QUANTITATIVE TECHNIQUES

Quantitative techniques may be defined as those techniques which provide the decision makes a systematic and powerful means of analysis, based on quantitative data. It is a scientific method employed for problem solving and decision making by the management. With the help of quantitative techniques, the decision maker is able to explore policies for attaining the predetermined objectives. In short, quantitative techniques are inevitable in decision-making process.

Classification of Quantitative Techniques:

There are different types of quantitative techniques. We can classify them into three categories. They are:

1. Mathematical Quantitative Techniques
2. Statistical Quantitative Techniques
3. Programming Quantitative Techniques

Mathematical Quantitative Techniques:

A technique in which quantitative data are used along with the principles of mathematics is known as mathematical quantitative techniques. Mathematical quantitative techniques involve:

1. Permutations and Combinations: Permutation means arrangement of objects in a definite order. The number of arrangements depends upon the total number of objects and the number of objects taken at a time for arrangement. Combination means selection or grouping objects without considering their order.

2. Set Theory:- Set theory is a modern mathematical device which solves various types of critical problems.

3. Matrix Algebra: Matrix is an orderly arrangement of certain given numbers or symbols in rows and columns. It is a mathematical device of finding out the results of different types of algebraic operations on the basis of the relevant matrices.

4. Determinants: It is a powerful device developed over the matrix algebra. This device is used for finding out values of different variables connected with a number of simultaneous equations.

5. Differentiation: It is a mathematical process of finding out changes in the dependent variable with reference to a small change in the independent variable.

6. Integration: Integration is the reverse process of differentiation.

7.Differential Equation: It is a mathematical equation which involves the differential coefficients of the dependent variables.

Statistical Quantitative Techniques:

Statistical techniques are those techniques which are used in conducting the statistical enquiry concerning to certain Phenomenon. They include all the statistical methods beginning from the collection of data till interpretation of those collected data. Statistical techniques involve:

1. Collection of data: One of the important statistical methods is collection of data. There are different methods for collecting primary and secondary data.

2. Measures of Central tendency, dispersion, skewness and Kurtosis: Measures of Central tendency is a method used for finding he average of a series while measures of dispersion used for finding out the variability in a series. Measures of Skewness measures asymmetry of a distribution while measures of Kurtosis measures the flatness of peakedness in a distribution.

3. Correlation and Regression Analysis: Correlation is used to study the degree of relationship among two or more variables. On the other hand, regression technique is used to estimate the value of one variable for a given value of another.

4. Index Numbers: Index numbers measure the fluctuations in various Phenomena like price, production etc over a period of time, They are described as economic barometers.

5. Time series Analysis: Analysis of time series helps us to know the effect of factors which are responsible for changes:

6. Interpolation and Extrapolation: Interpolation is the statistical technique of estimating under certain assumptions, the missing figures which may fall within the range of given figures. Extrapolation provides estimated figures outside the range of given data.

7. Statistical Quality Control Statistical quality control is used for ensuring the quality of items manufactured. The variations in quality because of assignable causes and chance causes can be known with the help of this tool. Different control charts are used in controlling the quality of products.

8. Ratio Analysis: Ratio analysis is used for analyzing financial statements of any business or industrial concerns which help to take appropriate decisions.

9. Probability Theory: Theory of probability provides numerical values of the likely hood of the occurrence of events.

10. Testing of Hypothesis Testing of hypothesis is an important statistical tool to judge the reliability of inferences drawn on the basis of sample studies.

Programming Techniques:

Programming techniques are also called operations research techniques. Programming techniques are model building techniques used by decision makers in modern times. Programming techniques involve:

1. **Linear Programming**: Linear programming technique is used in finding a solution for optimizing a given objective under certain constraints.
2. **Queuing Theory**: Queuing theory deals with mathematical study of queues. It aims at minimizing cost of both servicing and waiting.
3. **Game Theory**: Game theory is used to determine the optimum strategy in a competitive situation.
4. **Decision Theory**: This is concerned with making sound decisions under conditions of certainty, risk and uncertainty.
5. **Inventory Theory**: Inventory theory helps for optimizing the inventory levels. It focuses on minimizing cost associated with holding of inventories.
6. **Network programming**: It is a technique of planning, scheduling, controlling, monitoring and co-coordinating large and complex projects comprising of a number of activities and events. It serves as an instrument in resource allocation and adjustment of time and cost up to the optimum level. It includes CPM, PERT etc.
7. **Simulation**: It is a technique of testing a model which resembles a real life situations
8. **Replacement Theory**: It is concerned with the problems of replacement of machines, etc due to their deteriorating efficiency or breakdown. It helps to determine the most economic replacement policy.
9. **Non Linear Programming:** It is a programming technique which involves finding an optimum solution to a problem in which some or all variables are non-linear.
10. **Sequencing**: Sequencing tool is used to determine a sequence in which given jobs should be performed by minimizing the total efforts.
11. **Quadratic Programming**: Quadratic programming technique is designed to solve certain problems, the objective function of which takes the form of a quadratic equation.

