MADURAI KAMARAJ UNIVERSITY

(University with Potential for Excellence)

DIRECTORATE OF DISTANCE EDUCATION

M.B.A

AIRLINE & AIRPORT MANAGEMENT

QUANTITATIVE METHODS FOR MANAGEMENT

www.mkudde.org

6037

PMAM04

MADURAI KAMARAJ UNIVERSITY

(University with Potential for Excellence)
DIRECTORATE OF DISTANCE EDUCATION
Palkalai Nagar, Madurai – 625 021, India
Ph : 0452 – 2458471 (30 Lines)   Fax : 0452 – 2458265

E-mail	: mkudde@mkudde.org
General grievances	: mkudde grievances@gmail.com
UG Courses	: mkuddeug@gmail.com
PG Courses	: mkupg@gmail.com
MBA Courses	: mkuddembag@gmail.com
MCA Courses	: mkuddemcag@gmail.com
Education Courses	: mkuddeedu@gmail.com
Website	: www.mkudde.orgI
VRS	: 0452 – 2459990
	: 0452 - 2459596

Student Support Service : 0452 – 2458419

DDE – Examinations	
Fax	: 0452 - 2458261
E-mail	: mkuace@yahoo.com
Examn. Grievance Redress Cell	: 0452 – 2458471 – Extn. 515

Reading Material Disclaimer

This reading material, developed by **Dr.c.THangapandi, Associate Professor, Department of Mathematics, Madura College** is an aid for the students of Directorate of Distance Education, Madurai Kamaraj University, to understand the course content. It is only for the Registered Students of DDE, MKU and is not for Private Circulation.

MADURAI KAMARAJ UNIVERSITY

(University with Potential for Excellence)

Madurai – 625 021.

DIRECTORATE OF DISTANCE EDUCATION

M.B.A

AIRLINE & AIRPORT MANAGEMENT

FIRST YEAR

I - SEMESTER

QUANTITATIVE METHODS FOR MANAGEMENT

www.mkudde.org

6037

PMAM04

Printed at Vimala Note Book

Copies-1000 Year :2013 Fresh print

Paper 4 Quantitative Methods for Management

UNIT I

Linear, Non-Linear functions – graphical representation of functions, Constants, Variables – notion of Mathematical models – concept of trade off – notion of constants – concept of Interest.

Basic Concept of differentiation – integration – Optimization concepts – use of differentiation for optimization of business problem- Optimization

UNIT II

Data Analysis – Univariate – ungrouped and grouped data measures of central Tendencies, measures of dispersion – C V percentages (problem related to business applications). Bivariate – correlation and regression – problems related to business applications

UNIT III

Probability – definitions – addition and multiplication Rules (only statements) – simple business application problems – probability distribution – expected value concept – theoretical probability distributions – Binomial, Poison and Normal – Simple problems applied to business.

UNIT IV

Basic concept of index numbers – simple and weighted index numbers – concept of weights - types of index numbers – Business index number – CPT, WPI, Sensex, Niffy, Production Index, Time series – variations in Time Series for business forecasting.

UNIT V

Hypothesis testing of Proportion and Mean – single and two tailed tests – errors in Hypothesis Testing – Measuring the power of Hypothesis test. Chi-Square Tests

References:

1. Statistics for Management – Richard L Levin & Daid S Rubin
2. Statistical Methods – S P Gupta
3. Statistics for Business and Economics – R P Hoods – MacMillan India Limited
4. David M.Levine, Timothy C.Krehbiel and Mark L.Berenson
Course , Pearson Education Asia
5. Amir D. Aczel, Complete Business Statistics, 5th edition, Irwin McGraw-Hill.

PAPER – IV Quantitative Methods For Management

UNIT 1: MATHEMATICAL BASI CS OF DIFFERENTIATION AND INTEGRATION

Introduction

The technique for converting a real world problem into a problem expressed in concrete or mathematical terms is called quantitative techniques. The converted mathematical problem is called a mathematical model of the real world problem. The real world problem is in structured and complicated. We make certain assumptions regarding the real world problem. Based on the assumptions, mathematical model is constructed and solved using available mathematical technique. If the outcome arrived at using mathematical model do not differ significantly from the actual outcome, the model gets validated and approved.

Unit Objectives

The objectives of this unit are

To understand constant and variables

To understand basics of differentiation

To understand basics of integration

To calculate simple interest and compound interest

To apply derivatives in social and economic problems.

Linear function and Non-Linear function:

Linear function

If the degree of polynomial involved in the polynomial function is one then it is called a linear function.

Examples:

(i) $f(x)=2x-4$

(ii) $g(x)=3x-2$ are linear functions

Non-Linear function

If the polynomial involved in polynomial function is of degree more than 1, then it is called a non-linear function.

In particulars,

(i) If the polynomial involved in polynomial function is of degree 2 then it is called a quadratic function.

Example: (i) $f(x)=3x^2+2x+1$

(ii) $g(x)=2x^2-2x+1$

(ii) If the polynomial involved in polynomial function is of degree 3 then it is called a cubic function.

Example: (i) $f(x) =x^3+x^2+x+1$

(ii) $g(x)=x^3-x+1$

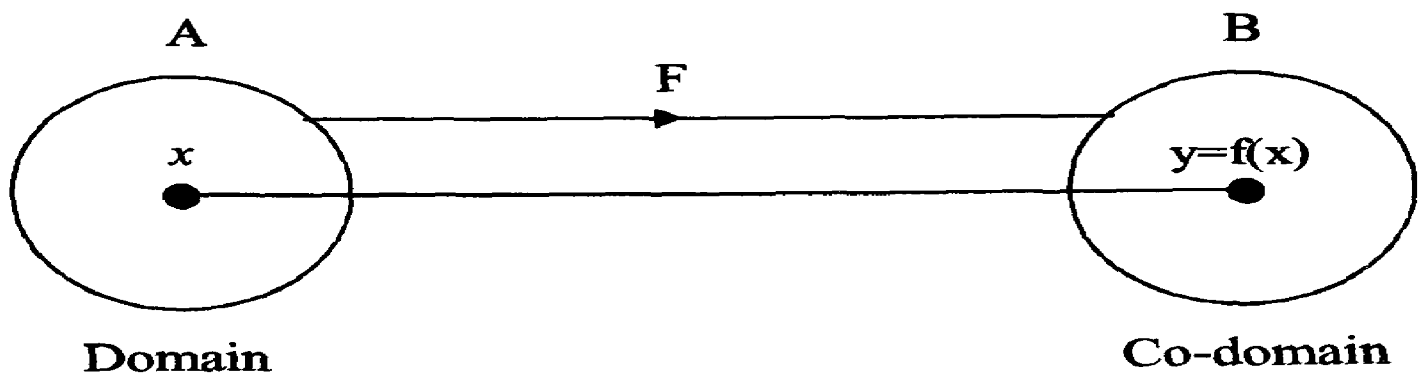
Constants and variables

1. A quantity which remains same in its value is called constant. If the value of the constant is a real number, then it is called is a real number, then it is called real constant.
2. A quantity which changes in its value is called a variable. If a variable takes the values which are real numbers only, then the variable is called a real variable.

The variables are usually denoted by u, v, x, y, z, etc.

Graphical representation of functions

Let A and B be two non empty sets. If to each element of x B, there corresponds a unique elements y B, by means of a rule denoted by 'f' then we say that a function or mapping "f" is defined from the set A into B.



Definition (Real values function

If to each value of a real variable x a unique real number y is associated, by means of a rule 'f', then we say the variable y is a real valued function of the real variable x. This we denote it by y=f(x).

The variable x is called the independent variable and the variable y is called the dependent variable.

Notion of Mathematical Modes

Definition

Let f(x) and g(x) be two real valued function of real variable x, defined on the same domain. Then the function f+g, f-g, f.g, f/g and k.f where k is the real constants are defined as below.

1. $(f+g)\ (x) = f(x) + g(x)$
2. $(f-g)\ (x) = f(x) - g(x)$
3. $(f.g)\ (x) = f(x).g(x)$
4. $\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)}\ where\ g(x) \neq 0$
5. $(k.g)(x) = k.f(x)$

Illustration

If $f(x) = x^2$ and $g(x) = 3x-1$ find the functions. $(f+g)(x)$, $(f-g)(x)$, $(f.g)(x)$, $\left(\frac{f}{g}\right)(x)$ and $(6f)(x)$.

Solution

Given that, $f(x)=x^2$, $g(x)=3x-1$

a) $(f+g)(x) = f(x) + g(x)$

$= x^2+3x-1$

b) $(f-g)(x) = f(x) - g(x)$

$= x^2-(3x-1)$

c) $(f.g)(x) = f(x).g(x)$

$= x^2(3x-1)$

$= 3x^3-x^2$

d) $\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)}$

$= \frac{x^2}{3x-1}$

Here note that, when $x = \frac{1}{3}$

$g(x)=0$

$\left(\frac{f}{g}\right)(x)$ is not defined for $x = \frac{1}{3}$

(ie) $(6f)(x) = 6f(x) = 6x^2$

Illustration

If $f(x)=2x^2-4$ find $f(1)$, $f\left(\frac{1}{2}\right)$, $f(-1)$

Solution

Given that $f(x)=2x^2-4$

a) $f(1)=2x1^2-4$

$= 2-4=-25$

b) $f\left(\frac{1}{2}\right) = 2\left(\frac{1}{2}\right)^2 - 4$

$$= 2x\frac{1}{4} - 4 = \frac{1}{2} - 4 = \frac{1-8}{2} = \frac{-7}{2}$$

c) $f(-1)=2x(-1)^2-4$

$= 2x1-4=-2$

Types of function

1. Constant function

A function y=f(x) is said to be constant function, if for every value of 'x' the value of f(x) of the function is a same constant say k.

(ie) y=f(x)=k for all real values of x.

Example

f(x)=2 is a constant function

f(0)=3, f(-1)=2, $f\left(\frac{1}{2}\right) = 2$ etc.,

2. Identity function

The function f(x)=x, for all real values of x is called the identity function defined on the set of real numbers.

In this case f(1)=1, f(x)=2, f(-5)=-5 etc.,

3. Polynomial function

A function of the form $f(x) = a_0x^n + a_1x^{n-1}+\ldots+a_n$ where n is a positive integer, $a_0, a_1, a_2, \ldots a_n$ are rational constants, is called a polynomial function.

The right hand side expression is called a polynomial in the variable 'x' of degree n.

The following are few polynomial functions

(i) $f(x)=3x^2-4x+2$ (ii) g(x)=3x-4

4. Exponential function

The function defined by $y=e^x$, for all real number x is called the exponential function with base.

5. Logarithmic function

The function defined by y=log e^x for all x>0, is called natural logarithm function of simply logarithm function.

Simple interest and compound interest

Simple Interest

Interest is the payment of extra money made by a borrower for having used the money lent by others for certain period of time. If a borrows Rs. 10,000 from B for one year, A has to pay back Rs. 10,000 plus some extra amount as a consideration for having used B's money. The extra payment made is known as Interest.

The money actually borrowed is called principal and the total sum comprising principal and interest is called the amount.

Simple interest is the interest always calculated on the principal. It is assumed that simple interest is paid at the end of a specified period regularly.

The formula for calculating simple interest is Pni.

'p' stands for principal

'n' stands for number of years (period)

i stands for rate of interest per unit

By using the above principle, the following formula can be obtained.

a) Total amount A=P(1+ni)

b) Simple interest (SI) = Pni

c) Principal $(P) = \frac{SI}{ni}$

d) Period $(P) = \frac{SI}{pi}$

e) Rate of interest per unit $i = \frac{SI}{pn}$

Illustration

Find simple interest on

(i) Rs. 1600 for one year at 10% p.a.

(ii) Rs. 2500 for 15 months at 15% p.a.

Solution

(i) Given that p=Rs. 1600, n=1, i=10% = 0.10

Simple interest SI= Pni

= 1600 x 1 x 0.10

= Rs. 160

(ii) Given that P=Rs. 2500

$$n = \frac{15}{12} \text{ years}$$

i = 15% = 0.15 per unit

Simple interest = Pni

$$= 2500x\frac{15}{12}x0.15$$

= Rs. 468.75

Illustration

At what rate percent per annum will

a) The simple interest of Rs. 75 for 9 months is Rs. 4.50?

b) Simple interest on Rs. 25 for 3 years is Rs. 4.50?

Solution

a) Given that SI=Rs. 4.50, P=Rs. 75; $n = \frac{9}{12}$

Rate of interest per unit $i = \frac{SI}{Pn}$

$$= \frac{4.50}{75x\frac{9}{12}}$$

$$= \frac{4.50}{56.25}$$

= 0.08 = 8%

Rate of interest = 8%

b) Given that P=Rs. 25, SI=Rs. 4.50; n=3 years

$$\text{Rate on interest per unit} = \frac{SI}{Pn} = \frac{4.50}{25x3} = \frac{4.5}{75} = 0.66$$

Rate of interest = 6%

Illustration

Find by using appropriate formulae,

a) The amount of Rs. 700 invested for 2½ years at 6¼% p.a.

b) Time in which Rs. 770 will amount to Rs. 847 at 5% p.a.

c) The sum which yield simple interest of Rs. 77 in 8 years at 3½% p.a.

Solution

a) Given that, P=Rs. 700; n=2.5 years i=0.0625

Total amount A = P (1+ni) = 700 (i+2.5 x 0.0625)

= 700 (1+0.15625)

= Rs. 809.38

b) Give that P=Rs. 770, SI=A-P=847-770=Rs. 77

i = 0.05

$$\text{Time } n = \frac{SI}{ni} = \frac{77}{7x0.05} = \frac{77}{38.5} = 2\ years$$

c) SI = Rs. 77, n=8 years, i=0.035

$$\text{Principal sum } P = \frac{SI}{ni} = \frac{77}{8 \times 0.035} = \frac{77}{0.028} = Rs. 275$$

Compound Interest

Compound interest is the interest calculated on the principal and accrued interest. The interest unpaid is added to the principal and for the next period of interest is computed. In other words, compound interest on the growing principal. Both principal and compound interest change from time to time. Compound interest can be computed annual, half yearly, quarterly or monthly.

The following formulae are kept in mind solving the problems concerning to compound interest

1) $CI=P(1+i)^n-P$
2) $A=P(1+i)^n$
3) $P=A+(1+i)^n$
4) $A = P\left(1+\frac{1}{2}\right)^n$ if CI is paid half yearly and 'n' indicates the number of 'half years' in the period.
5) $A = (1+i)^n\left(1+\frac{i}{2}\right)^n$ etc. if CI is paid annually but for each period o time an different rate of interest is paid.

Illustration

Find the compound interest in following cases

a) On Rs. 10,000 for 2 years at 10% p.a. paid annually
b) On Rs. 6,000 for 2½ years at 10% p.a. paid
c) On Rs. 5,000 for 1½ years at 5% p.a., CI paid half yearly
d) Rs. 1,000 for one year at 10% p.a., CI is paid quarterly

Solution

Give p=Rs. 10,000, n=2 years, i=0.10

Amount $A = P(1+i)^n$

$= 10000\ (1+0.10)^2$

$= 10000\ (1.1)^2$

$= 12,\ 00$

Compound interest CI = A-P

= 12100-10000

= Rs. 2100

b) P=Rs. 6000, n=2½ years; i=0.10, $\frac{i}{2} = 0.05$

$$A = P(1+i)^2\left(1+\frac{i}{2}\right)^1$$

$= 6000\ (1+0.10)^2\ (1+0.05)^1$

$= 6000 \times (1.10)^2 \times 1.05$

= Rs. 7,623

c) P=Rs. 5000, n=1½ years = 3 half years, i=0.05

$\frac{i}{2} = 0.025$

$$A = P\left(1+\frac{i}{2}\right)^n$$

$= 5000\ (1+0.025)^3$

$= 5000\ (1.025)^3$

$= 5000 \times 1.0769$

= Rs. 5384.45

CI = A-P = 5384.45 – 5000

= Rs. 384.45

d) P=Rs. 1,000; n=1 years or 4 quarters; i=0.10

$\frac{i}{4} = 0.025\ quarterly$

$$A = P\left(1+\frac{i}{4}\right)^n$$

$= 1000 \times (1+0.025)^4$

$= 1000 \times (1.025)^4$

= Rs. 1103.83

CI = A-P

= 1,103.83-1,000

= Rs. 103.83

Illustration

Find the amount if the principal is Rs. 10,000 and the rate of interest is 15% p.a. for 2 years if (a) CI is paid annually (b) CI is paid half yearly.

Solution

Given that P=Rs. 10,000, i=0.15 per unit, n=2 years

a) If CI is paid annually,

$$A=P\,(1+i)^n = 10,000\,(1+0.15)^2$$
$$= 10,000\,(1.15)^2$$
$$= \text{Rs. } 13.225$$

b) If CI is paid half yearly

$$A = P\left(1+\frac{1}{2}\right)^n = 10000(1+0.075)^4$$
$$= 1000\,(1.075)^4$$
$$= \text{Rs. } 13,354.69$$

BASIC CONCEPT OF DIFFERENTIATION

Differential coefficient

Consider the function y=f(x). Here x is the independent variable and y is the dependent variable. The value y of the function f(x) changes as the variable x changes. The rate at which the function changes, is of great importance in the study of the subject "calculus". The process of finding the rate of change of the function is called differentiation.

