothesis.

Test for Difference of Proportions
Used to determine if there is a significant difference between two population proportions.

Example:

- A study compares the proportion of smokers in two cities. City A has a sample proportion of 0.25, while City B has a sample proportion of 0.30. Test the difference at a

5% significance level.

Solution:

- H0: p1 - p2 = 0
- H1: p1 - p2 ≠ 0
- Sample proportions ($\hat{p}1$, $\hat{p}2$) = (0.25, 0.30)
- Standard error (SE) = $\sqrt{(\hat{p}1(1-\hat{p}1)/n1 + \hat{p}2(1-\hat{p}2)/n2)}$ = $\sqrt{(0.25(1-0.25)/100 + 0.30(1-0.30)/100)}$ = 0.058
- Z-score = ($\hat{p}1 - \hat{p}2$) / SE = (0.25 - 0.30) / 0.058 = -0.86
- P-value = 2P(Z > |z-score|) = 2P(Z > 0.86) = 0.389
- Since the p-value is greater than the significance level (0.05), we fail to reject the null hypothesis.

3. Testing Hypothesis for Small Samples

T-Test

Used to determine if there is a significant difference between the means of two groups.

Example:

- A study compares the average scores of two groups of students. Group A has a sample mean of 80, while Group B has a sample mean of 85. Test the difference at a 5% significance level.

Solution:

- H0: $\mu1 - \mu2 = 0$
- H1: $\mu1 - \mu2 \neq 0$
- Sample means ($\bar{x}1$, $\bar{x}2$) = (80, 85)
- Sample standard deviations (s1, s2) = (10, 12)
- Degrees of freedom (df) = n1 + n2 - 2 = 20 + 20 - 2 = 38
- T-score = ($\bar{x}1 - \bar{x}2$) / SE = (80 - 85) / $\sqrt{((s1^2/n1) + (s2^2/n2))}$ = -2.13
- P-value = 2P(T > |t-score|) = 2P(T > 2.13) = 0.039
- Since the p-value is less than the significance level (0.05), we reject the null hypothesis.

4. Calculating Analysis of Variance (ANOVA)

Two-Way Factorial ANOVA

Used to determine the effect of two independent variables on a continuous outcome variable.

Example:

- A study examines the effect of temperature (high, low) and pressure (high, low) on the yield of a chemical reaction.

Solution:

- H0: There is no significant interaction between temperature and pressure.
- H1: There

MODULE 3

CORRELATION ANALYSIS

Introduction:

- In practice, we may come across with lot of situations which need statistical analysis of either one or more variables. The data concerned with one variable only is called univariate data. For Example: Price, income, demand, production, weight, height marks etc are concerned with one variable only. The analysis of such data is called univariate analysis.
- The data concerned with two variables are called bivariate data. For example: rainfall and agriculture; income and consumption; price and demand; height and weight etc. The analysis of these two sets of data is called bivariate analysis.
- The date concerned with three or more variables are called multivariate date. For example: agricultural production is influenced by rainfall, quality of soil, fertilizer etc.

Definition:

Two or more variables are said to be correlated if the change in one variable results in a corresponding change in the other variable.

According to Simpson and Kafka, ―Correlation analysis deals with the association between two or more variables‖.

Lunchou defines,―Correlation analysis attempts to determine the degree of relationship between variables‖.

Boddingtonstates that ―Whenever some definite connection exists between two or more groups or classes of series of data, there is said to be correlation.‖

Correlation Coefficient:

Correlation analysis is actually an attempt to find a numerical value to express the extent of relationship exists between two or more variables. The numerical measurement showing the degree of correlation between two or more variables is called correlation coefficient. Correlation coefficient ranges between -1 and +1.

SIGNIFICANCE OF CORRELATION ANALYSIS

Correlation analysis is of immense use in practical life because of the following reasons:

1. Correlation analysis helps us to find a single figure to measure the degree of relationship exists between thevariables.
2. Correlation analysis helps to understand the economicbehavior
3. Correlation analysis enables the business executives to estimate cost, price and other variables.
4. Correlation analysis can be used as a basis for the study of regression. Once we know that two variables are closely related, we can estimate the value of one variable if the value of other isknown.
5. Correlation analysis helps to reduce the range of uncertainty associated with decision making. The prediction based on correlation analysis is always near toreality.
6. It helps to know whether the correlation is significant or not. This is possible by comparing the correlation co-efficient with 6PE. It ‗r' is more than 6 PE, the correlationissignificant.