12. Branch and Bound Technique : This is designed to solve the combinational problems of decision making where there are large number of feasible solutions. Problems of plant location, problems of determining minimum cost of production etc. are examples of combinational problems.

Functions of Quantitative Techniques:

The following are the important functions of quantitative techniques:

1. To facilitate the decision-making process
2. To provide tools for scientific research
3. To help in choosing an optimal strategy
4. To enable in proper deployment of resources

5. To help in minimizing costs
6. To help in minimizing the total processing time required for performing a set of jobs
7. helping in capital resource allocation

Limitations of Quantitative Techniques:

1. Quantitative techniques involves mathematical models, equations and other mathematical expressions
2. Quantitative techniques are based on number of assumptions.
3. Quantitative techniques are very expensive.
4. Quantitative techniques do not take into consideration intangible facts like skill, attitude etc.
5. Quantitative techniques are only tools for analysis and decision-making. They are not decisions itself.

CORRELEATION 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. The analysis of such data is called univariate analysis. The data concerned with two variables are called bivariate data. The analysis of these two sets of data is called bivariate analysis. The date concerned with three or more variables are called multivariate date.

The statistical technique which can be used to study the relationship between two or more variables is called correlation analysis.

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".

Lun chou defines, " Correlation analysis attempts to determine the degree of relationship between variables".

Boddington states that "Whenever some definite connection exists between two or more groups or classes of series of data, there is said to be correlation."

Classification of Correlation

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

1. Positive and Negative correlation
2. Simple, partial and multiple correlation
3. Linear and Non-linear correlation

Positive and Negative correlation

Positive Correlation: 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.

Negative Correlation: 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.

Multiple correlation In a correlation analysis, if three or more variables are studied simultaneously, it is called 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.

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.

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 non-linear correlation.

Degrees of correlation:

Correlation exists in various degrees

1. Perfect positive correlation: 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.

2. Perfect Negative correlation: 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.

3. Limited Degree of Positive correlation:: 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 positive correlation.

4. Limited degree of Negative correlation: 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.

5. Zero Correlation/ no correlation: 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

1) Scatter Diagram
2) Correlation graph
3) coefficient of correlation

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

Merits of Scatter Diagram method

1.it is easy to plot the points
2. simple to understand
3. abnormal values in the data can be easily detected
4. extreme value do not affect it.

Demerits of Scatter diagram method

1.the degree of correlation cannot be easily estimated
2. algebraic treatment is not possible

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 be negative.

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. Coefficient of correlation can be computed by applying the methods given below

1) Karl Pearson's Co-efficient of correlation
2) Spearman's Rank correlation method
3) Concurrent deviation method

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.

Pearson's coefficient of correlation is defined as the ratio of the covariance between X and Y to the product of their standard deviations. This is denoted by 'r'

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 between variables.
2. When r = -1, it means there is perfect negative relationship between variables.
3. When r = 0, it means there is no relationship between the variables.
4. When 'r' is closer to +1, it means there is high degree of positive correlation between variables.
5. When 'r' is closer to – 1, it means there is high degree of negative correlation between variables.
6. When 'r' is closer to 'O', it means there is less relationship between variables.

Properties of Co-efficient of Correlation

1.correlation coefficient has a well-defined formula
2. correlation coefficient is a pure number and is independent.
3. it lies between -1 and + 1.
4. correlation coefficient do not change with reference to change.
5. coefficient of correlation between X and Y is same as that between Y and X.

Computation of Pearson's Co-efficient of correlation:

Pearson's correlation co-efficient can be computed in different ways. They are:

a Arithmetic mean method
b Assumed mean method
c Direct method

Probable Error

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.

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.

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. The correlation coefficient obtained from ranks of the variables instead of their quantitative measurement is called rank correlation. This was developed by Charles Edward Spearman in 1904.

Spearman's coefficient correlation (R) = $1-\frac{6\sum D^2}{n(n^2-1)}$

Where D = difference of ranks between the two variables

N = number of pairs

Computation of Rank Correlation Coefficient when Ranks are Equal

There may be chances of obtaining same rank for two or more items. In such a situation, it is required to give average rank for all. Such items.

Merits of Rank Correlation method

1.it is easy to calculate
2. simple to understand
3. can be applied both quantitative and qualitative data.

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.

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: -

$$r = \sqrt{\pm\left(\frac{2C-N}{N}\right)}$$

Where N = No. of pairs of symbol

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

Co-efficient of determination

co-efficient determination gives the percentage variation in the dependent variable in relation with the independent variable. In other words co-efficient of determination gives the ratio of the explained variable to the total variance. The co-efficient of determination is the square of the correlation coefficient.