Differential coefficient of a certain basic functions

1. Differential coefficient of a constant function is 0

   (ie) $\frac{d(c)}{dx} = 0$

2. Differential coefficient of the function $y=x^n$ where n is any rational number,

   If $y=x^n$ then $\frac{dy}{dx} = nx^{n-1}$

3. Differential coefficient of the exponential function

   If $y=e^x$ then $\frac{dy}{dx} = e^x$

4. The differential coefficient of the product of two functions is given by

$$\text{If } y=uv \text{ then } \frac{dy}{dx} = u\frac{d}{dx}v + v\frac{d}{dx}u$$

5. The differential coefficient of the quotient of two function is given by

$$\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{v\frac{du}{dx} - v\frac{d}{dx}v}{v^2}$$

Illustration

Differentiate the following w.r.t.x

a) x^7 b) $x^{4/3}$ c) $\frac{1}{x^{10}}$ c) $\frac{1}{x^{5/2}}$

Solution

a) Let $y=x^7$

Differentiate with respect to x,

$$\frac{dy}{dx} = 7x^{7-1}$$

$$= 7x^6$$

b) Let $y = x^{\frac{4}{3}}$

Diff. w.r.t.x

$$\frac{dy}{dx} = \frac{4}{3}x^{\frac{4}{3}-1}$$

$$= \frac{4}{3}x^{\frac{1}{3}}$$

c) Let $y = \frac{1}{x^{10}} = x^{-10}$

Diff.w.r.t.x

$$\frac{dy}{dx} = -10^{-10-1} = -10x^{-11}$$

d) Let $y = \frac{1}{x^{5/2}} = x^{-5/2}$

Diff. w.r.t.x,

$$\frac{dy}{dx} = \frac{-5}{2} x^{\frac{-5}{2}-1}$$

$$= \frac{-5}{2} x^{\frac{-7}{2}}$$

Illustration

Differentiate the following w.r.t.x

a) $x^3 - 3x^2 + 4x + 3$ b) $x^4 + 3logx - 4e^x$

Solution

a) Let $y=x^3-3x^2+4x+3$

Diff.w.r.t.x,

$$\frac{dy}{dx} = 3x^2 - 6x + 4$$

b) Let $y=x^5+3logx-4e^x$

Diff.wr.t.x,

$$\frac{dy}{dx} = 5x^4 + \frac{3}{x} - 4e^x$$

Illustration

Differentiate the following w.r.t.x

a) $(x^3 - 3x + 2)e^x$ b) $\frac{1-x^2}{1+x^2}$

Solution

a) Let $y=(x^3-3x+2)e^x$

Diff. w.r.t.x

$$\frac{dy}{dx} = (x^3 - 3x + 2)\frac{d}{dx}(e^x) + e^x \frac{d}{dx}(x^3 - 3x + 2)$$

$$= (x^3 - 3x + 2)e^x + e^x(3x^2 - 3)$$

$$= (x^3 + 3x^2 - 3x - 1)e^x$$

b) Let $y = \frac{(1-x^2)}{(1+x^2)}$

Diff.w.r.t.x

$$\frac{dy}{dx} = \frac{(1+x)^2 \frac{d}{dx}(1-x^2) - (1-x^2)\frac{d}{dx}((1+x^2))}{(1+x^2)^2}$$

$$= \frac{(1+x^2)(-2x)-(1-x^2)(2x)}{(1+x^2)^2}$$

$$= \frac{-2x--2x^3-2x+2x^3}{(1+x^2)^2}$$

$$= \frac{-4x^3}{(1+x^2)^2}$$

2.2. Higher order derivatives

If a function is differentiated more than one time is called higher order derivative.

The second order derivative is denoted by $\frac{d^2y}{dx^2}$

Illustration

Find the second order derivative x^3-$3x^2$+3x+7

Let $y=x^3-3x^2+3x+7$

Diff. w.r.t.x

$$\frac{dy}{dx} = 3x^2 - 6x + 3$$

Again, diff. w.r.t.x

$$\frac{d^2y}{dx^2} = 6x - 6$$

Illustration

Find the second derivative of $2x^3$-$3x^2$-36x+10

Let $y=2x^3-3x^2-36x+10$

Diff. w.r.t.x

$$\frac{dy}{dx} = 6x^2 - 6x - 36$$

Again diff.w.r.tx,

$$\frac{d^2y}{dx^2} = 12x - 6$$

Integration

Definition of a integral

The integral of a function f(x) with respect to x is that function whose derivative with respect to x in f(x) and is written as f(x) dx.

Generally f(x)dx= (x)+c, since c being an arbitrary constant can take any value, f(x)dx is called the indefinite integral.

In this is the sign of integration f(x) is called the inegrand and x is called the variable of integration. The process of findings the integral is called integration standard forms.

1. $x^n dx = \frac{x^{n+1}}{n+1} + c$ for all values of n except n=-1
2. $\frac{dx}{x} = \log x + c$
3. $\int e^x dx = e^x + c$
4. $\int e^{ax} dx = \frac{e^{ax}}{n}$
5. $\int a^x dx = \frac{a^x}{\log a}$

Illustration

Evaluate $\int x^2 dx$

Solution

$$\int x^2 dx = \frac{x^{2+1}}{2+1} + c$$

$$= \frac{x^3}{3} + c$$

Illustration

Evaluate $\int \left(x^3 - 3x^{\frac{1}{2}} + \frac{1}{x^2}\right) dx$

Solution

$$\int \left(x^3 - 3x^{\frac{1}{2}} + \frac{1}{x^2}\right) dx$$

$$= \left(x^2 - 3x^{\frac{1}{2}} + x^{-2}\right) dx$$

$$= \frac{x^3}{2} - \frac{3x^{\frac{3}{2}}}{\frac{3}{2}} + \frac{x^{-1}}{-1} + c$$

$$= \frac{x^4}{2} - 2x^{3/2} - \frac{1}{x} + c$$

Definition

Let $\int f(x)dx = F(x) + c$ where c is the constant of integration.

The value of the integral when x=b is F(b)+c. The value of integral when x=a is F (a) +c. subtracting

F (b)-F (a) = value of $\int f(x)dx$ when x=b-value of $\int f(x)dx$ when x=a

The notation $\int_a^b f(x)$ is used to denote the value of integral when x=b the value of the integral when x=a.

$\int_a^b f(x)dx$ is called a definite integral.

Its value is F (b)-F (a) where F(x) is the integral of f(x) with respect to x. 'a' and 'b' are called the limits of integration, 'a' being the lower limit and b being the upper limit.

$\int_a^b f(x)dx$ is to be read as integral from a and b of f(x) dx the integral a to b is called the range of integration.

Illustration

Evaluate $\int_1^2 (x^3 + x^2 + x)dx$

Solution

$$\int_1^2 (x^3 + x^2 + x)dx = \left(\frac{x^4}{4} + \frac{x^3}{3} + \frac{x^2}{2}\right)^2$$

$$= \left(\frac{2^4}{4} + \frac{2^3}{3} + \frac{2^2}{2}\right) - \left(\frac{14}{4} + \frac{13}{3} + \frac{12}{2}\right)$$

$$= \left(\frac{16}{4} + \frac{8}{3} + 2\right) - \left(\frac{1}{4} + \frac{1}{3} + \frac{1}{2}\right)$$

$$= \left(\frac{48+32+24}{12}\right) - \left(\frac{3+4+6}{12}\right)$$

$$= \frac{104}{12} - \frac{13}{12}$$

$$= \frac{104-13}{12}$$

$$= \frac{91}{12}$$

Evaluate $\int_1^2 \left(x^2 - 3x^{\frac{1}{2}} + \frac{1}{x^2}\right) dx$

$$= \left[\frac{x^3}{3} - 2x^{\frac{3}{2}} - \frac{1}{x}\right]_1^2$$

$$= \left[\frac{8}{3} - 4\sqrt{2} - \frac{1}{2}\right] - \left[\frac{1}{3} - 2 - 1\right]$$

$$= \frac{29}{6} - 4\sqrt{2}$$

APPLICATION OF THE DERIVATIVE

The geometric meaning of $\left(\frac{dy}{dx}\right)$ is the slope of the tanget to the curve y=f(x) at the point (x,y).

$\left(\frac{dy}{dx}\right)$ is also called the rate of change of y with respect to x.The cost of C comprised to two component – fixed cost and the variable cost. It is a function of output x. (x is the number of units of output)

TC=FC+VC

C=f(x)

Average cost $(AC) = \frac{Totcal\ cost}{Output} = \frac{C}{x}$

Marginal cost (MC $= \frac{dC}{dx}$

Marginal average cost $(MAC) = \frac{d(AC)}{dx}$

Example

Find the average cost and the marginal cost functions from the total cost function.

$C = 60 + 10x + 15x^2$

$$Average\ cost = \frac{Total\ cost}{Output}$$

$$= \frac{60+10x+15x^2}{x}$$

$$= \frac{60}{x} + 10 + 15x$$

Marginal cost (MC) $= \frac{dC}{dx} = 0 + 10(1) + 15(2x)$

$= 10+30x$

Revenue

Total revenue R = Price x quantity sold

R = Px

$$Average\ revenue = \frac{Total\ revenue}{quantity\ sold} = P$$

$$Marginal\ Revenue = \frac{dR}{dx}$$

$= \frac{d}{dx}(px) = P + x\frac{dF}{dx}$ (Here the product rule of Calculate is applied)

Elasticity Function

The elasticity of the function y=f(x) at the point x is given by

$$\eta = \frac{d}{y}\frac{dy}{dx}$$

Example

If y=3x-6, find the elasticity of y

Solution

$$Elasticity = \eta = \frac{d}{y}\frac{dy}{dx}$$

A $y = 3x - 6, \frac{dy}{dx} = 3(1) - 0 = 3$

$$\therefore \eta = \frac{x}{3x-6}.3 = \frac{3x}{3x-6}$$
$$= \frac{3x}{3(x-2)} = \frac{x}{x-2}$$

Elasticity Demand

If x=f(p) is the demand function, where x is the demand, and p is the price then the elasticity of supply is

$$\eta_s = \frac{p}{x}\frac{dx}{dp}$$

Example

Find the elasticity of supply for the supply function $x = 2p^2 + 5$

Elasticity of supply is given by $\eta_s = \frac{p}{x}.\frac{dx}{dp}$

As $x = 2p^2 + 5, \frac{dx}{dp} = 2(2p) + 0 = 4p$

$$\eta_s = \frac{p}{x}.4p = \frac{4p^2}{2p^2+5}, since\ x = 2p^2 + 5$$

Example 2

If the demand law is x=20/p+1, find the elasticity of demand at the point where p=3.

Solution

x=20/p+1

$$\frac{dx}{dp} = \frac{(p+1)\frac{d}{dp}(20)-20\frac{d}{dp}(p+1)}{(p+1)^2}$$
$$= \frac{(p+1)(0)-20(1+0)}{(p+1)^2} = -\frac{20}{(p+1)^2}$$

Elasticity of demand $\eta_d = \frac{p}{x}.\frac{dx}{dp}$

$\eta_d = \left(p/\frac{20}{p+1}\right)\left(-\frac{20}{(p+1)^2}\right) = \frac{p}{p+1}$ (Considering only the magnitude)

Elasticity of demand when p=3 is

$$n_d = \frac{3}{3+1} = \frac{3}{4} = 0.75$$

Example 3

The total cost C of making x units of product is $C=0.00003x^3-0.045x^2+8x+25000$. Find the marginal cost at 1000 units output.

Solution

$C=0.0000x^3-0.45x^2+8x+25000$

$$Marginal\ cost = \frac{dC}{dx} = (0.00003)3x^2 - (0.45)2x + 8.1$$

$$= 0.00009x^2-0.09x+8$$

Marginal cost at x = 1000 is $0.00009\ (1000)^2 - 0.09(1000)+8=8$

Example 4

For the demand function $p=550-3x-6x^2$ where x is the quantity demanded and p is the price per unit, find the average revenue and marginal revenue.

The revenue function is given by

$$R = Px$$
$$= (550-3x-6x^2)\ x$$
$$= 550x-3x^2-6x^3$$

Average revenue is given by $R = \frac{R}{x} = 550 - 3x - 6x^2$

Marginal revenue $\frac{dR}{dx} = \frac{d}{dx}(550x - 3x^2 - 6x^2)$

$$= 550-6x-18x^2$$

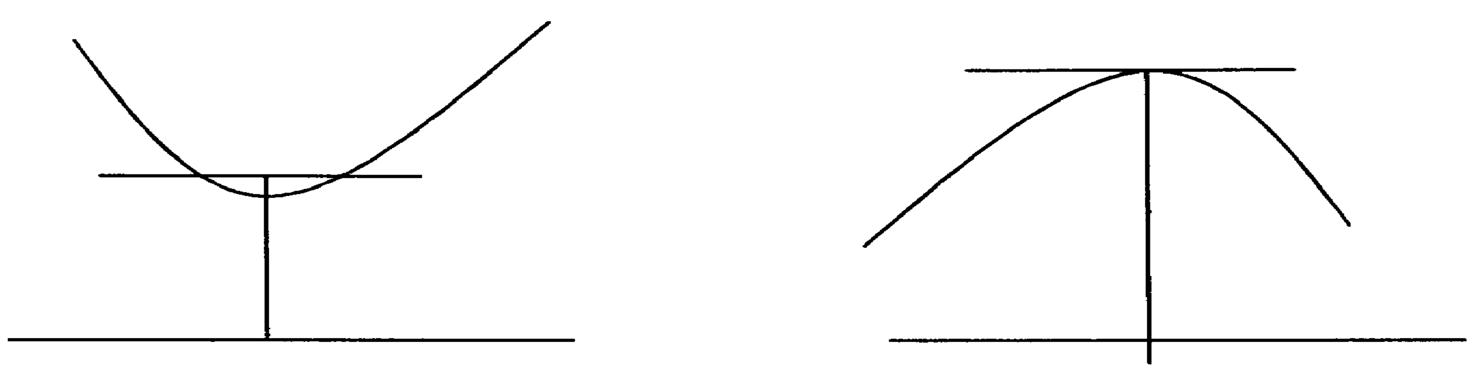
OPTIMIZATION CONCEPT

MAXIMA AND MINIMA

Everybody wants to maximize his gain and minimize his loss. Any decision to be taken by an organization, when quantified, reduced to a cost function or a profit function. We went to minimize the cost function and maximize the profit function.

When the cost / profit function is linear, we can draw the graph of the function easily and find the maximum / minimum value from the graph itself.

When the function is a quadratic function, we can use calculus to find the maximum or minimum. Fig 1.7 gives two functions having a minimum and maximum at $x=x_0$ respectively. In both cases, the tanget at $(x0,f(x_0))$ is parallel to the x-axis. As $\left(\frac{dy}{dx}\right)x = x_0$ is the slope of the tanget $x=x_0$, $\left(\frac{dy}{dx}\right)x = x_0$ has to zero,.



X_0 Fig 1 X_0

To distinguish between attainment of maximum and minimum, we evaluate $\left(\frac{d^2y}{dx^2}\right)x = x_0$. I $\left(\frac{d^2y}{dx^2}\right)x = x_0$ is positive, the function $y=f(x)$ has minimum at $x=x_0$. If $\left(\frac{d^2y}{dx^2}\right)x = x_0$ is negative, the function $y=f(x)$ has maximum at $x=x_0$. (If $\left(\frac{d^2y}{dx^2}\right)x = x_0 = 0$, then we have conditions for maximum in terms of higher derivative. This is beyond the scope of our study.

Procedure to find Maxima and Minima

Let $y=f(x)$ be a function, whose maxima and minima are to be obtained.

i) Find $\left(\frac{dy}{dx}\right)$ and equate it to zero

ii) Solve the equation $\frac{dy}{dx} = 0$ obtain its roots, say $x_1, x_2, \ldots$

iii) Find $\frac{d^2y}{dx^2}$

iv) Find the value of $\frac{d^2y}{dx^2}$ at $x=x_i$. If it is negative they y=f(x) attains a maximum at $x=x_i$. If it is positive then y=f(x) attains a minimum.

Perform step iv) for each root to get all the maxima and minima

Example 1

Find the points or maxima and minima of the function

$$y=x^3-3x^2+5$$

Solution

$$y=x^3-3x^2+5$$

$$\therefore \frac{dy}{dx} = 3x^2 - 6x$$

$$= 3x(x-2)$$

$\frac{dy}{dx} = 0$ when $3x(x-2)=0$

That is, x=0 or x=2

Differentiating $\frac{dy}{dx}$ again w.r.t x, we get

$$\frac{d^2y}{dx^2} = 6x - 6$$

When x=0, $\frac{d^2y}{dx^2} = -6 < 0$ (negative)

and when x=2, $\frac{d^2y}{dx^2} = 6 > 0$ (positive)

The function attains its maximum at x=0 and minimum at x=2. At x=0, y=0-0+5=5 is the maximum value and at x=2, y=8-12+5=1 is the minimum value.

Example 2 :

Investigate the maxima and minima of the function $y-2x^3+3x^2-36x+16$

Solution

$y=2x^3+3x^2-36x+16$

$\therefore \frac{dy}{dx} = 6x^2 + 6x - 36$

Equate $\frac{dy}{dx}$ *to zero*

(i) $\frac{dy}{dx} = 0 = 6x^2 + 63x - 36 = 0$

When divided by 6, we get $x^2+x-6=0$

While solving the equation $x^2+x-6=0$, wet get (x+3) (x-2)=0

Hence x=-3 (or) x=2

Differentiating $\frac{dy}{dx}$ once again w.r.t.x,

$$\frac{d^2y}{dx^2} = 12x + 6$$

At x=-3, $\frac{d^2y}{dx^2} = -36 + 6 = -30 < 0$

At x=-3, the function y has maximum value. To get the maximum value substitute x=-3 in y.

y=2(-27)+3(0)-36(-3)+10

= -54+27+108+10

y = 91

91 is a maximum value

At x=2, $\frac{d^2y}{dx^2} = 24 + 6 = 30 > 0$

The function y has a minimum value at x=2. Substitute x=2 in y.

y=2(8)+34-72+10

= 16+12-72+10

= -34

The minimum value is -34.