Classification of Correlation

Correlation can be classified in different ways. The following are the most important classifications

1. Positive and Negativecorrelation
2. Simple, partial and multiplecorrelation
3. Linear and Non-linear correlation

Positive and Negative Correlation PositiveCorrelation

When the variables are varying in the same direction, it is called positive correlation. In other words, if an increase in the value of one variable is accompanied by an increase in the value of other variable or if a decrease in the value of one variable is accompanied by a decree se in the value of other variable, it is called positive correlation.

Eg:1)	A:10	20	30	40	50
	B: 80	100	150	170	200

When the variables are moving in opposite direction, it is called negative correlation. In other words, if an increase in the value of one variable is accompanied by a decrease in the value of other variable or if a decrease in the value of one variable is accompanied by an increase in the

value of other variable, it is called negative correlation.

Simple, Partial and Multiple correlation

Simple Correlation

In a correlation analysis, if only two variables are studied it is called simple correlation. Eg.the study of the relationship between price & demand, of a product or price and supply of a product is a problem of simple correlation.

Multiple correlation

In a correlation analysis, if three or more variables are studied simultaneously, it is called multiple correlations. For example, when we study the relationship between the yield of rice with both rainfall and fertilizer together, it is a problem of multiple correlation.

Partial correlation

In a correlation analysis, we recognize more than two variable, but consider one dependent variable and one independent variable and keeping the other Independent variables as constant. For example yield of rice is influenced b the amount of rainfall and the amount of fertilizer used. But if we study the correlation between yield of rice and the amount of rainfall by keeping the amount of fertilizers used as constant, it is a problem of partial correlation.

Linear and Non-linear correlation

Linear Correlation

In a correlation analysis, if the ratio of change between the two sets of variables is same, then it is called linear correlation.

For example when 10% increase in one variable is accompanied by 10% increase in the other variable, it is the problem of linear correlation.

X:	10	15	30	60
Y:	50	75	150	300

Here the ratio of change between X and Y is the same. When we plot the data in graph paper, all the plotted points would fall on a straight line.

Non-linear correlation

In a correlation analysis if the amount of change in one variable does not bring the same ratio of change in the other variable, it is called nonlinear correlation.

X:	2	4	6	10	15
Y:	8	10	18	22	26

Here the change in the value of X does not being the same proportionate change in the value of Y

Degrees of correlation:

Correlation exists in various degrees

1. **Perfect positivecorrelation**

If an increase in the value of one variable is followed by the same proportion of increase in other related variable or if a decrease in the value of one variable is followed by the same proportion of decrease in other related variable, it is perfect positive correlation. eg: if 10% rise in price of a commodity results in 10% rise in its supply, the correlation is perfectly positive. Similarly, if 5% full in price results in 5% fall in supply, the correlation is perfectlypositive.

2. **Perfect Negativecorrelation**

If an increase in the value of one variable is followed by the same proportion of decrease in other related variable or if a decrease in the value of one variable is followed by the same proportion of increase in other related variably it is Perfect Negative Correlation. For example if 10% rise in price results in 10% fall in its demand the correlation is perfectly negative. Similarly if 5% fall in price results in 5% increase in demand, the correlation is perfectlynegative.

3. **Limited Degree of Positivecorrelation**:

When an increase in the value of one variable is followed by a non-proportional increase in other related variable, or when a decrease in the value of one variable is followed by a non-proportional decrease in other related variable, it is called limited degree of positivecorrelation.

For example, if 10% rise in price of a commodity results in 5% rise in its supply, it is limited degree of positive correlation. Similarly if 10% fall in price of a commodity results in 5% fall in its supply, it is limited degree of positivecorrelation.

4. **Limited degree of Negativecorrelation**

When an increase in the value of one variable is followed by a non-proportional decrease in other related variable, or when a decrease in the value of one variable is followed by a non-proportional increase in other related variable, it is called limited degree of negative correlation.