REGRESSION ANALYSIS

Introduction:-

Regression analysis means to analyse the average relationship between two variables and thereby provides a mechanism for estimation or predication or forecasting. 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:

According to M.M Blair "Regression is the measure of the average relationship between two or more variables in terms of the original units of the date".

Types of Regression:-

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

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 of correlation.

Properties of Regression lines:-

1. The two regression lines cut each other at the point of average of X and average of Y
2. When r = 1, the two regression lines coincide each other and give one line.
3. When r = 0, the two regression lines are mutually perpendicular.

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 in Y.

here equation is $X = a + by$

$\sum x = na + b \sum y$

$\sum xy = a\sum y + b \sum y^2$

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

here equation is $Y = a + bx$

$\sum y = na + b \sum x$

$\sum xy = a\sum x + b \sum x^2$

Difference between correlation and regression

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

THEORY OF PROBABILITY

INTRODUCTION

Probability refers to the chance of happening or not happening of an event. Any problem which contains uncertainty about the happening of the event is the problem of probability.

Definition of Probability

The probability of given event may be defined as the numerical value given to the likely hood of the occurrence of that event. It. is a number lying between '0' and '1' '0' denotes the even which cannot occur, and '1' denotes the event which is certain to occur

Terms use in Probability.

Random Experiment: A random experiment is an experiment that has two or more outcomes which vary in an unpredictable manner from trial to trail when conducted under uniform conditions. In a random experiment, all the possible outcomes are known in advance but none of the outcomes can be predicted with certainty. For example, tossing of a coin is a random experiment because it has two outcomes (head and tail), but we cannot predict any of them which certainty.

Sample Point: Every indecomposable outcome of a random experiment is called a sample point. It is also called simple event or elementary outcome. Eg. When a die is thrown, getting '3' is a sample point.

Sample space: Sample space of a random experiment is the set containing all the sample points of that random experiment. Eg:- When a coin is tossed, the sample space is (Head, Tail)

Event: An event is the result of a random experiment. It is a subset of the sample space of a
random experiment.

Sure Event (Certain Event): An event whose occurrence is inevitable is called sure even. Eg:- Getting a white ball from a box containing all while balls.

Impossible Events: An event whose occurrence is impossible, is called impossible event. Eg:- Getting a white ball from a box containing all red balls.

Uncertain Events: An event whose occurrence is neither sure nor impossible is called uncertain event. Eg:- Getting a white ball from a box containing white balls and black balls.

Equally likely Events: Two events are said to be equally likely if anyone of them cannot be expected to occur in preference to other. For example, getting herd and getting tail

when a coin is tossed are equally likely events.

Mutually exclusive events: A set of events are said to be mutually exclusive of the occurrence of one of them excludes the possibility of the occurrence of the others.

Exhaustive Events:: A group of events is said to be exhaustive when it includes all possible outcomes of the random experiment under consideration.

Dependent Events: Two or more events are said to be dependent if the happening of one of them affects the happening of the other.

DIFFERENT SCHOOLS OF THOUGHT ON PROBABILITY

Different Approaches/Definitions of Probability

There are 4 important schools of thought on probability :-

1. Classical or Priori Approach Objective Probability
2. Relative frequency or Empirical Approach
3. Modern or Axiomatic Approach

1. Classical or Priori Approach

If out of '*n*' exhaustive, mutually exclusive and equally likely outcomes of an experiment; '*m*' are favourable to the occurrence of an event 'A', then the probability of 'A' is defined as to be $\mathbf{P(A) = \frac{f}{n}}$

Frequency ratio : consider a random experiment. Let 'A' be an event associated with the random experiment. Let us repeat the experiment 'n' times. Let the event A happen 'f ' out of 'n' repetition of the experiment. Then f/n is called frequency ratio.

Law of statistical regularity : the outcomes of a random experiment vary in an unpredictable manner. But if we consider sequence of frequency ratios, we can observe a regularity. We notice that these frequency ratios have a tendency to become more and more stable as the number of repetitions of the experiment increases. The frequency ratio appears to approach a fixed number as the number repetitions is increased. This tendency of frequency ratio to stabilize, is common in all random experiments. This phenomenon is known as statistical regularity.

Relative Frequency Definition or Empirical Approach

According to Relative Frequency definition, the probability of an event can be defined as the relative frequency with which it occurs in an indefinitely large number of trials.

Here, probability has between 0 and 1, i.e. $0 \le P(A) \le 1$

Axiomatic Approach (Modern Approach) to Probability

Let 'S' be the sample space of a random experiment, and 'A' be an event of the random experiment, so that 'A' is the subset of 'S'. Then we can associate a real number to the event 'A'. This number will be called probability of 'A' if it satisfies the following three axioms or postulates :-

1.P (A) is a real number such that $P(A) \ge 0$ for every subset of S

2. P (S) = 1 where 'S' is the sample space

3. $P(A \cup B) = P(A) + P(B)$, where A and B are two non-intersecting subset of S

THEOREMS OF PROBABILITY

There are two important theorems of probability. They are :

1. Addition Theorem
2. Multiplication Theorem

Addition Theorem

Here, there are 2 situations.