Example 3

Find the maxima and minima value of the cost function $C=5+2x^2-x^3$

SOLUTION

$$C=5+2x^2-x^3$$

$$\frac{dC}{dx} = 4x - 3x^2$$

Equate $\frac{dC}{dx}$ to zero

(ie) $\frac{dC}{dx} = 0 \Rightarrow 4x - 3x^2 = 0$

$$\Rightarrow x(4-3x) = 0$$

$$\Rightarrow x = 0\ (or)\ x = 4/3$$

Differentiating $\frac{dC}{dx}$ once again w.r.t.x,

$$\frac{d^2C}{dx^2} = 4 - 6x$$

when x=0, $\frac{d^2C}{dx^2} = 4 - 6(0) = 4 > 0$. Hence C has a minimum at x=0.

The minimum cost = $5+2(0)^2-(0)^3=5$

When $x = \frac{4}{3}; \frac{d^2C}{dx^2} = 4 - 6\left(\frac{4}{3}\right)$

= 4-8=-4<0. Hence C has a maximum

When $x = \frac{4}{3}; C = 5 + 2\left(\frac{4}{3}\right)^2 - \left(\frac{4}{3}\right)^3 = 167/27$

Example 4

A cricket association in a city is arranging an inter-state cricket match. The association estimates that 3000 spectators would attend the match if the ticket costs Rs. 80. It feels for ever decrease of Rs. 10 in the ticket price, 600 additional persons will buy tickets. Find the optimal ticket price so that the association will get maximum revenue.

SOLUTION

Let p denote the price / ticket and n be the number of spectators. We can assume that the number of additional spectators is proportional to the decrease in the price of a ticket. So far decrease of one rupee in ticket price, the increase in number of spectators in 60.

n = 3000 if p=80

n = 3000 + 60x if p=80-x

Total revenue R=n.p

$= (3000+60x)(80-x)$

$= 240000+1800x-60x^2$

$$\frac{dR}{dx} = 1800 - 120x$$

If $\frac{dR}{dx} = 0, 1800 - 120x = 0, \text{ie } x = 15$

$$\left(\frac{d^2R}{dx^2}\right) = -120 < 0$$

Hence R attains maximum when x=15.

So the optimal ticket price is 80-15=65.

Maximum Revenue $= 240000+1800(15)-60(15)^2$

$= 253500$ rupees

When the profit / cost is linear or piecewise linear (that is, it is given by two different linear functions for different range of values), we cannot apply calculus to get maxima or minima. In this case, we can represent the function graphically and get the point at which maximum / minimum is attained.

SUMMARY

In this unit we came to know that the first and higher derivatives of a function, maxima and minima of a function and application of derivatives.

KEYWORDS

Fixed cost

Variable cost

Average cost

Marginal cost

KEY TO CHECK YOUR PROGRESS

1. Find the second derivative of $2x^3 \quad 3x^2 \quad 36x \quad 10$.

   Ans: $12x \quad 6$.

2. Find $\frac{dy}{dx}$ if y= $5x^3 \quad 6x^2 \quad 3$.

   Ans: $15x^2 \quad 12x$.

Exercise:

1. Find the maximum value of the function

$x^4 \quad 2x^3 \quad 3x^2 \quad 4x \quad 4$.

2. The total variable cost of monthly output x tons by firm producing a variable metal is Rs. $\frac{1}{10}x^3 \quad 3x^2 \quad 5x$ and the fixed cost is Rs. 300per month. Find the output of minimum average cost.

3. The unit demand function is x=$\frac{1}{3}(25 \quad 2p)$, where x is the number of units and p is the price. Let the average cost per unit be Rs. 40. Find

i. The revenue function of R in terms of p,

ii. The cost function C,

iii. The profit function P,

iv. The price per unit that maximizes the profit function,

v. The maximum profit.

UNIT 2 CENTRAL TENDENCY, DISPERSION AND CORRELATION

INTRODUCTION

One of the most important objectives of statistical analysis is to get one single value that describes the characteristic of the entire mass of unwieldy data, such a value is called the central value or an average or the expected value of the variable. The word average is very commonly used in day-to-day conversation. For example, we often talk of average boy in a class, average height or life of an Indian, average Income etc., when we say , 'he is an average student', what it means is that he is neither very good nor very bad, just a mediocre type of student. However in statistics the term average has different meaning "Average is an attempt to fine one single figure to describe whole of figures"

Unit Objectives

On learning this unit, you could able to

Study about Data analysis.

Discuss about measurers of Central Tendencies.

Discuss about measures of Dispersion

To find the relation between two variables.

To find the correlation between the values in terms of their ranks.

To calculate the unknown variable with the help of a known variable.

Unit Structure:

2.1 Data Analysis.

2.2 Measures of Central Tendency.

2.3 Measures of Dispersion.

2.4 Correlation Analysis.

2.5 Regression Analysis.

2.1 Data Analysis

Statistical data constitute the raw material for statistical methods. The statistical data may be classified into two types. They are

(i) Primary data and

(ii) Secondary data

Primary Data

The processed data which is given to the user is called information. If the data is collected for the first time then it is called Primary data.

Example:

1. Attitude survey on a particular brand of product available in the market.

2. The reasons for absenteeism among employees in an organization.

3. Opinion poll about the election results.

The above examples represent primary data and are generated for the first time buy the investigator.

Secondary Data

If data, which is already collected and used for some purpose is once again used for different study, then it is called secondary data. As far as secondary data is concerned the availability and reliability of existing data should be given more importance.

Methods of Collecting Primary Data

Observation Techniques   Interview Techniques

Questionnaires Survey Techniques

Primary data is collected by Observation Techniques or Interview Techniques or Questionnaires Survey Techniques.

Observation Method

In this method the researcher has collected the information himself. This is more intensive rather than extensive. It gives more reliable information, because the researcher has cross examined particular situation in several ways.

Merits:

1. It is in depth study.
2. More accurate date can be obtained.

Demerits:

1. It is not uniform. It differs from person to person.
2. It demands very high interpersonal skills of the researcher.
3. It involves more cost and time.

Interview Method

The interviewer and Interviewee are directly meeting each other. The interviewer asks questions and the responses from the interviewee were recorded.

Telephonic Interview:

In this method responses are collected through telephonic conversation.

Merits:

1. The information is unbiased.
2. Adequate information is collected from the respondent.
3. High degree of accuracy can be aimed.
4. Responses will be encouraging because of personal approach.

Demerits:

1. Costlier and time consuming.
2. In order to get real position sufficient numbers of persons are to be interviewed.
3. It needs high skills for the interviewer to elicit proper responses.

Questionnaire Survey

A printed questionnaire which contains the questions for eliciting the required information is designed. The questionnaires are supplied to the respondents and the respondents are asked to fill up.

Merits:

1. It is the cheapest and less time consuming.
2. It can be used extensively and the area coverage is very large.
3. Error involved in the survey is less due to extensive area coverage.

Demerits:

1. This is mainly used for literate people.
2. Inelasticity is more, and so acquiring further information is not possible.
3. We cam not get 100% response, since the rejection of questionnaire will be inevitable

Secondary Data

There are two types of sources of secondary data

(i) Published sources

(ii) Unpublished sources.

Published Sources:

Various governmental, international and local agencies publish statistical data and important among them are

(a) International publications.

(b) Official publications of central and state governments.

(c) Semi Official publications.

(d) Publications of research institution such as Indian Statistical Institute, Indian Council of agricultural research etc,

(e) Publication of Commercial and financial institutions.

(f) Reports of various committees and commissions appointed by the government.

(g) Journals and news papers.

Unpublished sources:

There are various sources of unpublished data. They are records maintained by the various government and private offices, the researches carried out by individual research scholars in the universities or research institutes.

2.2 Measures of Central Tendency

The measure of central tendency is called as measures of location. This measure of Central Tendency is called as "Average".

Average is defined as a single value with in the range of data that is used to represent all values in the series. If we consider the scope of average, the average is represented by the following measures.

1. Arithmetic mean.
2. Median
3. Mode
4. Geometric Mean
5. Harmonic Mean

Arithmetic Mean

(i) The arithmetic mean x of n observations $x_1, x_2, \ldots x_n$ is given by

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

(ii) The formula for calculating average for data with frequencies is mathematically expressed as arithmetic mean.

$$\bar{x} = \frac{\sum f_i x_i}{N} \quad \text{Where } N = \sum f_i$$

Where $\bar{x}$ - Arithmetic mean

f_i - frequency of i^{th} item

X_i - The individual value of i^{th} item.

(iii) In the case of grouped or continuous frequency distribution, the arithmetic mean x is given by

$$\bar{x} = A + \frac{\sum_{i=1}^{n} f_i x_i}{N} \times c$$

Where $di = \frac{Xi-A}{C}$

C – Length of the interval

A- Assumed mean.,

Illustration

The age of 10 retired pensioners are given as below

74, 62, 84, 72, 61, 83, 72, 81, 64, 71.

Calculate the average age of pensions who are getting pension.

Solution: Average age $= \frac{\sum_{i=1}^{n} x_i}{n}$

$= \frac{74+62+84+72+61+83+72+81+64+71.}{10}$

$= \frac{724}{10}$

$= 72.4$

Illustration:

Calculate the monthly average pension payable per month.

Monthly pension	20	25	30	35	40
No. of person receiving the pension	7	5	6	4	3

Solution

Monthly pension x	No. of person receiving the pension f	fx
20	7	140
25	5	125
30	6	180
35	4	140
40	3	120
	$N = \sum f$ =25	$\sum fx = 705$

Average pension $\bar{x} = \frac{\sum fx}{N}$

$$= \frac{705}{25}$$

$$= 28.20$$

Illustration:

Calculate the Arithmetic mean of the marks from the following table

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	12	18	27	20	17	6

Solution

Marks X	No. of students F	x = Mid X	$d = \frac{x-A}{c} = \frac{x-25}{10}$	fd
0-10	12	5	-2	-24
10-20	18	15	-1	-18
20-30	27	25	0	0
30-40	20	35	1	24
40-50	17	45	2	34
50-60	6	55	3	18
	N=100			$\sum fd = 30$

Arithmetic mean $= A + \frac{\sum fd}{N} \times c$

$= 25 + \frac{30}{100} \times 10$

$= 25 + \frac{300}{100}$

$= 25 + 3$

$= 28$

Note: If the grouped frequency distribution is not continuous, we first convert it into continuous distribution as in the following example,

(Subtract 0.5 from lower limit and add 0.5 to upper limit)

Calculate the Arithmetic mean of the marks from the following table

Marks	0-19	20-29	30-39	40-49	50-59
No. of students	9	11	10	44	40

Solution

Marks x	No. of students F	x = Mid X	$d = \frac{Xi-A}{c} = \frac{x-34.5}{10}$	fd
9.5-19.5	9	14.5	-2	-12
19.5-29.5	11	24.5	-1	-11
29.5-39.5	10	34.5	0	0
39.5-49.5	44	44.5	1	44
49.5-59.5	40	54.5	2	80
			3	
	N=114			$\sum fd = 95$

Arithmetic mean $= A + \frac{\sum fd}{N} \times C$

$$= 34.5 + \frac{95}{114} \times 10$$

$$= 34.5 + \frac{950}{114}$$

$$= 34.5 + 8.3$$

$$= 42.83$$

Median

(i) The median is the middle observation in data that have been arranged in ascending numerical sequence.

(ii) In the case of frequency distribution Xi/f_i, median is obtained by considering the cumulative frequencies. The steps for calculating median are

(a) Find N where N = $\sum f_i$ = total number of frequencies.

(b) See the Cumulative frequency just greater than N/2.

(c) The corresponding value of x is the Median.

(iii) In the case of continuous frequency distribution, the class corresponding to the cumulative frequency (C.F) just greater than N/2 is called the median class and the value of median is obtained by the formula

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - C\right)$$

Where l – Lower limit of the median class.

f – Frequency of the median class.

h – Magnitude of the median class.

c – Cumulative frequency of the class preceding the median class.

Illustration

In a telephone exchange the data on the inter arrival time in seconds was collected and the observations are given below:

18, 18, 20, 22, 10, 8, 7. Find the median value.

Solution:

Step 1 : Arrange the items in ascending order

7, 8, 10, 15, 18, 20, 22

Steps 2 : Median $= \left(\frac{n+1}{2}\right)^{th}$ item

$= \left(\frac{7+1}{2}\right)^{th}$ item

$= \left(\frac{8}{2}\right)^{th}$ item

$= 4^{th}$ item

$= 15$

Illustration:

In a telephone exchange, the data on the inter arrival time in seconds is given below:

15, 18, 20, 25, 22, 10, 8, 7 Find the median value.

Solution:

Step 1 : Arrange the items in ascending order

7, 8, 10, 15, 18, 20, 22, 25

Steps 2 : Median

$$= \left(\frac{n+1}{2}\right)^{th} \text{item}$$

$$= \left(\frac{8+1}{2}\right)^{th} \text{item}$$

$$= \left(\frac{9}{2}\right)^{th} \text{item}$$

$$= 4.5^{th} \text{ item}$$

$$= \frac{4^{th} \text{ item} + 5^{th} \text{ item}}{2}$$

$$= \frac{15+18}{2}$$

$$= \frac{33}{2}$$

$$= 16.5$$

Illustration:

Obtain the median for the following frequency distribution

X	1	2	3	4	5	6	7	8	9
f	8	10	11	16	20	25	15	9	6

Solution:

x	f	Cumulative frequency
1	8	8
2	10	18
3	11	29
4	16	45
5	20	65
6	25	90
7	15	105
8	9	114
9	6	120
	N=120	

Here $\frac{N}{2} = \frac{120}{2} = 60$

The c.f just greater than 60 is 65

Median class

The value of x corresponding to 65 is 5

Hence median = 5

Illustration:

Obtain the median wage of the following distribution

Wages	20-30	30-40	40-50	50-60	60-70
No of labours	3	5	20	10	5

Solution:

Wages X	f	C.F
20-30	3	3
30-40	5	8
40-50	20	28
50-60	10	38
60-70	5	43
	N = 43	

Here $\frac{N}{2} = \frac{43}{2} = 21.5$

The c.f just greater than 21.5 is 28

Median class

Median

$$= l + \frac{h}{f}\left(\frac{N}{2} - C\right)$$

Where l – Lower limit of the median class=40.

f – frequency of the median class=10

h – Magnitude of the median class. \= 20

c – is the cumulative frequency of the class preceding the median class= 8

$$\therefore \text{ Median} = 40 + \frac{10}{20}\left(21.5-10\right)$$

$$= 40 + 6.75 = 46.75$$

Mode

(i) Mode is the value which occurs most frequently in a set of observations and around which the other items of the set cluster closely.

(ii) In the case of discrete frequency distribution mode is the value of x corresponding to the maximum frequency

(iii) In the case of continuous frequency distribution mode is give by the formula

$$\text{Mode} = l + \frac{h\ (f_1 - f_o)}{2\ f_1 - f_o - f_2}$$

Where l – is the lower limit of the Modal class.

h – is the magnitude of the Modal class.

fi – is the frequency of the Modal class.

fo – is the frequency of the class preceding Modal class

f_2 – is the frequency of the class succeeding the Modal class

Modal class: - The class corresponding to the maximum frequency is called modal class

Illustration

Compute the mode for the following:

1, 9, 4, 2, 12, 4, 14, 15, 5, 14, 6, 14, 12

Solution: By data

1, 9, 4, 2, 12, 4, 14, 15, 5, 14, 6, 14, 12

Mode = 14

Illustration:

Find the mode for the following distribution:

X	1	2	3	4	5	6	7	8
F	4	9	6	25	22	18	7	3

Solution:

X	1	2	3	4	5	6	7	8
F	4	9	6	25	22	18	7	3

Maximum frequency

Here maximum frequency is 25.

The value of x corresponding to the maximum frequency 25 is 4

Hence Mode = 4

Illustration:

Find the mode for the following distribution

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No of Students	5	8	7	12	28	20	10	10

Solution:

Marks x	No. of students f	
0-10	5	Model Class = 28 (Maximum frequency)
10-20	8	
20-30	7	

30-40	12	
40-50	28	
50-60	20	
60-70	10	
70-80	10	

$$\text{Mode} = 1 + \frac{h\ (f_1 - f_o)}{2\ f_1 - f_o - f_2}$$

Where 1 – 40 (lower limit of the Modal class)

h – 10 (length of the class interval)

fi – 28 (frequency of the Modal class)

fo – 12 (frequency of the class preceeding the Modal class)

f_2 – 20 (frequency of the class succeeding the Modal clas

$$\text{Mode} = 1 + \frac{h\ (f_1 - f_o)}{2\ f_1 - f_o - f_2}$$

$$= 40 + \frac{10\ \ (28 - 12)}{2x28 - 12 - 20}$$

$$= 40 + \frac{10\ (16)}{56-32}$$

$$= 40 + \frac{160}{24}$$

$$= 46.667$$

Empirical relationship between Mean (M), Median (M_d) and Mode (M_o)

In case of symmetrical distribution Mean, Median and Mode coinside (ie $M = M_d = M_o$.

Suppose the distribution is asymmetrical, Mean, Median and Mode usually will co not inside but they obey the following important empirical relationship, given by Karl-Pearson.

Mode = 3 Median – 2 Mean

Illustration:

Compute Mean, Median and Mode for the following distribution.