For example, if 10% rise in price results in 5% fall in its demand, it is limited degree of negative correlation. Similarly, if 5% fall in price results in 10% increase in demand, it is limited degree of negative correlation.

5. **Zero Correlation (Zero Degreecorrelation)**

If there is no correlation between variables it is called zero correlation. In other words, if the values of one variable cannot be associated with the values of the other variable, it is zero correlation.

Methods of measuring correlation

Correlation between 2 variables can be measured by graphic methods and algebraic methods.

I GraphicMethods

1) ScatterDiagram

2) Correlationgraph

II Algebraic methods (Mathematical methods or statistical methods or Co-efficient of correlation methods):

1) Karl Pearson‘s Co-efficient ofcorrelation

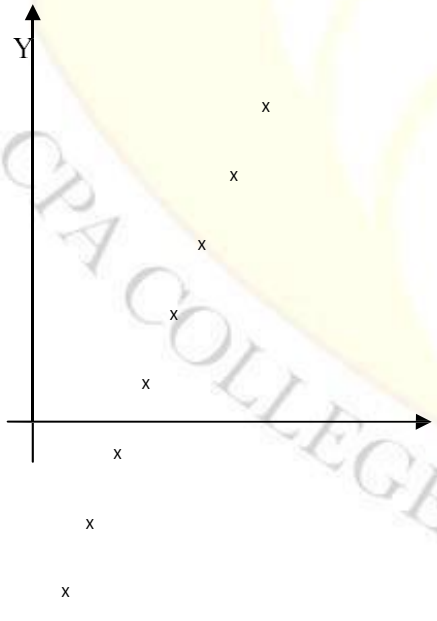
2) Spear man‘s Rank correlationmethod

3) Concurrent deviationmethod

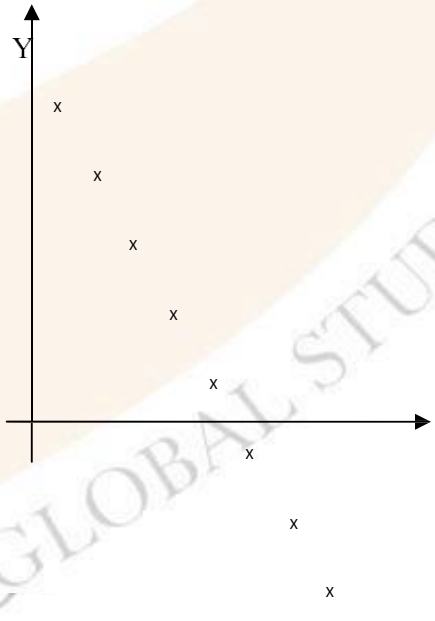
Scatter Diagram

This is the simplest method for ascertaining the correlation between variables. Under this method all the values of the two variable are plotted in a chart in the form of dots. Therefore, it is also known as dot chart. By observing the scatter of the various dots, we can form an idea that whether the variables are related or not.

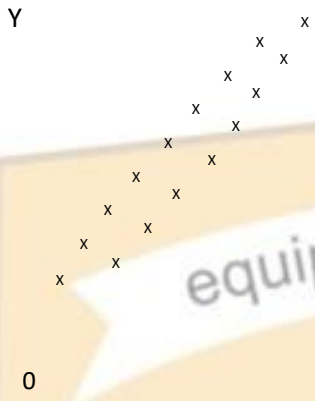
A scatter diagram indicates the direction of correlation and tells us how closely the two variables under study are related. The greater the scatter of the dots, the lower is the relationship