(a) Events are mutually exclusive

(b) Events are not mutually exclusive

(a) Addition theorem (Mutually Exclusive Events): If two events, 'A' and 'B', are mutually exclusive the probability of the occurrence of either 'A' or 'B' is the sum of the individual probability of A and B.

P(A or B) = P(A) + P(B)

i.e., $P(A \cup B) = P(A) + P(B)$

(b)Addition theorem (Not mutually exclusive events): If two events, A and B are not mutually exclusive the probability of the occurrence of either A or B is the sum of their individual probability minus

probability for both to happen.
P(A or B) = P(A) + P(B) – P(A and B)
i.e., P(A∪B) = P(A) + P(B) – P(A∩B)

MULTIPLICATION THEOREM

Here there are two situations:
(a) Events are independent
(b) Events are dependent

(a)Multiplication theorem (independent events): If two events are independent, then the probability of occurring both will be the product of the individual probability
P(A and B) = P(A) X P(B)
i.e., P(A∩B) = P(A) X P(B)

(b)Multiplication theorem (dependent Events):-If two events, A and B are dependent, the probability of occurring 2nd event will be affected by the outcome of the first.
P(A∩B) = P(A) X P(B/A)

CONDITIONAL PROBABILITY

Multiplication theorem states that if two events, A and B, are dependent events then, the probability of happening both will be the product of P(A) and P(B/A).
i.e., P(A and B) or P(A ∩ B) = P(A) X P(B/A)
Here, P (B/A) is called Conditional probability

Inverse Probability

If an event has happened as a result of several causes, then we may be interested to find out the probability of a particular cause of happening that events. This type of problem is called inverse probability. Baye's theorem is based upon inverse probability.

BAYE'S THEOREM:

Baye's theorem is based on the proposition that probabilities should revised on the basis of all the available information. The revision of probabilities based on available information will help to reduce the risk involved in decision-making. The probabilities before revision is called priori probabilities and the probabilities after revision is called posterior probabilities.

PROBABILITY DISTRIBUTION

DEFINITION

Probability distribution (Theoretical Distribution) can be defined as a distribution obtained for a random variable on the basis of a mathematical model. It is obtained not on the basis of actual observation or experiments, but on the basis of probability law.

Random variable

Random variable is a variable who value is determined by the outcome of a random experiment. Random variable is also called chance variable or stochastic variable.

Classification of Probability Distribution

Discrete probability Distribution
Continuous Probability Distribution
Binomial Distribution
Poisson distribution
Normal Distribution

Discrete Probability Distribution: If the random variable of a probability distribution assumes specific values only, it is called discrete probability distributions. Binomial distribution and poisson distribution are discrete probability distributions.

Continuous Probability Distributions:- If the random variable of a probability distribution assumes any value in a given interval, then it is called continuous probability distributions. Normal distributions is a continuous probability distribution.

BIONOMIAL DISTRIBUTION

Meaning & Definition:

Binomial Distribution is associated with James Bernoulli, a Swiss Mathematician. Therefore, it is also called Bernoulli distribution. Binomial distribution is the probability distribution expressing the probability of one set of dichotomous alternatives, i.e., success or failure. In other words, it is used to determine the probability of success in experiments on which there are only two mutually exclusive outcomes. Binomial distribution is discrete probability distribution.

Binomial Distribution can be defined as follows: "A random variable x is said to follow Binomial Distribution with

parameters n and p if its probability function is:

$$f(x) = nC_x\, p^x q^{n-x}$$

Where, P = probability of success in a single trial

q = 1 – p

n = number of trials

x = number of success in 'n' trials.

Assumption of Binomial Distribution OR (Situations where Binomial Distribution can be applied)

Binomial distribution can be applied when:-

1. The random experiment has two outcomes i.e., success and failure.
2. The probability of success in a single trial remains constant from trial to trial of the experiment.
3. The experiment is repeated for finite number of times.
4. The trials are independent.

Properties (features) of Binomial Distribution:

1. It is a discrete probability distribution.
2. The shape and location of Binomial distribution changes as 'p' changes for a given 'n'.
3. The mode of the Binomial distribution is equal to the value of 'r' which has the largest probability.
4. Mean of the Binomial distribution increases as 'n' increases with 'p' remaining constant.
5. The mean of Binomial distribution is np.
6. The Standard deviation of Binomial distribution is $\sqrt{npq}$
7. If 'n' is large and if neither 'p' nor 'q' is too close zero, Binomial distribution may be approximated to Normal Distribution.
8. If two independent random variables follow Binomial distribution, their sum also follows Binomial distribution.