Class	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50
frequency	2	28	125	270	303	197	65	10

Solution

x	f	x = Mid X	$d = \frac{x-A}{c} = \frac{x-25}{10}$	fd	c.f
10-15	2	12.5	-4	-8	2
15-20	28	17.5	-3	-84	30
20-25	125	22.5	-2	-250	155
25-30	270	27.5	-1	-270	425
30-35	303	32.5	0	0	728
35-40	197	37.5	1	197	925
40-45	65	42.5	2	130	990
45-50	10	47.5	3	30	1000
	N=100			Σfd = -255	

Arithmetic mean $= A + \frac{\sum fd}{N} \times c$

$= 32.5 \ \frac{-255}{1000} \times 5$

$= 32.5 \ \frac{-1275}{1000}$

$= 32.5 - 1.275$

$= 31.225$

Illustration:
Find the missing frequency from the following data

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	5	15	20	--	20	10

The arithmetic mean is 34 marks

Solution

The missing frequency be denoted by m

Marks X	No. of students F	x = Mid X	fx
0-10	5	5	25
10-20	15	15	225
20-30	20	25	500
30-40	m	35	35m
40-50	20	45	900
50-60	10	55	550
	N=70 + m		$\sum fx = 2200+35m$

Arithmetic mean $x = \frac{\sum fx}{N}$

$$= 34 + \frac{2200 + 35\,m}{70 + m}$$

$$34\,(70 + m) = 2200 + 35\,m$$

$$2380 + 34m, = 2200 + 35\,m$$

$$35m - 34m = 2380 - 2200$$

$$m = 180$$

Hence the missing value is 180

Median

$$\frac{N}{2} = \frac{1000}{2} = 500$$

The C.f. just greater than 500 is 728

$$\text{Median} = 1 + \frac{h}{f}\left(\frac{N}{2} - C\right)$$

Wherel – 30
f –
30
3
h – 5

$$\text{Median} = 30 + \frac{5}{303}\left(500 - 425\right)$$

$$= 31.238$$

Mode

The class corresponding to the maximum frequency 303 is 30-35 which is the 'Modal class'.

$$\text{Mode} = 1 + \frac{h\,(f_1 - f_o)}{2\,f_1 - f_o - f_2}$$

Where 1 – 3 (lower limit of the Modal class)

h – 5 (length of the class interval)

f_1 – 303 (frequency of the Modal class)

f_o – 270 (frequency of the class preceding the Modal class)

f_2 – 197 (frequency of the class succeeding the Modal clas

$$\text{Mode} = 1 + \frac{h\ (f_1 - f_o)}{2\ f_1 - f_o - f_2}$$

$$= 30 + \frac{5\ (303\text{-}270)}{606\text{-}207\text{-}197}$$

$$= 43 + \frac{5\ (3)}{139}$$

$$= 31.187$$

Combined Arithmetic Mean

If we have the arithmetic mean and numbers of items of two or more than two related groups, we can compute combined average of these groups by applying the following formula

$$\overline{X_{12}} \quad \frac{n_1\overline{X_1} \quad n_2\overline{X_2}}{n_1 \quad n_2}$$

Where

$\overline{X_{12}}$ - Combined mean of two groups

$\overline{X_1}$ - Arithmetic mean of first group

$\overline{X_2}$ - Arithmetic mean of Second group

n_1 - Number of items in the first group

n_2 - Number of items in the second group

Illustration

The mean height of 25 male workers in a factory is 61 inches and the mean height of 35 female workers in the same factory is 58 inches. Find the combined mean height of 60 workers in the factory.

Solution

Given,

$n_1 = 25$ $\quad$ $n_2 = 35$ $\quad$ $x_1 = 61$ $\quad$ $x_2 = 58$

Combined mean $\overline{X_{12}}$ $\dfrac{n_1\overline{X_1} \quad n_2\overline{X_2}}{n_1 \quad n_2}$

$$= \frac{25 \text{ x } 61 + 35 \text{ x } 58}{25 + 35}$$

$$= \frac{1525 + 2030}{60}$$

$$= \frac{3555}{60} = 59.25$$

Thus the Combined mean height of 60 workers is 59.25 inches.

Illustration

In a certain examination the average grade of all students in class A is 68.4 and students in class B is 71.2. If the average of both classes combined are 70. Find the ratio of the number of students in class A to the number of student in B.

Solution :-

Let us assume that the number of students in class A was X and in class B was Y.

We are given, $x_{12} = 70$ $\quad$ $x_1 = 68.4$ $\quad$ $x_2 = 71.2$

Combined mean $\overline{X_{12}}$ $\dfrac{n_1\overline{X_1} \quad n_2\overline{X_2}}{n_1 \quad n_2}$

$$70 = \frac{n_1(68.4) + n_1(71.2)}{n_1 + n_2}$$

$$70 = \frac{x(68.4) + y(71.2)}{x + y}$$

$$70(x+y) = 68.4x + 71.2y$$

$$7\text{-}0x + 70y - 68.4x - 71.2y = 0$$

$$1.6x - 1.2y = 0$$

$$1.6x = 1.2y$$

Suppose $X = 10$

$$1.6 \times 10 = 1.2y$$

$$1.2y = 16$$

$$y = \frac{16}{1.2}$$

$$y = \frac{40}{3}$$

Thus x and y are in the ratio of $10 : \frac{40}{3}$ or $30 : 40$

Hence for every 3 students in class A, there are 4 students in class B.

Correcting incorrect values:

If sometimes happens that due to an oversight or mistake in copying certain wrong items are taken while calculating mean. The problem is how to find out the correct mean. The process is very simple. From incorrect Ex deduct wrong items and add correct Ex by the number of observations. The result, so obtained, will give the value of correct mean.

Illustration

The mean marks of 100 students were found to be 40. Later on it was discovered that a score of 53 was misread as 83. Find the correct mean corresponding to the correct score

Solution :-

We are given $n = 100$, $x_1 = 40$

Since $x_1 = \frac{\sum x}{n}$

$\sum x = n\, x_1 = 100 \times 40 = 4000$

But this is not correct $\sum x$

Corrected $\sum x$ = Incorrect $\sum x$ – Wrong item + correct item

$= 4000 - 83 + 53$

$= 3970$

Correct x $= \frac{\text{Correct } \sum x}{n}$

$= \frac{3970}{100}$

$= 39.7$

Hence the correct average = 39.7

Geometric Mean

Geometric mean is defined as the n^{th} root of the product of n items.

Geometric mean $= \sqrt[n]{x_1 \times x_2 \times \ldots x_n}$

Where n = number of items

$x_1, x_2, x_3 \ldots\ldots$ = are varies values.

Calculation of geometric mean – individual series:

Steps:

1. Find out the logarithm of each value or the size of item from the log table – log x
2. Add all the values of log x - $\sum \log x$.
3. The sum of log (ie $\sum \log x$) is divided by the number of items ie $\frac{\sum \log x}{}$
4. Find out the antilog of the quotient (from step 3). This is the geometric mean of the data

Illustration :

Compute geometric mean of the following 50, 72, 54, 82, 93

Solution :

x	log x
50	1.6990
72	1.8573
54	1.7324
82	1.9138
93	1.9685
	$\sum \log x = 9.1710$

$$\text{Geometric mean} = \text{antilog}\left(\frac{\sum \log x}{n}\right)$$

$$= \text{antilog}\left(\frac{9.1710}{5}\right)$$

$$= \text{antilog } 1.8342$$

$$= 68.26$$

Calculation of Geometric mean – Discrete series:

Steps:

1. Find out the logarithm of each value - - log x
2. Multiply the log of each size by its frequency – f log x
3. Add all the products thus we get - f log x.
4. Divide the total product by the total frequency N (ie _ f log x)
   n
5. The antilog of the step & is the result (i.e) GM = antilog f log x)
   N

Illustration

The following table gives the weight of 31 persons in a sample survey calculate Geometric mean.

Weight (lbs)	130	135	140	145	146	148	149	150	157
No. of person	3	4	6	6	3	5	2	1	1

Solution

Calculation of Geometric Mean

size of items x	frequency f	log x	f log x
130	3	2.1139	6.3417
135	4	2.1303	8.5212
140	6	2.1461.	12.8766
145	6	2.1614	12.9684
146	3	2.1644	6.4932
148	5	2.1703	10.8515
149	2	2.1732	4.3464
150	1	2.1761	2.1761

157	1	2.1959	2.1959
	N = f 31		f log x = 66.7710

Geometric mean = antilog $\left(\frac{\text{f log x}}{\text{N}} \right)$

= antilog $\left(\frac{}{31} \right)$

= antilog 2.1539

= 142.5 lbs

calculation of Geometric mean

steps :-

1) Find the mid value of each - M

2) Find the logarithm of the mid value – log m

3) Multiply the log of M by their respective frequency f log m

4) Add up all the products f log m

5) Divide f log m by N (i.e) $\frac{\text{f log m}}{\text{N}}$

6) Find out the antilog of the result of step 5 and this will give the answer: GM = antilog $\frac{\text{f log m}}{\text{N}}$)

Find out Geometric mean for the following data:-

Yield of wheat mounds	8-10	11-13	14-16	17-19	20-22	23-25	26-28
No. of person	5	9	19	23	7	4	1

Solution
Calculation of Geometric Mean

X	f	M = Mid x	log m	f log m
7.5-10.5	5	9	0.9542	4.7710
10.5-13.5	9	12	1.0792	9.7128
13.5-16.5	19	15	1.1761	22.3459
16.5-19.5	23	18	1.2553	28.8719
19.5-22.5	7	21	1.3222	9.2554
22.5-25.5	.4	24	1.3802	5.5208
25.5-28.5	1	27	1.4314	1.4314
	N = 68			f log m = 81.9092

$$\text{Geometric mean} = \text{antilog}\left(\frac{f \log m}{N}\right)$$

$$= \text{antilog}\left(\frac{}{68}\right)$$

$$= \text{antilog } 1.045$$

$$= 16.02 \text{ mounds}$$

Harmonic mean:-

Harmonic mean is the reciprocal of the arithmetic average of the reciprocal of values of various items in the variable.

Calculation of Harmonic mean – Individual series
Steps:-

1) Find out the reciprocal of each size (ie \) 1/x
2) Add all the reciprocals of all values ( )
3) Apply the formula : Harmonic Mean $\left(\dfrac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}\right)$

Illustration:

The monthly incomes of 10 families in rupees in a certain village are given below:-

Family	1	2	3	4	5	6	7	8	9	10
Income	85	70	10	75	500	8	42	250	40	36

Solution: Calculation of Harmonic mean

Family	Income X	
1	85	0.01176
2	70	0.01426
3	10	0.10000
4	75	0.01333
5	500	0.00200
6	8	0.12500
7	42	0.02318
8	250	0.00400
9	40	0.02500
10	36	0.02778

$$H.M = \frac{n}{\frac{1}{x}}$$

$$= \frac{10}{0.34631}$$

$$= \text{Rs. } 28.87$$

Calculation of H.M. – discrete series

Steps:-

1) Find out the reciprocal of each size of item 1/x
2) Multiply the reciprocal 1/x of each size by its frequency f (1/x)
3) Add up all the products f (1/x)
4) Apply the formula $H,M = \left(\frac{N}{f(1/x)} \right)$

Illustration :

Calculate H.M. from the following data:

Size of the item	6	7	8	9	10	11
No. of person	4	6	9	5	2	8

Solution

x	F	$\frac{1}{x}$	f (1/x)
6	4	0.1667	0.6668
7	6	0.1429	0.8574
8	9	0.1250	1.1250
9	5	0.111	0.555
10	2	0.1000	0.2000
11	8	0.0909	0.7272
	N = 34		f (1/x) = 4.1319

$$H.M = \left(\frac{N}{f(1/x)} \right)$$

$$= \frac{34}{4.1319}$$

$$= 8.23$$

Calculation of H.M. – continuous series

Steps:-

1) Find out the mid value of each class m
2) Find out the reciprocal of each mid value 1/m
3) Multiply the reciprocal of each Midvale by its frequency f(1/m)

4) Add up all the products f(1/m)

5) Apply the formula $H,M = \left(\frac{N}{f(1/m)} \right)$

Illustration :

Calculate H.M. from the following data :

Marks	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Frequency	15	13	8	6	15	7	6

Solution

X	f	m = Mid x	$\frac{1}{m}$	f (1/m)
30-40	15	35	0.02857	0.42855
40-50	13	45	0.02222	0.28886
50-60	8	55	0.01818	0.14544
60-70	6	65	0.01534	0.09204
70-80	15	75	0.01333	0.19995
80-90	7	85	0.01176	0.08232
90-100	6	95	01.01053	0.06318
				f (1/m) = 1.30034

$$H.M = \left(\frac{N}{f(1/m)} \right)$$

$$= \frac{70}{1.30034}$$

$$= 53.83$$

2.3 Measures of Dispersion

In statistics, statistical dispersion (also called statistical variability or variation) is variability ore spread in a variable or a probability distribution. Common examples of measures of statistical dispersion are the variance, standard deviation and inter-quartile range.

Dispersion is the measure of the variation of the items:-

Methods of measuring dispersion

The following are the important methods of studying variation.

(i) Range

(ii) Quartile deviation

(iii) Mean deviation

(iv) Standard deviation

RANGE:

Range is the simplest dispersion It is defined as the difference between the value of the largest item and the value of the smallest item included in the distribution.

Range $= L - S$

Where L = Largest value

S = Smallest value

The relative measure corresponding to range is called the coefficient of range.

$$\text{Coefficient of range} = \frac{L - S}{L + S}$$

Illustration

Find the range of weights of 7 students from the following

27, 31, 35, 36, 40, 42, 43

Solution:-

L = Largest value = 43

S = Smallest value=27

Range $= L - S$

$= 43 - 27 = 16$

$$\text{Coefficient of range} = \frac{L - S}{L + S} = \frac{43 - 27}{43+27}$$

$= 16/70 = 0.23$

QUARTILE DEVIATION:-

Quartile deviation is an absolute measure of dispersion.

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2}$$

The relative measure of dispersion, known as coefficient of quartile deviation, is calculated as follows :

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Where, $Q_1 =$ Lower Quartile

$Q_3 =$ Upper Quartile

Formula for finding lower quartile Q_1

$$Q_1 = l + \frac{h}{f}\left(\frac{N}{4} - C\right)$$

where, l – lower limit of the Q_1 class

f – corresponding frequency of the Q_1 class h – magnitude of the Q_1 class

e – is the cumulative frequency of the class preceding the Q_1 class.

Formula for finding upper quartile Q_3

where, l – lower limit of the Q_3 class

f – Corresponding frequency of the Q_3 class

h – Magnitude of the Q_3 class

c - Cumulative frequency of the class preceding the Q_3 class.

Illustration :-

Calculate the range and quartile deviation of wages

Wages (Rs)	30-32	32-34	34-36	36-38	38-40	40-42	42-44
Laborers	12	18	16	14	12	8	6

Also calculate the quartile coefficient of dispersion

Solution : Range = L – S = 44 – 30 = 14 Rupees

X	f	c.f	
30-32	12	12	
32-34	18	30	Q_1 Class
34-36	16	46	
36-38	14	60	← Q_3 Class
38-40	12	72	
40-42	8	80	$\frac{N}{4} = \frac{86}{4} = 21.5$
42-44	6	86	(in c.f, just greater than 21.5 is Q_1 30 class)

$$Q_1 = l + \frac{h}{f}\left(\frac{N}{4} - C\right)$$

Where $l = 32$, $f = 18$, $h = 2$, $c = 12$

$$= 32 + \frac{2}{18}(21.5 - 12)$$

$$= 32 + \frac{2}{18}(9.5)$$

$$= 32 + \frac{19}{18}$$

$$= 32 + 1.06$$

$$= 33.06$$

$$\frac{3N}{4} = \frac{3 \times 86}{4} = 64.5$$

(in c.f, just greater than 64.5 is 72 Q_3 class)

$$Q_3 = l + \frac{h}{f}\left(\frac{3N}{4} - C\right)$$

Where $l = 38$, $f = 12$, $h = 2$, $c = 60$

$$= 38 + \frac{2}{12}(64.5 - 60)$$

$$= 38 + \frac{2}{18}(4.5)$$

$$= 38 + 0.75$$

$$= 38.75$$

Quartile Deviation $= \dfrac{Q_3 - Q_1}{2}$

$$= \frac{38.75 - 33.06}{2}$$

$$= \frac{5.69}{2}$$

$$= 2.85$$

Coefficient of Quartile Deviation $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

$$= \frac{38.75 - 33.06}{38.75 + 33.06}$$

$$= \frac{5.69}{71.81}$$

$$= 0.08$$

Illustration :-

Calculate the quartile deviation and coefficient of quartile deviation for the following:-

Age in years	20	30	40	50	60	70	80
Number of members	3	61	132	153	140	51	3

Solution:

x	f	c.f	
20	3	3	
30	61	64	
40	132	196	← Q_1 Class
50	153	349	
60	140	489	
70	51	540	
80	3	543	

Q_1 = Value of $\left(\frac{N+1}{4}\right)^{th}$ item

= Value of $\left(\frac{543+1}{4}\right)^{th}$ item

= Value of $\left(\frac{544}{4}\right)^{th}$ item

= 136^{th} item

= 40 years

(∴ Just greater that 136 is 196 – Q_1 class)

$$Q_3 = \text{Value of } 3\left(\frac{N+1}{4}\right)^{th} \text{ item}$$

$$= \text{Value of } 3\left(\frac{543+1}{4}\right)^{th} \text{ item}$$

$$= \text{Value of } 3\left(\frac{544}{4}\right)^{th} \text{ item}$$

$= \text{Value of } 3 \times 136^{th}$ item

$= \text{Value of } 408^{th}$ item

$=$ 60 years

(Just greater that 408 is 489 – Q_3 class)

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2}$$

$$= \frac{60-40}{2}$$

$$= \frac{20}{2}$$

$=$ 10 years

$$\text{Coefficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$= \frac{60 - 40}{60 + 40}$$

$$= \frac{20}{100}$$

$$= 0.02$$

3. MEAN DEVIATION OR AVERAGE DEVIATION

Mean deviation is the arithmetic mean of the deviation of a series computed from any measure of central tendency, (i.e) mean, media or mode; all the deviation are taken as positive

(i.e) + and – signs are ignored.

Coefficient of mean Deviation:

Mean deviation calculated by any measure of central tendency is an absolute measure.

For the purpose of comparing variation among different units, a relative mean deviation is required.