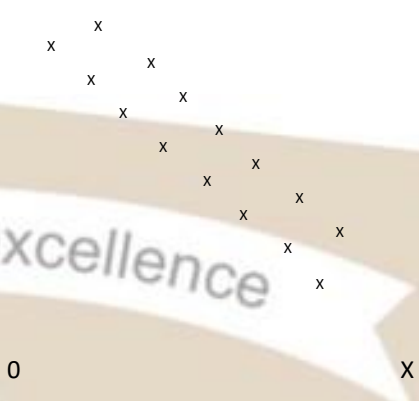
PerfectPositiveCorrelation



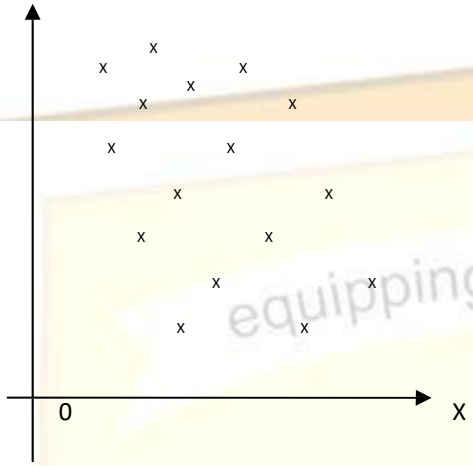
Perfect NegativeCorrelation



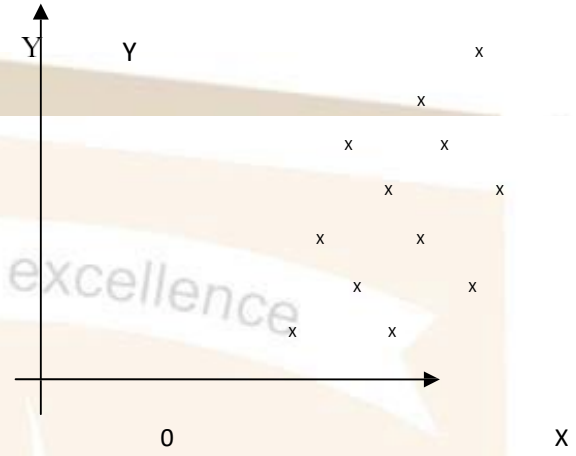
High Degree of Positive Correlation



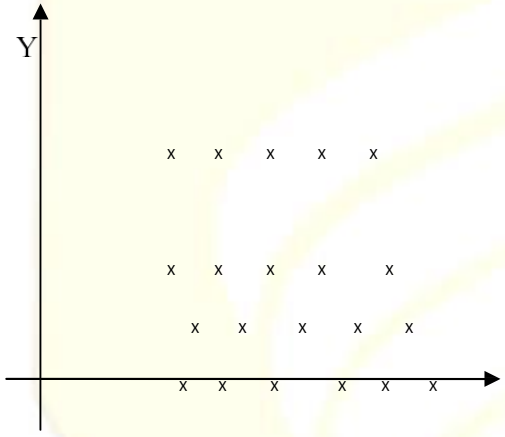
High Degree of Negative Correlation



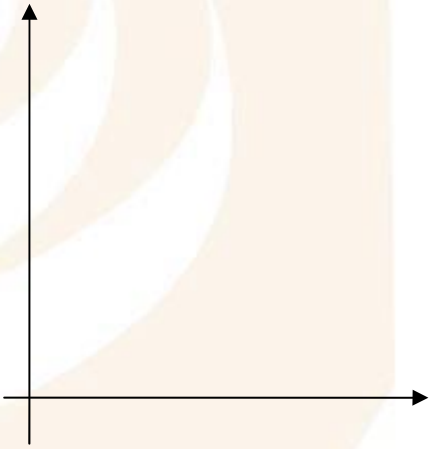
Low Degree ofPositiveCorrelation



Low Degree of NegativeCorrelation

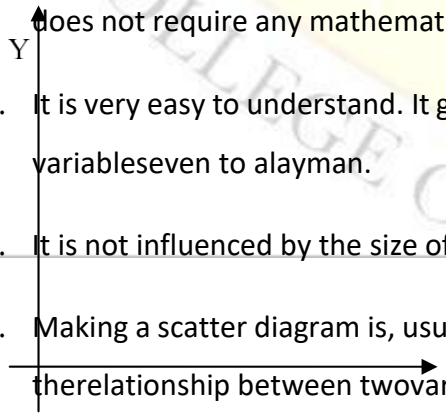


No Correlation (r = 0)



Merits of Scatter Diagram method

1. It is a simple method of studying correlation betweenvariables.
2. It is a non-mathematical method of studying correlation between the variables. It does not require any mathematicalcalculations.