Mean and Standard Deviation of Binomial Distribution

Mean of Binomial Distribution = np

Standard Deviation of Binomial Distribution = $\sqrt{npq}$

Fitting a Binomial Distribution

Steps:

1. Find the value of n, p and q
2. Substitute the values of n, p and q in the Binomial Distribution function of $f(x) = nC_x\, p^x q^{n-x}$
3. Put x = 0, 1, 2, ……. in the function $f(x) = nC_x\, p^x q^{n-x}$
4. Multiply each such terms by total frequency (N) to obtain the expected frequency.

POISSON DISTRIBUTION

Meaning and Definition:

Poisson Distribution is a limiting form of Binomial Distribution. In Binomial Distribution, the total number of trials are known previously. But in certain real life situations, it may be impossible to count the total number of times a particular event occurs or does not occur. In such cases Poisson Distribution is more suitable. Poison Distribution is a discrete probability distribution. It was originated by Simeon Denis Poisson. The Poisson Distribution is defined as:-

$$p(x) = \frac{e^{-m} m^x}{X!}$$

Where x = random variable (i.e., number of success in '**n**' trials.

e = 2.7183

m = mean of Poisson distribution

Properties of Poisson Distribution

1. Poisson Distribution is a discrete probability distribution.
2. Poisson Distribution has a single parameter 'm'. When 'm' is known all the terms can be found out.
3. It is a positively skewed distribution.
4. Mean and Variance of Poisson Distribution are equal to 'm'.
5. In Poisson Distribution, the number of success is relatively small.
6. The standard deviation of Poisson Distribution is $\sqrt{m}$.

Practical situations where Poisson Distribution can be used

1. To count the number of telephone calls arising at a telephone switch board in a unit of time.
2. To count the number of customers arising at the super market in a unit of time.
3. To count the number of defects in Statistical Quality Control.
4. To count the number of bacteria's per unit.
5. To count the number of defectives in a park of manufactured goods.
6. To count the number of persons dying due to heart attack in a year.
7. To count the number of accidents taking place in a day on a busy road.

NORMAL DISTRIBUTION

The normal distribution is a continuous probability distribution. It was first developed by De-Moivre in 1733 as limiting form of binomial distribution. Fundamental importance of normal distribution is that many populations seem to follow approximately a pattern of distribution as described by normal distribution. Numerous phenomena such as the age distribution of any species, height of adult persons, intelligent test scores of students, etc. are considered to be normally distributed.

Definition of Normal Distribution

A continuous random variable 'X' is said to follow Normal Distribution if its probability function is:

$$P(X) = \frac{1}{\sqrt{2\pi}\,\sigma} \times e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

π = 3.146
e = 2.71828
μ = mean of the distribution
σ = standard deviation of the distribution

Properties of Normal Distribution

1. Normal distribution is a continuous distribution.
2. Normal curve is symmetrical about the mean.
3. Both sides of normal curve coincide exactly.
4. Normal curve is a bell shaped curve.
5. Mean, Median and Mode coincide at the centre of the curve.
6. Quantities are equi-distant from median. Q3 – Q2 = Q2 – Q1
7. Normal curve is asymptotic to the base line.
8. Total area under a normal curve is 100%.
9. The ordinate at the mean divide the whole area under a normal curve into two equal parts. (50% on either side).
10. The height of normal curve is at its maximum at the mean.
11. The normal curve is unimodel, i.e., it has only one mode.
12. Normal curve is mesokurtic.
13. No portion of normal curve lies below the x-axis.
14. Theoretically, the range of normal curve is – α to + α . But practically the range is

Fitting of a Normal Distribution

Procedure :

1. Find the mean and standard deviation of the given distribution. (i.e., μ and σ)
2. Take the lower limit of each class.
3. Find Z value for each of the lower limit. $Z = \frac{X-\mu}{\sigma}$
4. Find the area for z values from the table. The first and the last values are taken as 0.5.
5. Find the area for each class. Take difference between 2 adjacent values if same signs and take total of adjacent values if opposite signs.
6. Find the expected frequency by multiplying area for each class by N.

TESTING OF HYPOTHESIS

Statistical Inference:

Statistical inference refers to the process of selecting and using a sample statistic to draw conclusions about the population parameter. Statistical inference deals with two types of problems.
They are:-
1. Testing of Hypothesis
2. Estimation

Hypothesis: Hypothesis is a statement subject to verification. More precisely, it is a quantitative statement about a population, the validity of which remains to be tested. In other words, hypothesis is an assumption made about a population parameter.

Testing of Hypothesis: Testing of hypothesis is a process of examining whether the hypothesis formulated by the researcher is valid or not. The main objective of hypothesis testing is whether to accept or reject the hypothesis.