The relative mean deviation or coefficient of mean deviation is obtained by dividing the mean deviation by the average

$$\text{Mean deviation about A} = \frac{1}{n} \quad |x - A|$$

Where A=mean or median or mode

$$\text{Coefficient of mean deviation} = \frac{\text{Mean deviation about A}}{A}$$

In the case of frequency distribution

$$\text{Mean deviation about A} = \frac{1}{n} \quad |x - A|$$

Illustration

Calculate the mean deviation from mean for the following data

Class interval	2-4	4-6	6-8	8-10
Frequency	3	4	2	1

Solution:

x	f	x=mid x	fx	|x-mean| |x-5.2|	f |x-mean|
2-4	3	3	9	2.2	6.6
4-6	4	5	20	0.2	0.8
6-8	2	7	14	1.8	3.6
8-10	1	9	9	3.8	3.8
	f=10		fx=52		14.8

$$\text{Mean} = \frac{\sum fx}{N}$$

$$= \frac{52}{10}$$

$$= 5.2$$

Mean deviation about $A = \frac{1}{n} \sum |x - A|$ Where

A=Mean

$$= \frac{14.8}{5.2}$$

$$= 1.48$$

Coefficient of mean deviation = $\dfrac{\text{Mean deviation about A}}{\text{Mean}}$

$$= \frac{1.48}{5.2}$$

$$= 0.28$$

Illustration

Find the mean deviation about median for the following distribution

Marks	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45
No. of students	5	6	15	10	5	4	2	2

Also compute the coefficient mean deviation about median of dispersion

Solution:-

We know that,

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - C\right)$$

Where

l – Lower limit of the median class

f – Corresponding frequency of the median class

h – Magnitude of the interval

c - Cumulative frequency of the class preceding median class.

Range = L – S = 44 – 30 = 14 Rupees

x	f	Cf	x=mid x	x –A = x-19.5	| x- A |	f |x-A|
5-10	5	5	7.5	-12	12	60
10-15	6	11	7.5	-7	7	42
15-20	15	26	12.5	-2	2	30
20-25	10	36	17.5	3	3	30
25-30	5	41	22.5	8	8	40
30-35	4	45	27.5	13	13	52
35-40	2	47	32.5	18	18	36
40-45	2	49	37.5	23	23	46
	f=49					f |x-A| =336

Here N = 49, N/2 = 24.5

Thus the median class is 15-20

l = 15, f = 15, h= 5, c = 11

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - C\right)$$

$$= 15 + \frac{5}{15}(24.5 - 11)$$

$$= 15 + \frac{5}{15}(13.5)$$

$$= 15 + \frac{67.5}{15}$$

$$= 15 + 4.5$$

$$= 19.5$$

Mean deviation about A = $\frac{1}{n}$ $|x - A|$ Where

A=Mean

$$= \frac{336}{49}$$

$$= 6.85$$

Coefficient of mean deviation = $\frac{\text{Mean deviation about Mean}}{\text{Mean}}$

$$= \frac{6.85}{19.5}$$

$$= 0.351$$

Standard deviation:

A common measure of dispersion which is preferred in most circumstance in statistics is the standard deviation.

It is denoted by σ and defined as the square root of the average of square of deviation about mean

The formula to find the standard deviation is

$$\sqrt{\frac{(x \quad \bar{x})^2}{n}}$$

In the case of frequency distribution standard deviation

$$\sqrt{\frac{f(x \quad \bar{x})^2}{n}} \qquad f$$

Coefficient of variation = $\frac{\text{Standard deviation}}{}$ X 100

Mean

$$= \frac{\sigma}{\bar{x}} \text{ X } 100$$

Note : The formula of standard deviation can be simplified as,

$$\text{S.D} \quad \sqrt{\frac{1}{n} \; d^2}$$

$$\text{S.D} \quad \sqrt{\frac{1}{n} \; fd^2}$$

Sometime we write the formula as

$$\text{(i)} \quad \text{S.D} \quad \sqrt{\frac{x^2}{n} \quad \frac{x}{n}}^{\,2}$$

$$\text{(ii)} \quad \text{S.D} \quad \sqrt{\frac{1}{N} \; fd^2 \quad \left(\frac{1}{N} \; fd\right)^2}$$

C

For comparing the variability of two sets of a frequency distribution we calculate the coefficient of variation for each of the distribution and that distribution having smaller coefficient of variation is considered as more consistent or more stable or more reliable then the other.

Illustration:

Calculate the mean and standard deviation from the following table giving the age distribution of 542 members.

Age	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of Members	3	61	132	153	140	51	2

Solution :

$$\text{Mean} = A + \frac{fd}{N} \times C$$

$$= 55 + \frac{10}{542} \times (-15)$$

$$= 54.723$$

x	f	x=mid x	$d = \frac{x-A}{C} = \frac{x-55}{10}$	fd	fd^2
20-30	3	25	-3	-9	27
30-40	61	35	-2	-122	244
40-50	132	45	-1	-132	132
50-60	153	55	0	0	0
60-70	140	65	1	140	140
70-80	51	75	2	102	204
80-90	2	85	3	6	18
	f=542				fd^2 =765

σ

$$\sqrt{\frac{1}{N} fd^2 - \left[\frac{1}{N} fd\right]^2}$$

$$= \sqrt{\frac{765}{542} - \frac{-15}{542}^2} \times 10$$

$$= \sqrt{\frac{765}{542} - \frac{225}{542}} \times 10$$

$= \quad 141.067$

$= \quad 11.877$

Illustration :

The scores of two golfers X and Y in 12 rounds are given below. Who is the more consistent player..

X	74	75	78	72	78	77	79	81	79	76	72	71
Y	87	84	80	88	89	85	86	82	82	79	86	80

Solution :

x	$X \quad \overline{X}$ =X-72	$X \quad \overline{X}^2$	Y	$Y - \tilde{Y}$ = Y - 84	$(Y - \tilde{Y})^2$
74	-2	4	87	3	9
75	-1	1	84	0	0
78	2	4	80	-4	16
72	-4	16	88	4	16
78	2	4	89	5	25
77	1	1	85	1	1
79	3	9	86	2	4
81	5	25	82	-2	4
79	3	9	82	-2	4
76	0	0	79	-5	25
72	-4	16	86	2	4
71	-5	25	80	-4	16
912		114	1008		124

$\overline{X} = \frac{x}{n} = \frac{912}{12} = 76$	$\overline{y} = \frac{y}{n} = \frac{1008}{12} = 84$
$\sigma \quad \sqrt{\frac{(x - x^{-})^2}{n}}$	$\sigma \quad \sqrt{\frac{(y - y)^2}{n}}$
$= \frac{114}{12}$	$= \frac{124}{12}$

$= 3.08$ C.V (x) $= \frac{\sigma}{x}$ $= \frac{3.08}{76} \times 100$ $= 4.05$	$= 3.2$ C.V (y) $= \frac{\sigma}{y}$ $= \frac{3.2}{84} \times 100$ $= 3.08$

Since coefficient of variance of the player Y is less. Y is the better player and more consistent player.

Combined Standard Deviation:-

If a sample of n_1 items has mean x_1 and standard deviation σ_1 and another sample has mean x2 and standard deviation σ_2, we can find out the combined mean and combined standard deviation by using the formula,

$$\text{Combined mean } \overline{X_{12}} \quad \frac{n_1\overline{X_1} \quad n_2\overline{X_2}}{n_1 \quad n_2}$$

$$\text{Combined standard deviation} = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$

Illustration:-

The mean of 2 samples of sizes 50 and 100 respectively are 54.1 and 50.3 and standard respectively are 54.1 and 50.3 and standard deviations are 8 and 7. Find the mean and the standard deviation of the sample size of 150.

Solution:-

Given that $n_1 = 50$ $n_2 = 100$ $x_1 = 54.1$ $x_2 =$ 503

$\sigma_1 = 8$ $\sigma_2 = 7$

Combined mean $\overline{X_{12}}$ $\frac{n_1\overline{X_1} \quad n_2\overline{X_2}}{n_1 \quad n_2}$

$$= \frac{50 \times 54.1 + 100 \times 50.3}{20 \times 100}$$

$$= 51.57$$

$d_1 = 54.1 - 51.57 = 2.53$

$d_2 = 50.3 - 51.57 = -1.27$

Combined standard deviation $= \sqrt{\dfrac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$

$= \sqrt{\dfrac{50(\ ^2 + (2.53)^2 + 100(\ ^2 + (-1.27^2)}{50 \times 100}}$

$= 7.5637$

CORRELATION AND REGRESSION

Definition: Correlation

Correlation is an analysis of the relation between two or more variables.

Types of Correlation:

Positive Correlation

Two variables are said to be positively correlated if for an increase in the value of one variable there is a increase in the value of the other variable or for a decrease in one variable, there is a decrease in the other variable. That is, the two variables change in the same direction.

For example, the prices of the product increase due to increase in quality. The salary of employees and the number of years of experience are positively correlated.

Negative correlation

Two variables are said to be negatively correlated if for an increase in the value of one variable there is a decrease in the value of the other variable. That is two variables change in opposite direction.

For example: The number of children who receive Polio drops and the number of children affected by polio are negatively correlated.

Also price and supply are negatively correlated.

No correlation

The variables are said to be uncorrelated if the change in the value of one variable has not effect with the change in the value of the other variable. For example, In an organization the profit earned by the company and the age of the chairman of the company are not correlated.

Karl Pearson coefficient of Correlation

Correlation coefficient between two random variables X and Y, usually denoted by (x,y) is defined as

$$\gamma(x, y) = \frac{Cov(x,y)}{\sigma_x . \sigma_y}$$

Where $\mathrm{cov}(x,y) = \frac{1}{n}\Sigma xy - \bar{x}\,\bar{y}$

$$\sigma_x = \sqrt{\frac{1}{n}\Sigma x^2 - \bar{x}^2}$$

$$\sigma_x = \sqrt{\frac{1}{n}\Sigma y^2 - \bar{y}^2}$$

(OR)

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

The value of r is lies between -1 and +1

1. If r=+1 there is a perfect positive correlation between the two variables.
2. If r=-1 there is a perfect negative correlation between the two variables
3. If r=0 implies there is no correlation between the two variables

Illustration 1

Calculate the correlation coefficient for the following heights (in inches) of fathers x and their son's y.

x	65	66	67	67	68	69	70	73
y	67	68	65	68	72	72	69	71

Solution

$$\bar{X} = \frac{\Sigma x}{n} = \frac{544}{8} = 68$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{552}{8} = 69$$

X	Y	xy	x^2	y^2
65	67	4355	4225	4489
66	68	4488	4356	4624
67	65	4355	4489	4225
67	68	4556	4489	4624
68	72	4896	4624	5184
69	72	4968	4761	5184
70	68	4830	4900	4761
72	71	5112	5184	5041
x=544	y=552	xy=37560	x^2=37028	y^2=38132

$$\sigma_x = \sqrt{\frac{1}{n}\Sigma x^2 - (\bar{x})^2} = \sqrt{\frac{37028}{8} - (68)^2}$$

$$= \sqrt{\frac{37028}{8} - 4624}$$

$$= 2.121$$

$$\sigma_x = \sqrt{\frac{1}{n}\Sigma y^2 - (\bar{y})^2} = \sqrt{\frac{38132}{8} - (69)^2}$$

$$= \sqrt{\frac{38132}{8} - 4761}$$

$$= 2.345$$

$$r(x,y) = \frac{\frac{1}{n}\Sigma xy - \bar{x}\bar{y}}{\sigma_x \sigma_y}$$

$$= \frac{\frac{1}{8}(37560) - (68)(69)}{2.121\, x\, 2.345}$$

$$= 0.6030$$

Illustration 2

Calculate the coefficient correlation between the two firms A and B from the following sales figure in thousands of rupees per week.

X	1	5	8	3	9	10
Y	3	8	10	4	12	11

Solution

x	y	$(x-\bar{x}) = x -$	$(y-\bar{y}) = y -$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y}$
1	3	-5	-5	25	25	25
5	8	-1	0	1	0	0
8	10	2	2	4	4	4
3	4	-3	-4	9	9	12
9	12	3	4	9	9	12
10	N	4	3	16	4	12
x=3 6	y= 36	$\Sigma(x-\bar{x}) = 0$	$\Sigma(y-\bar{y}) = 0$	$\Sigma(x-\bar{x})(y-\bar{y}) =$	$\Sigma(y-\bar{y})^2 =$	$\Sigma(x-\bar{x})^2 = 64$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{36}{6} = 6;\ \bar{y} = \frac{\Sigma y}{n} = \frac{48}{8} = 48;$$

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

$$= \frac{65}{\sqrt{64}\sqrt{70}}$$

$$= 0.97$$

REGRESSION

Regression is a mathematical measure of the average relationship between two or more variables in terms of the original limits of the data.

Lines of Regression

If the variables in a bi-variate distributions are related we will find that the points in the scattered diagram will cluster around some curve called the curve of regression. If the curve is a straight line, it is called the line of regression and there is said to be linear regression between the variables. Otherwise regression is said to be curvilinear.

(h) Line of regression of y on x is given by

$$(x - \bar{x}) = bxy(y - \bar{y})$$

Where by $bxy = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-y)^2}$

= Regression coefficient of y on x

$= \frac{r\sigma_y}{\sigma_x}$

Where $_x$ = standard deviation of x

$_y$ = standard deviation of y

r = correlation of coefficient between x and y

(ii) Line of regression of x on y is given by

$$(x - \bar{x}) = bxy(y - \bar{y})$$

Where $bxy = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-y)^2}$

= Regression coefficient of x on y

$= r\frac{\sigma_x}{\sigma_y}$

Where σ_x=standard deviation of x

σ_y=standard deviation of y

r = correlation coefficient between x and y

Note: Both the lines of regression pass through $(\bar{x}, \bar{y})$

Relation between correlation coefficient and regression coefficient

We know that regression coefficient of y and x $byx = r\frac{\sigma_y}{\sigma_x}$

and regression coefficient of x and y

$$bxy = r\frac{\sigma_x}{\sigma_y}$$

$$\therefore byx \; x \; bxy = r\frac{\sigma_y}{\sigma_x} . r\frac{\sigma_x}{\sigma_y} = r^2$$

$$\therefore r^2 = byx \; x \; bxy$$

$$r = \pm\sqrt{byx \; x \; bxy}$$

Correlation coefficient

$$r = \pm\sqrt{byx . bxy}$$

Illustration

From the following data find,

(i) The two regression equations

(ii) The coefficient of correlation between the marks in Economics and statistics

(iii) The most likely marks in statistics when marks in Economic are 30

Marks in Economics	25	28	35	32	31	36	29	38	34	32
Marks in Statistics	43	46	49	41	36	32	31	30	33	39

Solution

$$\bar{x} = \frac{\Sigma x}{n} = \frac{320}{10} = 32; \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{380}{10} = 38;$$

Marks in Economics x	Marks in Statistics y	$(x-\bar{x}) = x-32$	$(y-\bar{y}) = y-$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
25	43	-7	5	49	25	-35
28	46	-4	8	16	64	-32
35	49	3	11	9	121	33
32	41	0	3	0	9	0
31	36	-1	-2	1	4	2
36	32	4	-6	16	36	-24
29	31	-3	-7	8	49	21
38	30	6	-8	36	64	-48
34	33	2	-5	4	25	-10
32	39	0	1	0	1	0
x=320	y=380			$\Sigma(x-\bar{x})^2 = 14$	$\Sigma(y-\bar{y})^2 = 39$	$\Sigma(x-\bar{x})(y-\bar{y}) = 64$

Regression coefficient of y on x

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

$$= \frac{-93}{140}$$

= -0.6643

Regression coefficient of x on y

$$bxy = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-y)^2} = \frac{-93}{398} = -0.2337$$

(i) Equation of the line of regression of x on y is

$$(x-\bar{x}) = bxy(y-y)$$

x-32 = -0.2337 (y-38)

= -0.2337y+0.2337 x 38

(ie) x=-0.2337y + 8.8806 + 32

x=-0.2237+40.8806

Equation of the line of regression of y on x is

$$(y - \bar{y}) = byx(x - \bar{x})$$

Y-38=-0.6643 (x-32)

=-0.6643x + 0.6643 x 32

(ie) y = -0.6643x + 21.2576+38

y = -0.6643x + 59.2576

(ii) Coefficient of correlation

$$r = \pm\sqrt{byx\ x\ bxy}$$

$$= \pm\sqrt{(-0.6643).(-0.2337)}$$

$$= \pm\sqrt{0.1552}$$

$= \pm 0.394$

(iii) Now we have to find the most likely marks in statistic (y) when marks in economics (x) are 30.

We use the line of regression of y on x

(ie) y=-0.6643x + 59.2576

Put x = 30 we get

y = (-0.6443) x 30 + 59.2596

= -19.329 + 59.2596

= 39.9306

$\simeq$ 39

Illustration

In a correlation study the following values are obtained

	x	y
Mean	65	67
Standard Deviation	2.5	3.5

Coefficient of correlation 0.8

Find the two regression equations that are associated with the above values.

Solution :

By data,

$\bar{x} = 65$ $\bar{y} = 67$

$\sigma_x = 2.5$ $\sigma_y = 3.5$

= 0.8

(i) Line of regression of x on y

$$(x - \bar{x}) = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

$$x - 65 = 0.8x\frac{2.5}{3.5}(y - 67)$$

x- 65 = 0.5714 (y-67)

x=0.5714y – 0.5714 x 64 + 65

x=0.5714y + 26.72

(ii) Line of regression of y on x

$$(y - \bar{y}) = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

$$y-67 = 0.8x\frac{3.5}{2.5}(y - 65)$$

x-67 = 1.12 (x-65)

y-67 = 1.12x – 72.8

y=1.12x – 72.8+67

y=1.12x – 5.8

Illustration

In a partially destroyed laboratory record of an analysis of correlation data, the following results only are legible.