3. It is very easy to understand. It gives an idea about the correlation between variableseven to alayman.
4. It is not influenced by the size of extremeitems.
5. Making a scatter diagram is, usually, the first step in investigating therelationship between twovariables.

Demerits of Scatter diagram method

1. It gives only a rough idea about the correlation betweenvariables.
2. The numerical measurement of correlation co-efficient cannot be calculated under this method.
3. It is not possible to establish the exact degree of relationship between thevariables.

Correlation graph Method

Under correlation graph method the individual values of the two variables are plotted on a graph paper. Then dots relating to these variables are joined separately so as to get two curves. By examining the direction and closeness of the two curves, we can infer whether the variables are related or not. If both the curves are moving in the same direction(either upward or downward) correlation is said to be positive. If the curves are moving in the opposite directions, correlation is said to benegative.

Merits of Correlation Graph Method

1. This is a simple method of studying relationship between thevariable
2. This does not require mathematicalcalculations.
3. This method is very easy tounderstand

Demerits of correlation graph method:

1. A numerical value of correlation cannot becalculated.
2. It is only a pictorial presentation of the relationship betweenvariables.
3. It is not possible to establish the exact degree of relationship between thevariables.

Karl Pearson's Co-efficient of Correlation

Karl Pearson's Coefficient of Correlation is the most popular method among the algebraic methods for measuring correlation. This method was developed by Prof. Karl Pearson in 1896. It is also called product moment correlation coefficient.

$$\Sigma(s-\bar{s})(y-\bar{y})$$

$$r = \frac{}{\sum (x-\bar{x})^2 \times (y-\bar{y})}$$

Interpretation of Co-efficient of Correlation

Pearson's Co-efficient of correlation always lies between +1 and -1. The following general rules will help to interpret the Co-efficient of correlation:

1. When r - +1, It means there is perfect positive relationship betweenvariables.
2. Whenr=-1,itmeansthereisperfectnegativerelationshipbetweenvariables.
3. When r = 0, it means there is no relationship between thevariables.
4. When_r'iscloserto+1,itmeansthereishighdegreeofpositivecorrelationbe tween variables.
5. When_r'iscloserto– 1,itmeansthereishighdegreeofnegativecorrelationbetween variables.
6. When_r'iscloserto_O', itmeansthereislessrelationshipbetweenvariables.

Properties of Pearson's Co-efficient of Correlation

1. If there is correlation between variables, the Co-efficient of correlation liesbetween +1 and -1.
2. If there is no correlation, the coefficient of correlation is denoted by zero (ier=0)
3. It measures the degree and direction ofchange
4. If simply measures the correlation and does not help to predictcansation.
5. It is the geometric mean of two regressionsco-efficient.

Probable Error and Coefficient of Correlation

Probable error (PE) of the Co-efficient of correlation is a statistical device which measures the reliability and dependability of the value of co-efficient of correlation.

$$PE = 0.6745 \times \frac{1-r^2}{\sqrt{}}$$

If the value of coefficient of correlation (r) is less than the PE, then there is no evidence of correlation.

If the value of _r' is more than 6 times of PE, the correlation is certain and significant.

By adding and submitting PE from coefficient of correlation, we can find out the

upper and lower limits within which the population coefficient of correlation may be expected to lie. **Uses of PE**:

1) PE is used to determine the limits within which the population coefficientof correlation may be expected tolie.
2) It can be used to test whether the value of correlation coefficient of a sampleis significant with that of thepopulation

If r = 0.6 and N = 64, find out the PE and SE of the correlation coefficient. Also determine the limits of population correlation coefficient.