Procedure for Testing of Hypothesis:

The various steps in testing of hypothesis involves the following :-

1. **Set Up a Hypothesis:**: The first step in testing of hypothesis is to set up a hypothesis about population parameter. Normally, the researcher has to fix two types of hypothesis. They are null hypothesis and alternative hypothesis.

Null Hypothesis:-Null hypothesis is the original hypothesis. It states that there is no significant difference between the sample and population regarding a particular matter under consideration. The word "null" means 'invalid' of 'void' or 'amounting to nothing'. Null hypothesis is denoted by Ho.

Alternative Hypothesis:-Any hypothesis other than null hypothesis is called alternative hypothesis. When a null hypothesis is rejected, we accept the other hypothesis, known as alternative hypothesis. Alternative hypothesis is denoted by H1.

2. **Set up a suitable level of significance**: After setting up the hypothesis, the researcher has to set up a suitable level of significance. The level of significance is the probability with which we may reject a null hypothesis when it is true. For example, if level of significance is 5%, it means that in the long run, the researcher is rejecting true null hypothesis 5 times out of every 100 times. Level of significance is denoted by α (alpha).

α = Probability of rejecting H0 when it is true.

Generally, the level of significance is fixed at 1% or 5%.

3. **Decide a test criterion**: The third step in testing of hypothesis is to select an appropriate test criterion. Commonly used tests are z-test, t-test, X2 – test, F-test, etc.

4. **Calculation of test statistic**: The next step is to calculate the value of the test statistic using appropriate formula. The general form for computing the value of test statistic is:-

Value of Test statistic = Difference Standard Error

5. **Making Decision**: Finally, we may draw conclusions and take decisions. The decision may be either to accept or reject the null hypothesis. If the calculated value is more than the table value, we reject the null hypothesis and accept the alternative hypothesis. If the calculated value is less than the table value, we accept the null hypothesis.

Sampling Distribution

The distribution of all possible values which can be assumed by some statistic, computed from samples of the same size randomly drawn from the same population is called Sampling distribution of that statistic.

Standard Error (S.E)

Standard Error is the standard deviation of the sampling distribution of a statistic. Standard error plays a very important role in the large sample theory. The following are the important **uses of standard errors:-**

1. Standard Error is used for testing a given hypothesis
2. S.E. gives an idea about the reliability of a sample, because the reciprocal of S.E. is a measure of reliability of the sample.
3. S.E. can be used to determine the confidence limits within which the population parameters are expected to lie.

Test Statistic

The decision to accept or to reject a null hypothesis is made on the basis of a statistic computed from the sample. Such a statistic is called the test statistic. There are different

types of test statistics. All these test statistics can be classified into two groups. They are
a. Parametric Tests
b. Non-Parametric Tests

PARAMETRIC TESTS

The statistical tests based on the assumption that population or population parameter is normally distributed are called parametric tests. The important parametric tests are:-
1. z-test
2. t-test
3. f-test

Z-test:

Z-test is applied when the test statistic follows normal distribution. It was developed by Prof.R.A.Fisher. The following are the important **uses of z-test:-**
1. To test the population mean when the sample is large or when the population standard deviation is known.
2. To test the equality of two sample means when the samples are large or when the population standard deviation is known.
3. To test the population proportion.
4. To test the equality of two sample proportions.
5. To test the population standard deviation when the sample is large.
6. To test the equality of two sample standard deviations when the samples are large or when population standard deviations are known.
7. To test the equality of correlation coefficients.

Z-test is used in testing of hypothesis on the basis of some assumptions. The important assumptions in z-test are:-
1. Sampling distribution of test statistic is normal.
2. Sample statistics are dose the population parameter and therefore, for finding standard error, sample statistics are used in place where population parameters are to be used.

T-test:

t-distribution was originated by W.S.Gosset in the early 1900. t-test is applied when the test statistic follows t-distribution. **Uses of t-test are:-**
1. To test the population mean when the sample is small and the population s.D.is unknown.
2. To test the equality of two sample means when the samples are small and population S.D. is unknown.
3. To test the difference in values of two dependent samples.
4. To test the significance of correlation coefficients.

The following are the important assumptions in t-test:-
1. The population from which the sample drawn is normal.
2. The sample observations are independent.
3. The population S.D.is known.
4. When the equality of two population means is tested, the samples are assumed to be independent and the population variance are assumed to be equal and unknown.

F-test:

F-test is used to determine whether two independent estimates of population variance significantly differ or to establish both have come from the same population. For carrying out the test of significance, we calculate a ration, called F-ratio. F-test is named in honour of the great statistician R.A.Fisher. It is also called Variance Ration Test.

Uses of F-distribution:-
1. To test the equality of variances of two populations.
2. To test the equality of means of three or more populations.
3. To test the linearity of regression

Assumptions of F-distribution:-
1. The values in each group are normally distributed.
2. The variance within each group should be equal for all groups.
3. The error (Variation of each value around its own group mean) should be independent for each value.