Variance of x=9

Regression equation: 8 x- 10y+66=0

40x-18y = 214

Find on the basis of the above information

(i) The mean values of x and y

(ii) Coefficient of correlation between x and y

(iii) Standard deviation of y

(i) The mean values of x and y

$$2x\text{-}10y\text{=-}66 \qquad (1)$$

$$40x\text{-}18t\text{=}214 \qquad (2)$$

(1) x 5 $\quad 40x\text{-}50y\text{=-}30$

(2) $\quad 40x\text{-}183y = 214$

Subtracting $\quad -32y = -554$

$$y = \frac{554}{31} = 17$$

$$y = 117 \text{ or } \bar{y} = 17$$

Substituting the value of y in equation (1)

$$8x\text{-}10y\text{=-}66$$

$$8x\text{-}10x17 = -66$$

$$8x\text{=-}66+170$$

$$8x\text{=}104$$

$$x = \frac{104}{8} = 13$$

$$x\text{=}13 \text{ or } \bar{x} = 13$$

Mean value of $x\text{=}\bar{x} = 13$; Mean value of $y\text{=}\bar{y}\text{=}17$,

(ii) For finding out the correlation coefficient, we will have to findout the regression coefficient.

Let us take equation (1) as the regression equation of x and y

$$8x\text{-}10y\text{=-}66$$

$$7x\text{=+}10y\text{-}66$$

$$x = \frac{10y - 66}{8}$$

$$x = \frac{10}{8}y - \frac{66}{8}$$

bxy=y coefficient $= \frac{10}{8} = 1.25$

And take equation (2) as the regression equation of y on x.

$$40x\text{-}18y = 214$$

$$-18y = 214\text{-}40x$$

$$y = \frac{-40x+214}{-18}$$

$$y = \frac{40}{18}x \frac{-214}{18}$$

byx = x coefficient $= \frac{40}{18} = 2.22$

$$\therefore r = \sqrt{byx\ x\ bxy} = \sqrt{1.25\ x\ 2.22} > 1$$

Our assumption is wrong.

Hence we take, the equation (i) is taken as y on x.

8x-10y=-66

-10y = -8x-66

$$y = \frac{-8x-66}{1-10}$$

$$y = \frac{8}{10}\ x + \frac{66}{10}$$

The equation (2) is taken as regression eqn. of x in y

40x-18t = 214

40x=18y+214

$$x = \frac{18y+214}{40}$$

$$x = \frac{18}{40}y + \frac{214}{40}$$

bxy = y coefficient $= \frac{18}{40}$

$$\therefore r = \sqrt{bxy\ x\ byx} = \sqrt{\frac{18}{40}x\frac{18}{10}}$$

$$= \sqrt{36}$$

$$= 0.6$$

R=06

Given variance of x=9

$$\sigma_x^2 = 9$$

$$\sigma_x = \sqrt{9} = 3$$

Regression coefficient of x on y

$$bxy = r\frac{\sigma_x}{\sigma_y}$$

$$\frac{18}{40} = \frac{0.6 \ x\ 8}{\sigma_y}$$

$$0.45 = \frac{(0.6)^3}{\sigma_y}$$

$$\sigma_y = \frac{1.8}{0.45}$$

$$\sigma_y = 4$$

Standard deviation of y=4

SPREARMAN'S RANK CORRELATION

In chapter on correlation and regression analysis, we introduced a collection of descriptive and inference techniques for analyzing the relation between two variables, a less formal non-parametric approach is often used to measure the strength of the relation between two variables. The spearman's rank correlation co-efficient is a measure of association based on the ordinal feature of data.

Among the various statistical methods based on ranks, the spearman's rank correlation procedure was the earliest to be developed. This method is simple to use and easy to apply. Also it has proved to be almost as powerful as its classical counter part the Pearson's produce moment correlation method under conditions favourable to the latter and even more powerful than the parametric method when its assumptions are violated.

The spearman's coefficient of rank correlation R is given by the formula.

$$R = 1 - \frac{6\Sigma D^2}{N^3 - N}$$

The coefficient of rank correlation is a relative measure which varies from -1 to +1.

The standard error of the rank correlation coefficient is given by

The spearman rank correlation coefficient may be employed as a test statistic to test a hypothesis of no association between the two populations. We assume that pairs of observations (x_i, y_i) have been randomly selected and therefore the hypothesis of no association between the populations implies a random assignment to ranks within each sample.

Each random assignment (for the two samples) represents a sample points associated with the experiment and a value of R_1 (stringer relationship between X and Y) could be calculated for each. Thus it is possible to calculate the probability that R_1 assumes a large positive or negative value due solely to chance and thereby suggests an association between populations when done exists.

SPEARMAN'S LIMITATIONS OF NON-PRAMETRIC TESTS

1. It should be noted that the statistical methods which require no assumptions about the populations for which we are sampling are usually less efficient (or powerful) than the corresponding standard techniques. Assertions made with equal confidence require larger samples if they are made without knowledge of the form of the underlying distribution than if they are made with such knowledge.

   It is generally true that the more we assume, the more we can infer from a sample. However, at the same time the more we assume the more. We have to be cautious about the applicability of the methods and the inferences based on them, it is disadvantages to use the non-parametric methods when all the assumptions of the classical produces can be not and the data are measured either on interval on ratio scale.

2. As the sample size gets larger, data manipulations required for non-parametric procedures are sometimes laborious unless appropriate computer software is available.
3. A collection of tabulated critical values for a variety of non-parametric tests under situations dealing with small and large n is not readily available.

Rank Correlation

Spearman Rank correlation

$$\text{Coefficient } R = 1 - \frac{6\Sigma D^2}{N^3 - N}$$

Where $D=R_1-R_2$

De denotes difference between the ranks for each pair of observations.

Illustration

Find the rank correlation coefficient from the following data.

Ran in X	1	2	3	4	5	6	7
Rank in Y	4	3	1	2	6	5	7

Solution

R_1=Rank in X	R_2=Rank in X	$D=R_1-R_2$	D^2
1	4	-3	9
2	3	-1	1
3	1	2	4
4	2	2	4
5	6	-1	1
6	5	1	1
7	7	0	0
			$D^2=20$

Ranks correlation coefficient $R = 1 - \frac{6\Sigma D^2}{N^3 - N}$

$= 1 - \frac{6x20}{7^3 - 7}$

$= 1 - \frac{120}{336}$

$= 1 - 0.3571$

$= 0.6429$

Illustration 4

Ten competitors in a musical test were ranked by the 3 judges x,y,z in the following order

	A	B	V	D	E	F	G	H	I	J
Rank by x	1	6	5	10	3	2	4	9	7	8
Rank by y	3	5	8	4	7	10	2	1	6	9
Rank by z	6	4	9	8	1	2	3	10	5	7

Using rank correlation method, discuss which pair of judges has the nearest approach to common links of music.

Solution

R_1 = Rank by x

R_2 = Rank by y

R_3 = Rank by z

Rank by x R1	Rank by y R2	Rank by z R3	$D_1 = R_1 - R_2$	$D_2 = R_2 - R_3$	$D_3 = R_1 - R_3$	D_1^2	D_2^2	D_3^2
1	3	6	-2	-3	-5	4	9	25
6	5	4	1	1	2	1	1	4
5	8	9	-3	-1	-4	9	1	16
10	4	8	6	-4	2	36	16	4
3	7	1	-4	6	2	16	36	4

2	10	2	-8	8	0	64	64	0
4	2	3	2	-1	1	4	1	1
9	1	10	8	-9	-1	64	81	1
7	6	5	1	1	2	1	1	4
8	9	7	-1	2	1	1	4	1
						$\Sigma D_1^2 = 2$	$\Sigma D_2^2 = 2$	$\Sigma D_3^2 =$

Rank Correlation between R_1 and R_2

$$R = 1 - \frac{6\Sigma D_1^2}{N^3 - N}$$

$$= 1 - \frac{6x200}{1000-16}$$

$$= 1 - \frac{1200}{990}$$

$$= 1\text{-}1.212$$

$$= -0.212$$

Rank correlation between R_2 and R_3

$$R = 1 - \frac{6\Sigma D_2^2}{N^3 - N}$$

$$= 1 - \frac{6x214}{1000-10}$$

$$= 1 - \frac{1284}{990} = 1 - 1.296 = -0.296$$

The rank correlation coefficient between R_1 and R_3

$$R = 1 - \frac{6\Sigma D_3^2}{N^3 - N}$$

$$= 1 - \frac{6x60}{1000-16}$$

$$= 1 - \frac{360}{990}$$

$$= 1\text{-}0.3636$$

$$= 0.6364$$

Since the rank correlation between R1 and R3 is maximum and also positive, we conclude that the pair of find x and z has the nearest approach to common links in music.

Illustration 5

Find the rank correlation coefficient for the following data

Marks in Tamil	39	65	62	90	82	75	25	98	36	78
Marks in English	47	53	58	86	62	68	60	91	51	84

Solution

Let x = Marks in Tamil

y = Marks in English

R_1 = Rank in x

R_2 = Rank in y

Marks in Tamil x	**Marks in English x**	**Rank in x R_1**	**Rank in y R_2**	**$D=R_1-R_2$**	**D^2**
39	47	7	10	-3	9
65	53	5	8	-3	9
62	58	6	7	-1	1
20	86	10	2	8	64
82	62	2	5	-3	9
75	68	4	4	0	0
25	60	9	6	3	9
98	91	1	1	0	0
36	51	8	9	-1	1
78	84	3	3	0	0
					D^2=102

Rank correlation coefficient

$$R = 1 - \frac{6\Sigma D^2}{N^3 - N}$$

$$= 1 - \frac{6x102}{1000-10}$$

$$= 1 - \frac{612}{990}$$

$$= 1\text{-}0.618$$

$$= 0.382$$

Repeated Ranks

In any two or more individuals are equal in any classification with respect to characteristic. A or B or if there is more than one item with the same value in the series then spear's man formula for calculating the rank correlation coefficients breaks down. In this case common ranks are given for repeated ranks. This common rank is the average of the ranks which these items would have assumed if they are slightly different from each other and the next item will get the ranks next to the ranks already assumed. As a result of this, following adjustment or correction is made in the correlation formula.

In the correction formula, we added the factor $\frac{m(m^2-1)}{12}$ to d^2 where m is the number of item is repeated. This correction factor is to be added for each repeated values.

Illustration 6

A sample of 12 fathers and their eldest sons have the following data about their heights in inches.

Father Height	65	63	67	64	68	62	70	66	68	67	69	71
Sons Height	68	66	68	65	69	66	68	65	71	67	68	70

Calculate the rank correlation coefficient

Let x = fathers height

y = sons height

R_1 = Rank in father's height

R_2 = Rank in Son's height

x	y	R_1	R_2	$D=R_1-R_2$	D_2
65	68	9	5.5	3.5	12.25
63	66	11	9.5	1.5	2.25
67	68	6.5	5.5	1	1
64	65	10	11.5	-1.5	2.25
68	69	4.5	3.	1.5	2.25
62	66	12	9.5	2.5	6.25
70	68	2	5.5	-3.5	12.25
66	65	8	11.5	-3.5	12.25
68	71	4.5	1	3.5	12.25
67	67	6.5	8	-1.5	2.25
69	68	3	5.5	-2.5	6.25
71	70	1	2	-1	1
					$D^2=72.5$

Correction Factors

In x series 68 is repeated twice

$$C.F = \frac{m(m^2-1)}{12} = \frac{2(2^2-1)}{12} = \frac{2(4-1)}{12} = \frac{2x3}{12} = \frac{1}{2}$$

In x series 67 is repeated twice

$$C.F = \frac{m(m^2-1)}{12} = \frac{2(2^2-1)}{12} = \frac{1}{2}$$

In y series 68 is repeated four times

$$C.F = \frac{m(m^2-1)}{12} = \frac{4(4^2-1)}{12} = \frac{4(16-1)}{12} = \frac{4x15}{12} = 5$$

In y series 66 is repeated twice

$$C.F = \frac{m(m^2-1)}{12} = \frac{2(2^2-1)}{12} = \frac{1}{2}$$

Sum of correction factors $= \frac{1}{2} + \frac{1}{2} + 5 + \frac{1}{2} + \frac{1}{2} = 7$

Rank correlation coefficient

$$R = 1 - \frac{6[\Sigma D^2 + sum\ of\ correction\ factors]}{N^3 - N}$$

$$= 1 - \frac{6[72.5+7]}{12^3-12}$$

$$= 1 - \frac{6[79.5]}{1728-12}$$

$$= 1 - \frac{477}{1716}$$

$$= 1\text{-}0.277$$

$$= 0.722$$

SUMMARY

Statistical methods are frequently used for taking good and reliable decisions in management. On the way the measures of central tendency and measures of dispersions are the fundamental platforms for getting more knowledge on correlation, regression and various statistical measurements. This chapter described about various measures of central tendency and dispersion. A measure of location is a value used to describe the centre of a set of data. The arithmetic mean is the most widely reported measure of location.

KEYWORDS

The keywords of this unit are

Mean

Median

Mode

Harmonic mean

Geometric mean

Correlation

Rank correlation

Regression

Check Your Progress Questions:

1. Calculate the rank correlation coefficient for the following data

Person	A	B	C	D	E
Rank in income	1	4	2	3	5
Battle axe rank	3	1	2	5	4

2. Compute spearman's rank correlation coefficient for the following data

Marks in statistics	45	60	72	62	56	40	39	52	30
Marks in Accountancy	62	78	65	70	38	54	60	32	21

3. Two judges in a beauty competition rank for the 12 entires as follows :

X	1	2	3	4	5	6	7	8	9	10	11	12
Y	12	9	6	10	3	5	4	7	8	2	11	1

4. Calculate coefficient of correlation for the following data

Marks in statistics	45	56	39	54	45	40	56	60	30	36
Marks in Maths	40	36	30	44	36	32	45	42	20	36

5. Find the coefficient of correlation between industrial production and export using the following data.

Production x	55	56	58	59	60	60	62
Export y	35	38	37	39	44	43	44

6. The two regression equations of the variable x and y are $x=19.93-0.87y$ and $y = 11.64 - 0.50x$ find

(i) Mean of x (ii) Mean of y

(iii) Correlation between x and y

7. Find the regression line of y and x if

x	1	4	2	3	5
y	3	1	2	5	4

8. Find the regression lines for the following data

x	40	70	50	60	80	50	90	40	60	60
y	2.5	6.0	4.5	5.0	4.5	2.0	5.5	3.0	4.0	3.0

9. Heights of fathers and sons are given in centimeters

x:Heights of father	150	152	155	157	160	161	164	166
y : Height of son	154	156	158	159	160	162	161	164

Find the two regression and calculate the expected average height of the son when the height of the father is 164.

Questions/ Exercise

1. What is meant by measures of dispersion?
2. Explain the characteristic of Karl pearson coefficient of correlation.
3. Enlist the uses of regression.
4. Differentiate regression and correlation.

UNIT 3 : PROBABILITY AND THEORITICAL DISTRIBUTIONS

INTRODUCTION

Let a coin be tossed. Nobody knows what will get: whether a head or tail. But it is certain that either a head or tail will occur. In a similar way, if a dice is thrown, we many get any of the faces 1, 2,3,4,5 and 6. But nobody knows which one will actually occur experiments of this type where the outcomes cannot be predicted are called 'random' experiments.

The word probability or chance is used commonly in day to day life. For example the chances of India and Pakistan winning the world cup cricket, before the start of the game are equal (ie 50:50). It is likely that Mr. Anbu may not come for taking his class today. We often say that it is very probable that it will rain tomorrow. Probably Mr. Anu will not come to tea party tomorrow. All these terms, chance, likely, probable etc., convey the same meaning, i.e. that event is not taken to take place. In other words, there is an uncertainty about the happening of the event.

In such of these cases we talk about chance or probability which is taken to be a quantitative measure of certainty.

The term probability refers to the study of randomness and uncertain. Before going to study the mathematical definition of probability we will define some terms of given below.

UNIT OBJECTIVES

After completing this unit, you should be able to calculate the probability of an event

to apply addition theorem of probability

to apply multiplication of probability

to calculate the expected value

to calculate the probability of an event using binomial distribution

to calculate the probability of an event using poisson distributor.

UNIT STRUCTURE

Mathematical probability

Calculation of probability

Addition theorem of probability

Multiplication of probability

Conditional probability

Expectation

Binomial distribution

Poisson distribution

Normal distribution

Exercise

TRIAL AND EVENT

Consider an experiment of throwing a coin. When tossing a coin, we may get a Head (H) or Tail (T).

EXHAUSTIVE EVENTS

The total number of possible outcomes in any trial is known as exhaustive events.

Example

In tossing a coin the possible outcomes are getting a head or tail.

Hence we have 2 exhaustive events in throwing a coin.

MUTUALLY EXCLUSIVE EVENTS

Two events are said to be mutually exclusive when the occurrence of one affects the occurrences of the other.

Example :

In tossing a coin the events head or tail are mutually exclusive, since both tail and head cannot appear in the same time.

EQUALLY LIKELY EVENTS

Two events are said to be equally likely if one of them cannot be expected in preference to the other.

Example :

1. In tossing a coin, head or tail are equally likely events.
2. In throwing a die, all the six faces are equally likely events.

INDEPENDENT EVENTS

Two events are said to be independent when the actual happening of one does not influence in any way the happening of the other.

Example : In tossing a coin, the event of getting a head in the first toss is independent of getting a head in the second toss, third toss, etc.

Mathematical or Apriori definition of probability

Let S be a sample space (the set of all possible outcomes which are assumed equally likely) and A be an event (a sub-set of S consisting of possible outcomes) associated with a random experiments Let N(S) and N(A) be the number of elements of S and A. Then the probability of event A occurring denoted as P(A), is defined by

$$P(A) = \frac{Number\ of\ cases\ favourable\ to\ A}{Exhaustive\ number\ of\ cases\ in\ S}$$

$$= \frac{n(A)}{n(S)}$$

Illustration 1 :

From a bag containing 10 red and 20 white balls, a ball is drawn at random. What is the probability that it is red.