Merits of Pearson's Coefficient of Correlation:-

1. This is the most widely used algebraic method to measure coefficient ofcorrelation.
2. It gives a numerical value to express the relationship betweenvariables
3. It gives both direction and degree of relationship betweenvariables
4. It can be used for further algebraic treatment such as coefficient ofdetermination coefficient of non-determinationetc.
5. It gives a single figure to explain the accurate degree of correlation betweentwo variables

Demerits of Pearson's Coefficient of correlation

1. It is very difficult to compute the value of coefficient ofcorrelation.
2. It is very difficult tounderstand

Spearman's Rank Correlation Method

Pearson's coefficient of correlation method is applicable when variables are measured in quantitative form. But there were many cases where measurement is not possible because of the qualitative nature of the variable. For example, we cannot measure the beauty, morality, intelligence, honesty etc in quantitative terms. However it is possible to rank these qualitative characteristics in some order.

$$(R) = 1-\frac{6XD2}{N3—N}$$

Merits of Rank Correlation method

1. Rank correlation coefficient is only an approximate measure as the actual

values are not used for calculations

2. It is very simple to understand the method.
3. It can be applied to any type of data, ie quantitative and qualitative
4. It is the only way of studying correlation between qualitative data such as honesty, beauty etc.
5. As the sum of rank differences of the two qualitative data is always equal to zero, this method facilitates a cross check on the calculation.

Demerits of Rank Correlation method

1. Rank correlation coefficient is only an approximate measure as the actual values are not used for calculations.
2. It is not convenient when number of pairs (ie. N) is large
3. Further algebraic treatment is not possible.
4. Combined correlation coefficient of different series cannot be obtained as in the case of mean and standard deviation. In case of mean and standard deviation, it is possible to compute combine arithmetic mean standard deviation.

Concurrent Deviation Method:

Concurrent deviation method is a very simple method of measuring correlation. Under this method, we consider only the directions of deviations. The magnitudes of the values are completely ignored. Therefore, this method is useful when we are interested in studying correlation between two variables in a casual manner and not interested in degree (or precision).

Under this method, the nature of correlation is known from the direction of deviation in the values of variables. If deviations of 2 variables are concurrent, then they move in the same direction, otherwise in the opposite direction.

The formula for computing the coefficient of concurrent deviation is: -

$$r = \pm\sqrt{\pm\frac{(2c-N)}{N}}$$

Where N = No. of pairs of symbol

C = No. of concurrent deviations (ie, No. of + signs in _dx dy‘ column)

Steps:

1. Every value of _X‘ series is compared with its proceeding value. Increase is shownby

_+‘ symbol and decrease is shown by _-_

2. The above step is repeated for _Y‘ series and we get_dy‘
3. Multiply_dx‘by_dy‘andtheproductisshowninthenextcolumn.Theco lumn heading is_dxdy‘.
4. Takethetotalnumberof_+‘signsin_dxdy‘column._+‘signsin_dxdy‘colu mn denotes the concurrent deviations, and it is indicated by_C‘.
5. Apply theformula:

$$r = \pm J \pm \left(\frac{2c - N}{N}\right)$$

If2cΣN,thenr=+veandif2c€N,thenr=—ve.

REGRESSION ANALYSIS

Introduction:-

Correlation analysis analyses whether two variables are correlated or not. After having established the fact that two variables are closely related, we may be interested in estimating the value of one variable, given the value of another. Hence, regression analysis means to analyses the average relationship between two variables and thereby provides a mechanism for estimation or predication orforecasting.

The term _Regression‖ was firstly used by Sir Francis Galton in 1877. The dictionary meaning of the term _regression‖ is ―stepping back‖ to the average.

Definition:

―Regression is the measure of the average relationship between two or more variables in terms of the original units of the date‖.

―Regression analysisisan attempt to establish the nature of the relationship between variables-that is to study the functional relationship between the variables and thereby provides a mechanism for prediction orforecasting‖.

Types of Regression:-

There are two types of regression. They are linear regression and multiple regressions.

Linear Regression:

It is a type of regression which uses one independent variable to explain and/or predict the dependent variable.

Multiple Regression:

It is a type of regression which uses two or more independent variable to explain and/or predict the dependent variable.

Regression Lines:

Regression line is a graphic technique to show the functional relationship between the two variables X and Y. It is a line which shows the average relationship between two variables X and Y.

If there is perfect positive correlation between 2 variables, then the two regression lines are winding each other and to give one line. There would be two regression lines when there is no perfect correlation between two variables. The nearer the two regression lines to each other, the higher is the degree of correlation and the farther the regression lines from each other, the lesser is the degree ofcorrelation.