TYPES OF ERRORS IN TESTING OF HYPOTHESIS:

In any test of hypothesis, the decision is to accept or reject a null hypothesis. The four possibilities of the decision are:-

1. Accepting a null hypothesis when it is true.
2. Rejecting a null hypothesis when it is false.
3. Rejecting a null hypothesis when it is true.
4. Accepting a null hypothesis when it is false.

Out of the above 4 possibilities, 1 and 2 are correct, while 3 and 4 are errors. The error included in the above 3rd possibility is called type I error and that in the 4th possibility is called type II error.

Type I Error

The error committed by rejecting a null hypothesis when it is true, is called Type I error. The probability of committing Type I error is denoted by α (alpha).

α = Prob. (Type I error)
= Prob. (Rejecting H0 when it is true)

Type II Error

The error committed by accepting a null hypothesis when it is false is called Type II error. The probability of committing Type II error is denoted by β (beta).

β = Prob. (Type II error)
β = Prob. (Accepting H0 when it is false)

Small and Large samples

The size of sample is 30 or less than 30, the sample is called small sample. When the size of sample exceeds 30, the sample is called large sample.

Degree of freedom

Degree of freedom is defined as the number of independent observations which is obtained by subtracting the number of constraints from the total number of observations. Degree of freedom = Total number of observations – Number of constraints.

Rejection region and Acceptance region

The entire area under a normal curve may be divided into two parts. They are rejection region and acceptance region.

Rejection Region: Rejection region is the area which corresponds to the predetermined level of significance. If the calculated value of the test statistic falls in the rejection region, we reject the null hypothesis. Rejection region is also called critical region. It is denoted by α (alpha).

Acceptance Region: Acceptance region is the area which corresponds to 1 – α.

Acceptance region = 1 – rejection region = 1-α.

If the calculated value of the test statistic falls in the acceptance region, we accept the null hypothesis.

TWO TAILED AND ONE TAILED TESTS:

A two tailed test is one in which we reject the null hypothesis if the computed value of the test statistic is significantly greater or lower than the critical value (table value) of the test statistic. Thus, in two tailed test the critical region is represented by both tails of the normal curve. If we are testing hypothesis at 5 % level of significance, the size of the acceptance region is 0.95 and the size of the rejection region is 0.05 on both sides together. (i.e. 0.025 on left side and 0.025 on right side of the curve).

TESTING OF EQUALITY OF TWO SAMPLE MEANS

This test is used to test whether there is significant difference between two sample means. It there is no significant difference, we can consider the samples are drawn from the same population.

Procedure:

1. Set up null hypothesis that there is no significant difference between the two means.

H0 : $\mu_1 = \mu_2$
H1 : $\mu_1 \neq \mu_2$

2. Decide the test criterion:

- If sample is large, apply z – test
- If sample is small, but population S.D. is known, apply z-test.
- If sample is small and population S.D. is unknown, apply t-test.

3. Apply the formula:

Z or t = $\frac{X-\mu}{\sigma}$

4. Fix the degree of freedom:
5. Obtain the table value.
6. Decide whether to accept or reject the H0.

TESTING OF EQUALITY OF TWO SAMPLE STANDARD DEVIATIONS

This test is used to test whether there is any significant difference between the standard between the standard deviation of two samples.

Procedure:

1. Set the null hypothesis that there is no significant difference between two standard deviations.
2. Decide the test criterion:
If sample is large, apply Z – test
If sample is sample, apply F – test
3. Apply the formula:
4. Fix the degree of freedom
5. Obtain the table value.
6. Decide whether to accept or reject the null hypothesis.

NON-PARAMETRIC TESTS

A non-parametric test is a test which is not concerned with testing of parameters. Nonparametric tests do not make any assumption regarding the form of the population. Therefore, non-parametric tests are also called distribution free tests.

Following are the important **non-parametric tests:-**

1. Chi-square test (χ2 – test)
2. Sign test
3. Signed rank test (Wilcoxon matched pairs test)
4. Rank sum test (Mann-whitney U-test and Kruskal-Wallis H test)

CHI-SQUARE TEST

The value of chi-square describes the magnitude of difference between observed frequencies and expected frequencies under certain assumptions. X^2 value (χ2 quantity) ranges from zero to infinity. It is zero when the expected frequencies and observed frequencies completely coincide. So greater the value of χ2, greater is the discrepancy between observed and expected frequencies. χ2-test is a statistical test which tests the significance of difference between observed frequencies and corresponding theoretical frequencies of a distribution without any assumption about the distribution of the population. This is one of the simplest and most widely used nonparametric test in statistical work. This test was developed by Prof. Karl Pearson in 1990.