Solution :

Total number of balls in the bag = 10+20 = 30

Number of red balls = 10

Let A be the event of drawing a red ball

$$\therefore P(A) = \frac{Number\ of\ favourable\ cases}{Total\ number\ of\ equally\ likely\ cases}$$

$$= \frac{10}{30}$$

$$= \frac{1}{3}$$

Illustration 2 :

Find the probability of throwing : (a) 4 (a) an odd number (c) an even number with an ordinary die.

Solution :

Sample Space = S = {1,2,3,4,5,6}

a) When throwing a die there is only one way of getting 4.

$$\therefore P(getting\ 4) = \frac{number\ of\ favourable\ events}{number\ of\ exhansive\ events}$$

$$= \frac{1}{6}$$

b) Sample space S = {1,2,3,4,5,6}

n(S) = 6

Let A be event of getting an odd number

A = {1,3,5}

n(A) = 3.

$$P(A) = \frac{number\ of\ favourable\ cases}{Total\ number\ of\ equally\ likely\ cases}$$

$$= \frac{n(A)}{n(S)}$$

$$= \frac{3}{6}$$

$$= \frac{1}{2}$$

c) Sample space S={1,2,3,4,5,6}

n(S) = 6.

Let B be event of getting an ever number

B = {2,4,6}

n(B) = 3

$$P(B) = \frac{number\ of\ favourable\ cases}{Total\ number\ of\ equally\ likely\ cases}$$

$$= \frac{n(B)}{n(S)}$$

$$= \frac{3}{6}$$

$$= \frac{1}{2}$$

Illustration 3

What is the probability that a leap year selected at random will contain 53 Sundays?

Solution : A leap year consists of 366 days.

366 days = 52 full weeks + 2 days

= 52 x 7 + 2 days

(52 full weeks we have 52 Sunday definitely)

These two days may be

i. Monday, Tuesday

ii. Tuesday, Wednesday

iii. Wednesday, Thursday

iv. Thursday, Friday

v. Friday, Saturday

vi. Saturday, Sunday

vii. Sunday, Monday

of these 7 cases the last two cases contain Sunday and hence we have 2 favourable cases.

$$\therefore \textit{The required probability} = \frac{\textit{Number of favourable cases}}{\textit{Total number of equally liked cases}}$$

$$= \frac{2}{7}$$

AXIOMATIC APPROACH TO PROBABILITY

Let S be the sample space and A be an event associated with a random experiment. Then the probability of the event A, denoted by P(A) is defined as a real number satisfying the following axioms.

i) $O \leq P(A) \leq 1$

ii) $P(S)=1$

iii) If $A_1, A_2, \ldots A_n$ are a set of 'n' nutually exclusive events,

$$P(A_1 \quad A_2 \quad \ldots \quad A_n) = P(A_1) + P(A_2) + \ldots\ldots + P(A_n)$$

Results

1. Probability of impossible event is zero. (ie) P( )=0
2. Probability of the complement event $\bar{A}$ of A is given by

$P(\bar{A}) = 1 - P(A)$

Illustration 4 :

A ball is drawn at random from a box containing 6 red balls, 4 white balls and 5 blue balls. Determine the probability that it is (i) Red (ii) White (iii) Blue (iv) not red.

Solution :

$$Probability\ of\ an\ event = \frac{Number\ of\ favourable\ cases}{Total\ number\ of\ equally\ likely\ cases}$$

(i) $P(a\ red) = \frac{6}{15} = 0.40$

(ii) $P(a\ white) = \frac{4}{15} = 0.267$

(iii) $P(a\ blue\ ball) = \frac{5}{15} = 0.333$

(iv) $P(not\ red) = 1 - P(a\ red)$

$= 1 - 0.40$

$= 0.60$

Illustration 5

The probability that A can solve a. Problem is 0.7 and the probability that B can solve that problem is 0.6. Considering that these two events are independent, Find the probability that the problem gets solved by either of them.

Solution

$P(A) + P(\bar{A}) = 1$

P(success) + P(failure) = 1

P(failure) = 1-P(success)

P(A tails to solve the problem) = 1-P(A solve the problem)

= 1-0.7

= 0.3

P (B fails to solve the problem) = 1-P(B solve the problem)

= 1-0.6

= 0.4

Since the events are independent, the probability that both fail to solve the problem

= 0.3 x 0.4

= 0.12

The probability that the problem will be solved = 1-0.12

= 0.88

Illustration 6

A bag contains 6 white, 4 red and 10 black balls. Two balls are drawn at random. Find the probability that they will both be black.

Solution :

Total number of balls in the bag = 6+4+10 = 20

Two balls are drawn at random

$$P(both\ will\ be\ black) = \frac{10C_2}{20C_2}$$

$$= \frac{\frac{10x9}{1x2}}{\frac{20x19}{1x2}}$$

$$= \frac{\frac{90}{2}}{\frac{380}{2}}$$

$$= \frac{90}{2} \times \frac{2}{380}$$

$$= \frac{9}{38}$$

$$= 0.237$$

THEOREMS OF PROBABILITY

There are two important theorems of probability. They are

1. The addition theorem of probability
2. The multiplication theorem of probability

1. Addition theorem of probability

(a) The addition theorem states that 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.

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

The theorem can be extended to three or more mutually exclusive events

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

(b) For any two events A and B, the probability that either A or B occurs is given by the sum of individual probabilities of A and B less the probability of simultaneous occurrence of the events A and B.

$$P(A \quad B) = P(A) + P(B) - P(A \quad B)$$

(i.e) P(A or B) = P(A) + P(B) – P(A and B)

(c) Multiplication theorem of probability

This theorem states that if two events A and B are independent, the probability that they both will occur is equal to the product of their individual probability.

Symbolically, If A and B are independent

Then P(A and B) = P(A) x P (B)

The theorem can be extended to three or more independent events. Thus

$$P(A,B \text{ and } C) = P(A) \times P(B) \times P(C)$$

Illustration

One card is drawn from a standard pack of 52. What is the probability that it is either a king or a queen.

Solution

Total number cards in a pack = 52 cards n(S)=52

Let A be the event of drawing a king n(A) = 4.

$$P(drawing\ a\ king) = P(A) = \frac{n(A)}{n(S)} = \frac{4}{52}$$

Since the events are mutually exclusive.

$$P(A \text{ or } B) = P(A) + P(B)$$

$$= \frac{4}{52} + \frac{4}{52}$$

$$= \frac{4+4}{52}$$

$$= \frac{8}{52} = \frac{2}{13}$$

Illustration 8

The managing committee of vaishalli welfare association formed a sub-committee of 5 person to look into electricity problem. Profiles of the 5 persons are

1. Male age 40
2. Male age 43
3. Female age 38
4. Female age 27
5. Male age 65

If a chairperson has to be selected from this, what is the probability that he would be either female or over 30 years

Solution

$$P(A) = P(female) = \frac{2}{5}$$

$$P(B) = P(over\ 30) = \frac{4}{5}$$

$$P(A\ and\ B) = P(female\ and\ over\ 30)$$

$$= \frac{1}{5}$$

By addition theorem of probability

P(A or B) = P(A) + P(B) – P(A and B)

P(female or over 30) = P(female) + P(over 30) – P (female and over 30)

$$= \frac{2}{5} + \frac{4}{5} - \frac{1}{5}$$

$$= \frac{2+4-1}{5} = \frac{5}{5}$$

$$= 1$$

Illustration 9

A coin is tossed thrice. What is the probability of getting 2 or more heads.

Solution : A coin is tossed three times, then we have

S=Sample space = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

$$P(getting\ 2\ heads) = \frac{3}{8}$$

$$P(getting\ 3\ heads) = \frac{1}{8}$$

As A and B are mutually exclusive,

P (getting 2 or more heads)

$$= P(\textit{getting 2 heads or 3 heads})$$

$$= P(2 \textit{ heads} + P\ (3 \textit{ heads})$$

$$= \frac{3}{8} + \frac{1}{8} = \frac{3+1}{8} = \frac{4}{8} = 0.5$$

Illustration 10

A bag contains 30 balls numbered from 1 to 30. One ball is drawn at random. Find the probability that the number of the ball drawn will be multiple of (a) 5 or 7 (b) 3 or 7

Solution : Total number of balls = 30 in a bag

(a) P(A) = P(number being multiple of 5)

= P(5,10,15,20,25,30)

$$= \frac{6}{30}$$

P(B) = P(number being multiple of 7)

= P(7,14,21,28)

$$= \frac{4}{30}$$

Since the events are mutually exclusive, by addition theorem of probability

P(A or B) = P (multiple of 5 or multiple of 7)

= P(A) + P(B)

$$= \frac{3}{30} + \frac{4}{30}$$

$$= \frac{6+4}{30}$$

$$= \frac{10}{30}$$

(b) P(A) = P(number being multiple of 3)

= P(3,6,9,12,15,18,21,24,27,30)

$$= \frac{10}{30}$$

P(B) = P(number being multiple of 7}

= P(7,14,21,28)

$= \frac{4}{30}$

Since Z1 is a multiple of 3 wells as 7

P(A and B) = P(multiple of 3 and multiple of 7)

$= \frac{1}{30}$

By addition theorem of probability,

P (multiple of 3 or multiple of 7)

$= P(A\ or\ B)$

$= P(A) + P(B) - P(A\ and\ B)$

$= \frac{10}{30} + \frac{4}{30} - \frac{1}{30}$

$= \frac{10 + 4 - 1}{30}$

$= \frac{13}{30}$

Illustration 11

The probability that a boy will get a scholarship is 0.9 and that a girl will get is 0.8. What is the probability that atleast one of them will get the scholarship?

Solution :

P(A) = P(a boy will get a scholarship) = 0.9

P(B) = P(a girl will get a scholarship) = 0.8

P(A and B) = P(A) x P(B)

= 0.9 x 0.8

= 0.72

By addition theorem of probability

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

= 0.9 + 0.8 – 0.72

$= 1.7 - 0.72$

$= 0.98$

Illustration 12

A class consists of 80 students, 25 of them are girls and 55 boys, 10 of them are rich and remaining poor, 20 of them are fair complexioned what is the probability of selecting a fair complexioned rich girl?

Solution : Number of students in a class = 80

$P(A) = P(selecting\ a\ fair\ complexioned\ person) = \frac{20}{80}$

$P(B) = P(selecting\ a\ rich\ person) = \frac{10}{80}$

$P(C) = P(selecting\ a\ girl) = \frac{25}{80}$

Since the events are independent, by multiplication theorem of probability P(Selecting a fair, complexioned and rich girl)

$= P(A \text{ and } B \text{ and } C)$

$= P(A) \times P(B) \times P(C)$

$= \frac{20}{80} \times \frac{10}{80} \times \frac{25}{80}$

$= 0.0098$

Illustration 13

A university has to select and examiner from a list of 50 persons, 20 of them women and 30 men, 10 of them knowing Hindi and 40 or not, 15 of them being teachers and the reaming 35 not. What is the probability of the university selecting a Hindi knowing women Teacher?

Solution

Total number of person in a list = 0

P(A) = P(Selecting a Hindu knowing candidate)

$$= \frac{10}{50}$$

P(B) = P (selecting a woman)

$$= \frac{20}{50}$$

P(C) = P (selecting a teacher)

$$= \frac{15}{50}$$

Since the events are independent, by multiplication theorem of probability

P(selecting Hindi knowing woman teacher)

= P(A and B and C)

= P(A) x P(B) x P(C)

$$= \frac{10}{50} \times \frac{20}{50} \times \frac{15}{50}$$

$$= \frac{3}{125}$$

= 0.024

Illustration 14

A man wants to marry a girl having qualities; where complexion – the probability of such a girl is one in twenty; handsome dowry-the probability of getting this is one is fifty; Westernized manners – the probability here is one in hundred. Find out the probability of his getting married to such a girl when the possession of these three attributes are independents.

Solution

$$P(A) = P(a\ girl\ with\ white\ complexion) = \frac{1}{20} = 0.05$$

$$P(B) = P(a\ girl\ with\ handsome\ dowry) = \frac{1}{50} = 0.02$$

$$P(A) = P(a\ girl\ with\ westernised\ manners) = \frac{1}{100} = 0.01$$

Since the events are independent, by multiplication theorem of probability

P(getting all these qualities)

= P(A and B and C)

= P(A) x P(B) x P(C)

= 0.05 x 0.02 x 0.01

= 0.00001

Illustration 15

The probability that machine A will be performing an usual function in 5 years time is 1/4 while the probability of machine B will still be operating usefully at the end of the same period is 1/3. Find the probability that both machines will be performing an usual function.

Solution

$$P(A) = P(machine\ A\ operating\ usefully) = \frac{1}{4}$$

$$P(A) = P(machine\ B\ operating\ usefully) = \frac{1}{3}$$

By multiplication theorem of probability, P(both A and B will operate usefully)

$$= P(A\ and\ B)$$

$$= P(A) x P(B)$$

$$= \frac{1}{4} x \frac{1}{3}$$

$$= \frac{1}{12}$$

Illustration 16

A bag contains 8 white and 10 black balls. Two balls are drawn is succession. What is the probability that first is white and second is black.

Solution

Total number of balls in a bag = 8+10=18

$P(A) = P(drawing\ a\ white\ ball) = \frac{8}{18}$

$P(B) = P(drawing\ a\ black\ ball) = \frac{10}{18}$

Since the events are independent, by multiplication theorem of probability

$$\begin{aligned} P(A \text{ and } B) &= P(A) \times P(B) \\ &= \frac{8}{18} x \frac{10}{18} \\ &= \frac{80}{18} \\ &= 0.247 \end{aligned}$$

Illustration 17

A problem is mathematics is given to 3 students A,B,C whose chances of solving it are $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}$ respectively. What is the probability that the problem will be solved?

Solution

$P(A\ will\ solve\ the\ problem) = \frac{1}{2}$

$P(A\ will\ not\ solve\ the\ problem) = 1 - \frac{1}{2} = \frac{1}{2}$

$P(B\ will\ solve\ the\ problem) = \frac{1}{3}$

$P(B\ will\ not\ solve\ the\ problem) = 1 - \frac{1}{3} = \frac{2}{3}$

$P(B\ will\ solve\ the\ problem) = \frac{1}{4}$

$P(B\ will\ not\ solve\ the\ problem) = 1 - \frac{1}{4} = \frac{3}{4}$

$$P(all\ the\ three\ will\ not\ solve\ the\ problem) = \frac{1}{2} x \frac{2}{3} x \frac{3}{4}$$

$$= \frac{1}{4}$$

P(all the three will solve the problem)

= 1-P(all the three will not solve the problem)

$$= 1 - \frac{1}{4}$$

$$= \frac{4-1}{4}$$

$$= \frac{3}{4}$$

Illustration 18

Give two independent events A and B such that P(A)=0.2 and P(B)=0.5 Find

(i) P(A and B) (ii) P(A or B) (iii) P(A and Not B)

Given P(A) = 0.2

P(B) = 0.5

Since A and B are independent

(i) P (A and B)	=	P(A) x P(B)
	=	0.2 x 0.5
	=	0.1
(ii) P (A or B)	=	P(A) + P(B) – P(A and B)
	=	0.2 x 0.5-0.1
	=	0.7-0.1
	=	0.6
(iii) P (not B)	=	1-P(B)
	=	1-0.5
	=	0.5
P (A and not B)	=	P(A) x P(not B)
	=	0.2 x 0.5
	=	0.1

Definition

The conditional probability of A given that B is defined as

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}$$

2(b) multiplication rule of probability

The probability of simultaneous occurrence of two events A and B denoted by P(A B) and defined by

P(A B) = P(A) x P(B/A)

Illustration A

From a bag containing 3 red and 4 black balls two balls are drawn in succession, without replacement. Find the probability that both the balls are (i) red (ii) black (iii) of the same colour.

Solution

i) Let A = {first ball drawn is red}

B = {second ball drawn is red}

Then A B = {both balls are red}

$Now\ P(A) = \frac{3}{7}$ $\quad\quad$ $Now\ P(B/A) = \frac{2}{6}$

$\therefore P\ (A \cap B) = P(A) P\left(\frac{B}{A}\right) = \frac{3}{7} x \frac{2}{6} = \frac{1}{7}$

ii) Let C = {first ball drawn is black}

D = {second ball drawn is black}

Then C B = {both balls are black}

$Now\ P(C) = \frac{4}{7}$ $\quad\quad$ $Now\ P(D/C) = \frac{3}{6}$

P(C B0 = P(C) P(D/C)

$= \frac{4}{7} x \frac{3}{6} = \frac{2}{7}$

iii) When balls are o the same colour either both are red or both are black and the two cases are mutually exclusive.

= P[(A B) or (C D)]

= P(A B) + P(C D)

$= \frac{1}{7} + \frac{2}{7}$

$= \frac{1+2}{7}$

$= \frac{3}{7}$

Illustration 20

A bag contains 5 white and 3 black balls. Two balls are drawn at random one after the other without replacement. Find the probability that both balls drawn are black.

Solution

Probability of drawing a black ball in the first attempt is

$$P(A) = \frac{3}{5+3} = \frac{3}{8}$$

Probability of drawing the second ball given that the first ball drawn is black

$$P\left(\frac{B}{A}\right) = \frac{2}{5+2} = \frac{2}{7}$$

The probability that both balls drawn are black is given by

$$P(A \cap B) = P(A).P(B/A)$$

$$= \frac{2}{8} \times \frac{2}{7} = \frac{3}{28}$$

MATHEMATICAL EXPECTATION

The concept of mathematical expectation is of great importance in statistical work. The mathematical expectation (also called the expected value) of a random variable is the weighted arithmetic mean of the variable, the eights used to find the mathematical expectation are all the respective probabilities of the values that the variable can possibly assume.