Properties of Regression lines:-

1. The two regression lines cut each other at the point of average of X and average of Y (i.e $\bar{X}$ and $\bar{Y}$)
2. When r = 1, the two regression lines coincide each other and give oneline.
3. When r = 0, the two regression lines are mutuallyperpendicular.

Regression Equations (Estimating Equations)

Regression equations are algebraic expressions of the regression lines. Since there are two regression lines, therefore two regression equations. They are :-

1. Regression Equation of X on Y:- This is used to describe the variations in the values of X for given changes inY.
2. Regression Equation of Y on X :- This is used to describe the variations in the value of Y for given changes inX.

Regression Equation of Y on X:-

$$Y = a + bx$$

The normal equations to compute _a' and _b' are: -

$$\Sigma y = Na +$$

$$b\Sigma x\Sigma xy=a\Sigma x+b\Sigma$$

$$x^2$$

Regression Equation of X on Y:-

$$X = a + by$$

Thenormalequationstocompute_a'and_b'are:-

$$\Sigma x = Na +$$

$$n\Sigma y\Sigma xy=a\Sigma y+b$$

$$\Sigma y^2$$

Properties of Regression Coefficient:

1. There are two regression coefficients. They are b_{xy} and b_{yx}
2. Both the regression coefficients must have the same signs. If one is +ve, the other will also be a +vevalue.
3. Th~~egeome~~tricmeanofregressioncoefficientswillbethecoefficientof correlation.$r = \sqrt{b_{sy}.b_{ys}}$.
4. If$x\bar{x}$and$y\bar{y}$arethesame,then the regressioncoefficient and correlation coefficient will be thesame.

Computation of Regression Co-efficient

Regression co-efficient can be calculated in 3 different ways:

1. Actual meanmethod

- Regression coefficient x on y (b_{xy}) = $\frac{\sum sy}{\sum y^2}$

- Regression coefficient y on x (b_{yx}) = $\frac{\sum sy}{\sum s^2}$

Correlation	Regression
It studies degree of relationship between variables	It studies the nature of relationship between variables
It is not used for prediction purposes	It is basically used for prediction purposes
It is basically used as a tool for determining the degree of relationship	It is basically used as a tool for studying cause and effect relationship
There may be nonsense correlation between two variables	There is no such nonsense regression
There is no question of dependent and independent variables	There must be dependent and independent variables

Module 4
Time Series Analysis

1. Objectives and Uses of Time Series Analysis

Time series analysis is a statistical technique used to analyze and forecast data collected over time. The main objectives of time series analysis are:

- Identify patterns and trends in the data
- Make predictions about future values
- Understand the underlying factors that affect the data

Time series analysis has various uses, including:

- Forecasting sales and revenue
- Analyzing stock prices and market trends
- Understanding climate patterns and weather forecasting
- Modeling population growth and demographic changes

2. Components of Time Series

A time series can be broken down into four main components:

1. Trend: The overall direction or pattern in the data over time.
2. Seasonality: Regular fluctuations that occur at fixed intervals, such as daily, weekly, monthly, or yearly cycles.
3. Cycles: Long-term fluctuations that are not regular or predictable.
4. Irregular variations: Random and unpredictable events that affect the data.

3. Measuring Trend Using Free Hand Curve Method and Method of Semi-Averages

Free Hand Curve Method

This method involves plotting the data points on a graph and drawing a smooth curve that best fits the data. The curve represents the trend.

Method of Semi-Averages

This method involves dividing the data into two halves and calculating the average of each half. The two averages are then plotted on the graph, and a line is drawn through them to represent the trend.

4. Measurement of Trend by Moving Averages Method

The moving averages method involves calculating the average of a fixed number of consecutive data points. The average is then plotted on the graph, and the process is repeated for the next set of data points.

Simple Moving Average

The simple moving average is calculated by taking the average of a fixed number of consecutive data points.

Weighted Moving Average
The weighted moving average gives more weight to recent data points than to older data points.

5. Least Squares Method
The least squares method involves finding the best-fitting line to the data points by minimizing the sum of the squared errors.