Uses of χ2 - test

The uses of chi-square test are:-

1. Useful for the test of goodness of fit:- χ2 - test can be used to test whether there is goodness of fit between the observed frequencies and expected frequencies.
2. Useful for the test of independence of attributes:- χ2 test can be used to test whether two attributes are associated or not.
3. Useful for the test of homogeneity:- χ2 - test is very useful t5o test whether two attributes are homogeneous or not.
4. Useful for testing given population variance:- χ2-test can be used for testing whether the given population variance is acceptable on the basis of samples drawn from that population.

χ2 -test as a test of goodness of fit:

As a non-parametric test, χ2-test is mainly used to test the goodness of fit between the observed frequencies and expected frequencies.

Procedure:-

1. Set up mull hypothesis that there is goodness of fit between observed and expected frequencies.
2. Find the χ2 value using the following formula:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Where O = Observed frequencies E = Expected frequencies

3. Compute the degree of freedom. d. f. = n – r – 1

Where 'r' is the number of independent constraints to be satisfied by the frequencies

4. Obtain the table value corresponding to the lord of significance and degrees of freedom.
5. Decide whether to accept or reject the null hypothesis. If the calculated value is less than the table value, we accept the null hypothesis and conclude that there is goodness of fit. If the calculated value is more than the table value we reject the null hypothesis and conclude that there is no goodness of fit.

χ2 - test as a test of independence:

χ2 – test is used to find out whether one or more attributes are associated or not.

Procedure:-

1. Set up null and alternative hypothesis.
H0: Two attributes are independent (i.e., there is no association between the attributes)
H1: Two attributes are dependent (i.e., there is an association between the attributes)
2. Find the χ2 value.

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

3. Find the degree of freedom d.f. = (r-1)(c-1)
Where r = Number of rows
c = Number of columns
4. Obtain table value corresponding to the level of significance and degree of freedom.
5. Describe whether to accept or reject the H0 . If the calculated value is less than the table value, we accept the H0 and conclude that the attributes are independent.

χ2 – test as a test of homogeneity

χ2 – test is used to find whether the samples are homogeneous as far as a particular attribute is concerned.

Steps:

1. Set up null and alternative hypotheses:
H0: There is homogeneity.
H1: There is no homogeneity (heterogeneity)
2. Find the χ2 value.

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

3. Find the degree of freedom d.f. = (r-1)(c-1)
4. Obtain the table value
5. Decide whether to accept or reject the null hypothesis.

χ2 – test for Population Variance:

χ2 – test can be used for testing the given population when the sample is small.

Steps:-

1. Set up null and alternative hypotheses:
H0: There is no significant difference between sample variance and population variance.
H0: There is significant difference between sample variance and population variance.
2. Find the χ2 value.

$$\chi^2 = \frac{ns^2}{\sigma^2}$$

Where S^2 = Sample variance , $\boldsymbol{\sigma 2}$ = Population variance
3. Find the degree of freedom d.f. = n-1
4. Obtain the table value
5. Decide whether to accept or reject the null hypothesis

Limitations of Chi-square tests:-

1. It is not as reliable as a parametric test. Hence it should be used only when parametric tests cannot be used.
2. χ2 value cannot be computed when the given values are proportions or percentages.

wilcoxon matched pairs test (signed rank test)

Signed rank test was developed by Frank Wilcoxon. It is an important non-parametric test. This method is used when we can determine both direction and magnitude of difference between matched values.
Here there are two cases:-
a) When the number of matched pairs are less than or equal to 25.
b) When the number of matched pairs are more than 25.

ANALYSIS OF VARIANCE

Definition of Analysis of Variance

Analysis of variance may be defined as a technique which analyses the variance of two or more comparable series (or samples)

for determining the significance of differences in their arithmetic means and for determining whether different samples under study are drawn from same population or not, with the of the statistical technique, called F – test.

Characteristics of Analysis of Variance:

1. It makes statistical analysis of variance of two or more samples.

2. It tests whether the difference in the means of different sample is due to chance or due to any significance cause.

3. It uses the statistical test called, F – Ratio.

Types of Variance Analysis:

There are two types of variance analysis. They are:-

1. One way Analysis of Variance
2. Two way analysis of Variance

One way Analysis of Variance:

In one way analysis of variance, observations are classified into groups on the basis of a single criterion. For example, yield of a crop is influenced by quality of soil, availability of rainfall, quantity of seed, use of fertilizer, etc. It we study the influence of one factor, It is called one way analysis of variance.

If we want to study the effect of fertilizer of yield of crop, we apply different kinds of fertilizers on different paddy fields and try to find out the difference in the effect of these different kinds of fertilizers on yield.

two way analysis of variance

Two way analysis of variance is used to test the effect of two factors simultaneously on a particular variable.

Study well...

MUHAMMED RIYAS N
ASST. PROF.
AL JAMIA ARTS AND SCIENCE COLLEGE
POOPALAM
PERINTHALMANNA
PH : 9747799772
E-mail- riyasmuhammed89@gmail.com