Advantages
- Provides a more accurate representation of the trend
- Can handle non-linear relationships
- Can be used for forecasting

Disadvantages
- Can be sensitive to outliers
- Requires a large amount of data
- Can be computationally intensive

ILLUSTRATIONS AND SOLUTIONS

Measuring Trend Using Free Hand Curve Method and Method of Semi-Averages

Free Hand Curve Method

Illustration:

Plot data points on a graph
Draw a smooth curve that best fits the data

Example:

Year	Sales
2010	100
2011	120
2012	150
2013	180
2014	200

Draw a smooth curve through the data points to represent the trend.

Method of Semi-Averages

Illustration:

Divide data into two halves
Calculate average of each half
Plot averages on graph and draw a line through them

Example:

Year	Sales
2010	100
2011	120
2012	150
2013	180
2014	200

Divide data into two halves:

Half 1: 2010-2012
Half 2: 2013-2014

Calculate averages:

Half 1: (100 + 120 + 150) / 3 = 123.33
Half 2: (180 + 200) / 2 = 190

Plot averages on graph and draw a line through them.

4. Measurement of Trend by Moving Averages Method
Simple Moving Average

Illustration:

Calculate average of fixed number of consecutive data points
Plot average on graph

Example:

Year	Sales
2010	100
2011	120
2012	150
2013	180
2014	200

Calculate simple moving average with a window size of 3:

(100 + 120 + 150) / 3 = 123.33
(120 + 150 + 180) / 3 = 150
(150 + 180 + 200) / 3 = 176.67

Plot averages on graph.

Weighted Moving Average
Illustration:

Calculate weighted average of fixed number of consecutive data points
Plot average on graph

Example:

Year	Sales
2010	100
2011	120
2012	150
2013	180
2014	200

Calculate weighted moving average with a window size of 3 and weights of 0.2, 0.3, and 0.5:

(0.2 x 100 + 0.3 x 120 + 0.5 x 150) = 130
(0.2 x 120 + 0.3 x 150 + 0.5 x 180) = 155
(0.2 x 150 + 0.3 x 180 + 0.5 x 200) = 180

Plot averages on graph.

5. Least Squares Method
Illustration:

Plot data points on a graph
Draw a best-fitting line through the data points

Example:

Year	Sales
2010	100
2011	120
2012	150
2013	180
2014	200

Plot data points on a graph and draw a best-fitting line through the data points using the least squares method.

Solutions:

1. Time Series Analysis:

Objective: Identify patterns and trends in sales data.

Method: Use moving averages method or least squares method.

1. Measuring Trend:

Method: Use free hand curve method, method of semi-averages, or moving averages method.

1. Forecasting:

Method: Use least squares method or moving averages method.

1. Objectives and Uses of Time Series Analysis

Objectives
1. Identify patterns and trends: Analyze data to identify patterns and trends.
2. Make predictions: Use historical data to forecast future values.
3. Understand underlying factors: Identify factors that affect the data.

Uses
1. Forecasting sales and revenue
2. Analyzing stock prices and market trends
3. Understanding climate patterns and weather forecasting
4. Modeling population growth and demographic changes

2. Components of Time Series

Components
1. Trend: The overall direction or pattern in the data over time.
2. Seasonality: Regular fluctuations that occur at fixed intervals, such as daily, weekly, monthly, or yearly cycles.
3. Cycles: Long-term fluctuations that are not regular or predictable.
4. Irregular variations: Random and unpredictable events that affect the data.

3. Measuring Trend Using Free Hand Curve Method and Method of Semi-Averages

Free Hand Curve Method
1. Plot data points on a graph.
2. Draw a smooth curve that best fits the data.

Method of Semi-Averages
1. Divide data into two halves.
2. Calculate the average of each half.
3. Plot averages on graph and draw a line through them.

4. Measurement of Trend by Moving Averages Method

Simple Moving Average
1. Calculate the average of a fixed number of consecutive data points.
2. Plot average on graph.

Weighted Moving Average

1. Calculate the weighted average of a fixed number of consecutive data points.
2. Plot average on graph.

5. Least Squares Method

Least Squares Method

1. Plot data points on a graph.
2. Draw a best-fitting line through the data points using the least squares method.


equipping with excellence
CPA COLLEGE OF GLOBAL STUDIES