Definition

gf x denote a discrete random variable which can assume the values of $x_1, x_2, \ldots x_n$, with respective probabilities $P_1, P_2, \ldots P_n$ where $P_1+P_2+\ldots P_n=1$, the mathematical expectation of x denoted by E(x) is defined as

$$E(X) = P_1X_1 + P_2X_2 + \ldots + P_nX_n$$

Illustration 21

A petrol pump proprietor sells on an average Rs. 70,000 worth of petrol on rainy days and an average of Rs. 90,000 on clear days. Statistics from the meteorological department show that the probability is 0.7 for clear weather and 0.3 for rainy weather on coming Sunday. Find the expected value of petrol sale on coming Sunday.

By data,

$X_1 = 70{,}000,\ P_1 = 0.3$

$X_2 = 90000,\ P_2 = 0.7$

Expected value of petrol sale on coming Sunday is

$$\begin{aligned} E(X) &= P_1X_1 + P_2X_2 \\ &= 0.3 \text{ x } 70{,}000 + 0.7 \text{ x } 90{,}000 \\ &= 21{,}000 + 63{,}000 \\ &= \text{Rs. } 84{,}000 \end{aligned}$$

Illustration 22

A dealer in refrigerator estimates from his past experience the probabilities of his selling refrigerator in a day. These are as follows

No. of refrigerators sold in a day	0	1	2	3	4	5	6
Probability	0.03	0.20	0.23	0.25	0.12	0.10	0.07

Find the mean number of refrigerators sold in a day.

By data, $X_1=0, X_2=1, X_3=2, X_4=3, X_5=4, X_6=5, X_7=6$

$P_1=0.03 \quad P_2=0.20 \quad P_3=0.2 \quad P+=0.25 \quad P_5=0.12 \quad P_6=0.10$

$P_7=0.07$

$= 0 + 0.2 + 0.46 + 0.75 + 0.48 + 0.5 + 0.42$

$= 2.81$

$= 3.$

Illustration 23

The probability that a man fishing at a particular place will catch 1,2,3,4 fish is 0.4, 0.13, 0.2 and 0.1 respectively. What is the expected number of fish caught?

By data, $X_1=1$, $X_2=2$, $X_3=3$, $X_4=4$

$P_1=0.4$ $P_2=0.3$ $P_3=0.2$ $P_4=0.1$

Expected number of fish caught would be given by

$$E(x) = P_1X_1 + P_2X_2 + P_3X_3 + P_4X_4$$
$$= 0.4x1 + 0.3x2 + 0.2x3 + 0.1x4$$
$$= 0.4 + 0.6 + 0.6 + 0.4$$
$$= 2$$

Illustration 24

A salesman wants to know the average number of units he sells per sales call. He check his past sales records and comes up with the following probabilities

Sales in units :	0	1	2	3	4	5
Probability :	0.15	0.20	0.10	0.05	0.30	0.20

What is the average (expected) number of units he sells per sales call.

Solution :

By data, $X_1=0$ $X_2=1$ $X_3=2$ $X_4=3$ $X_5=4$ $X_6=5$

$P_1=0.15$ $P_2=0.20$ $P_3=0.00$ $P_4=0.05$ $P_5=0.30$ $P_6=0.2$

$$E(x) = P_1X_1 + P_2X_2 + P_3X_3 + P_4X_4 + P_5X_5 + P_6X_6$$

$= 0.15x0+0.20 \times 1+0.10 \times 2+0.05x3 + 0.30x4 +$
$0.2x5$

$= 0 + 0.2 + 0.2 + 0.15 + 1.2 + 1.0$

$= 2.75$

Thus he would expect to sell 2.75 or 3 units on each sales Call.

Illustration 25

A consignment is offered to two firms X and Y for Rs. 1,00,000. The following table shows the probability at which the firms will be able to sell it at different prices

Probability	Selling price (Rs.)			
	80,000	90,000	1,05,000	1,10,000
X	0.2	0.3	0.4	0.1
Y	0.25	0.2	0.5	0.05

Which firm X or Y will be more inclined towards the offer

Solution : Firm X

X_1=80,000 X_2=9,000 X_3=1,05,000 X_4=1,10,000

P_1=0.2 P_2=0.3 P_3=0.4 P_4=0.1

$E(X) = P_1X_1 + P_2X_2 + P_3X_3 + P_4X_4$

$= 0.2 \times 80{,}000 + 0.3 \times 90{,}000 + 0.4 \times 1{,}05{,}000 + 0.1 \times 1{,}10{,}000$

$= 16{,}000 + 27{,}000 + 42{,}000 + 11{,}000 = 96{,}000$

Firm Y X_1=80,000 X_2=90,000 X_3=1,05,000 X_4=1,10,000

$P_1=0.25$ $P_2=0.2$ $P_3=0.5$ $P_4=0.05$

$E(Y) = P_1X_1 + P_2X_2 + P_3X_3 + P_4X_4$

$= 0.25 \times 80{,}000 + 0.2 \times 40{,}000 + 0.5 \times 1{,}05{,}000 + 0.05 \times 1{,}10{,}000$

$= 20000 + 18000 + 52500 + 5500 = 96{,}000$

Since the expectation is the same, both the firms would be equally inclined towards the offer.

Illustration 26

There are three alternative proposals before a businessman to start a new project.

Proposal A : Profit of Rs. 5 Lakhs with a probability of 0.6 or a loss of Rs. 80,000 with a probability of 0.4.

Proposal B : Profit of Rs. 10 Lakhs with a probability of 0.4 or a loss of 2 Lakhs with a probability of 0.6

Proposal C : Profit of Rs. 4.5 Lakhs with a probability of 0.8 or a loss of Rs. 50,000 with a probability of 0.2.

If he wants to maximize the profits and minimize the loss, which proposal should be prefer.

Solution

Mathematical expectation of each of the proposal $E(X) = P_1X_1+P_2X_2$

Proposal A : $X_1=5{,}00{,}000$ $X_2 = -80{,}000$

$P_1=0.6$ $P_2 = 0.4$

Expected value $= P_1X_1 + P_2X_2$

$= 0.6 \times 5{,}00{,}000 \times 0.4 \times (-80{,}000)$

$= 3{,}00{,}000 - 32{,}000$

$= 2{,}68{,}000$

Proposal B :

$X_1 = 10,00,000$ $\qquad$ $X_2 = -2,00,000$

$P_1 = 0.4$ $\qquad$ $P_2 = 0.6$

Expected value $E(X) = P_1X_1 + P_2X_2$

$= 0.4 \times 10,00,000 + 0.6 \times (-2,00,000)$

$= 4,00,000 - 1,20,000 = 2,80,000$

Proposal C

$X1 = 4,50,000$ $\qquad$ $X2 = -50,000$ (Loss)

$P1 = 0.8$ $\qquad$ $P2 = 0.2$

Expected value $E(X) = P_1X_1 + P_2X_2$

$= 0.8 \times 4,50,000 + 0.2 \times (-50,000)$

$= 3,60,000 - 10,000$

$= 3,50,000$

Since expected value is heighest in the case of proposal C, hence he should prefer proposal C.

Baye's Theorem

Statement : Let $A_1, A_2, \ldots . A_n$ be n mutually exclusive events and B be any event then

$$P(A_i|B) = \frac{P(A_i)P(B/A_i)}{\sum_{i=1}^{n} P(A_i)P(B/A_i)}$$

Proof

Let $A_1, A_2, \ldots A_n$ be n mutually exclusive events and B be any event.

$A_1 \quad A_2 \quad \ldots . \quad An = S$ where S-sample space and hence

$B = B \quad S$

$= B_n(A_1 \quad A_2 \ldots \quad A_n)$

$= (B \quad A_1) \quad (B \quad A_2) \ldots . (B \quad A_n)$

$= (A_1 \quad B) \quad (A_2 \quad B) \ldots . \quad (A_n \quad B)$

Now since $A_1, A_2, \ldots . A_n$ are mutually exclusive so are the events $(B \quad A_1), B \quad A_2, \ldots (B \quad A_n)$.

$$P(B) = P(A_1 \quad B) + P(A_2 \quad B) + \ldots . + P(A_n \quad B)$$

$$= \sum_{j=1}^{n} P(A_j \cap B)$$

$$= \sum_{j=1}^{n} P(A_j) P(B / A_j)$$

Thus for a fixed j,

$$P\left(\frac{A_j}{B}\right) = \frac{P(A_j \cap B)}{P(B)}$$

$$= \frac{P(A_j) P(B/A_j)}{\sum_j^n P(A_j) P(B/A_j)}$$

Note :

(i) Let A_1, A_2 be 2 mutually exclusive events and B be any event then

(i) $P(A_1 | B) = \frac{P(A_1)P(B/A_1)}{P(A_1)P(B/A_1)+P(A_2)P(B/A_2)}$

(ii) $P(A_2 | B) = \frac{P(A_2)P(B/A_2)}{P(A_1)P(B/A_1)+P(A_2)P(B/A_2)}$

b) Let A_1, A_2, A_3 be 3 mutually exclusive events and B be any event then

(i) $P(A_1 | B) = \frac{P(A_1)P(B/A_1)}{P(A_1)P(B/A_1)+P(A_2)P(B/A_2)+P(A_3)P(B/A_3)}$

(ii) $P(A_2 | B) = \frac{P(A_2)P(B/A_2)}{P(A_1)P(B/A_1)+P(A_2)P(B/A_2)+P(A_3)P(B/A_3)}$

(iii) $P(A_3 | B) = \frac{P(A_3)P(B/A_3)}{P(A_1)P(B/A_1)+P(A_2)P(B/A_2)+P(A_3)P(B/A_3)}$

Illustration

In a bolt factory machines A,B and C manufacture respectively 25%, 35% and 40% of the total of their output 5,4 and 2 percent are defective bolts. A bolt is drawn at random from the product and is found to be defective. What is the probability that it was manufactured by machines A,B and C?

Solution

Let A1 = the event of drawing a bold produced by first machine

A2 = the event of drawing a bolt produced by second machine

A3 = The event of drawing a bold produced by third machine

B = the event of drawing a defective bolts.

Then from the first information

Prior probability

$$P(A_1) = 25\% = \frac{25}{100} = 0.25$$

$$P(A_2) = 35\% = \frac{35}{100} = 0.35$$

$$P(A_3) = 40\% = \frac{40}{100} = 0.40$$

Conditional probability

$$P(B/A_1) = 5\% = \frac{5}{100} = 0.05$$

$$P(B/A_2) = 4\% = \frac{4}{100} = 0.04$$

$$P(B/A_3) = 2\% = \frac{2}{100} = 0.02$$

Joint probability

$$P(A_1)\,P(B/A_1) = 0.25x0.05 = 0.0125$$

$$P(A_2)\,P(B/A_2) = 0.35x0.04 = 0.0140$$

$$P(A_3)\,P(B/A_3) = 0.40x0.02 = 0.0080$$

By Baye's theorem

Posterior Probability

$$P(A_1|B) = \frac{P(A_1)P(B/A_1)}{P(A_1)P(B/A_1) + P(A_2)P(B/A_2) + P(A_3)P(B/A_3)}$$

$$= \frac{0.0125}{0.0125+0.0140+0.0080}$$

$$= \frac{0.0125}{0.0345}$$

$$= 0.362$$

$$P(A_2|B) = \frac{P(A_2)P(B/A_2)}{P(A_1)P(B/A_1) + P(A_2)P(B/A_2) + P(A_3)P(B/A_3)}$$

$$= \frac{0.0140}{0.0125+0.0140+0.0080}$$

$$= \frac{0.0140}{0.0345}$$

$$= 0.406$$

$$P(A_3|B) = \frac{P(A_3)P(B/A_3)}{P(A_1)P(B/A_1) + P(A_2)P(B/A_2) + P(A_3)P(B/A_3)}$$

$$= \frac{0.008}{0.0125+0.0140+0.0080}$$

$$= \frac{0.008}{0.0345}$$

$$= 0.232$$

Event	Prior Probability	Conditional Probability	Joint Probability	Posterior Probability
(1)	(2)	(3)	(4)	(5)
A1	0.25	0.05	0.0125	0.362
A2	0.35	0.04	0.0140	0.406
A3	0.40	0.02	0.008	0.232
	1.0		0.0345 (6)	1.00

Illustration

Assume that a factory has two machines. Past records show that machine 1 produces 30% of the items of output and machine 2 produces 70% of the items. Further 5% of the items produced by machine 1 were defective and only 1% produced by machine 2 were defective. If a defective item is drawn at random, what is the probability that the defective item was produced by machine 1 or machine 2?

Solution

Let A_1 = the event of drawing an item produced by machine 1

A_2 = the event of drawing an item produced by machine 2

B = the event of drawing a defective item produced either by machine 1 or machine 2

Then from the first information,

Prior probability

$P(A_1) = 30\% = 0.30$

$P(A_2) = 70\% = 0.70$

Conditional probability

$P(B/A_1) = 5\% = 0.05$

$P(B/A_2) = 1\% = 0.01$

Joint Probability

$P(A_1)P(B/A_1) = 0.30 \times 0.05 = 0.015$

$P(A_2)P(B/A_2) = 0.70 \times 0.01 = 0.007$

By Baye's theorem,

Posterior probability

$$P(A_1/B) = \frac{P(A_1)P(B/A_1)}{P(A_1)P(B/A_1) + P(A_2)P(B/A_2)}$$

$$= \frac{0.015}{0.015+0.007}$$

$$= \frac{0.015}{0.022}$$

$$= 0.682$$

$$P(A_2/B) = \frac{P(A_2)P(B/A_2)}{P(A_1)P(B/A_1) + P(A_2)P(B/A_2)}$$

$$= \frac{0.007}{0.015+0.007}$$

$$= \frac{0.007}{0.022}$$

$$= 0.318$$

Event	Prior Probability	Conditional Probability	Joint Probability (2) x (3)	Posterior Probability (4) / (6)
(1)	(2)	(3)	(4)	(5)
A1	0.30	0.05	0.015	0.682
A2	0.70	0.01	0.007	0.318
Total	1.0		0.022	1.0

Probability that the defective item was produced by machine 1 or machine 2

$$= P(A_1/(B)+P(A_2/B)$$

$$= 0.682 + 0.318$$

$$= 1$$

THORETICAL DISTRIBUTIONS

The students will be introduced in this section to the techniques of developing discrete and continuous probability distributions and its applications. We may think of a probability distribution. Where just like distributing the total frequency to different class intervals, the total probability (i.e, one) is distributed to different mass points in case of discrete random variable or to different class intervals in case of continuous random variable. Such a probability distribution, since such a distribution exists only in theory. We need study theoretical distribution for the following factors.

(a) An observed frequency distribution in many a case, may be regarded a sample i.e a representative part of a large, unknown, boundealess universe on population and we may be interested to know the form of such distribution. By fitting a theoretical probability distribution so say the lamps produced by a manufacturer, it may be possible for the manufacturer to specify the length of life of the lamps produced by him upto a reasonable degree of accuracy. By studying the effect of a particular type missiles necessary to destroy an army position. By knowing the distribution of smokers, a social activist may warm the people of locality about the nuisance of active and passive smoking and so on.

(b) Theoretical probability distribution may be profitably employed to make short term projections for the future.

(c) Statistical analysis is possible only on the basis of theoretical probability distribution. Setting confidence limits or testing statistical hypothesis about population parameter(s) is based on the probability distribution of the population under consideration.

A probability distribution also possesses all the characteristic of observed distribution. Again a probability distribution may be either a discrete probability distribution or continuous probability distribution depending on the random variable under study. The two important discrete probability distributions are:

a) Binomial distribution
b) Poisson distribution

Some important continuous probability distributions are

a) Normal distribution
b) Chi-square distribution
c) Student's t-distribution
d) F-distribution

Binomial Distribution

Definition

A random variable X is said to follow binomial distribution if it assumes only non-negative values and its probability mass function is given by

P(X=x)=n($_x$P^x q^{n-x}, x=0,1,2,...n and p+q=1

Notation

X~B(n,p), n abd P are parameters of Binomial distribution i.e., X follows Binomial distribution with parameters n and p.

Properties of Binomial distribution

1. Mean of the binomial distribution is no.
2. Variance of the binomial distribution is np
3. Standard deviation of the binomial distribution is $\sqrt{npq}$
4. Moment coefficient of skewness $\gamma_1 = \frac{q-p}{\sqrt{npq}}$
5. Moment coefficient of kurtosis $\gamma_1 = \frac{1-6pq}{npq}$

Note :

(i) $\sum P(X = x) = \sum_{x=0}^{n} nc_x p^x q^{n-x} = (q + p)^n = \mathbf{1}$ since p+q=1

(ii) If N is the total frequency, the expected frequencies of 0,1,2,...n success are the successive terms of N(q+p)n.

(ie) f(x)=N.P(X=x)

(iii) P(X<1) = P(X=0)

(iv) P(X≤1) = P(X=0) + P(X=1)

(v) P(X≤2) = P(X=0) + P(X=1)

(vi) P(X≤2) = P(X=0) + P(X=1) + P(X=2)

(vii) P(X<3) = P(X=0) + P(X=1) + P(X=2)

(viii) P(X≤3) = P(X=0) + P(X=1) + P(X=3)

(ix) P(X≥1) = 1-P(X<1)

(i.e.) P(X≥1) = 1-P(X=0)

(x) $P(X \geq 2) = 1-P(X<2)$

(i.e.) $P(X \geq 2) = 1-[P(X=0) + P(X=1)]$

(xi) $P(X \geq 3) = 1-P(X<3)$

(i.e.) $P(X \geq 3) = 1-[P(X=0) + P(X=1) + P(X=2)]$

Illustration 27

In a large consignment of electric bulbs 10% are defective. A random sample of 20 is taken for inspection. Find the probability that

i) All are good bulbs

ii) At most there are 3 defective bulbs

iii) Exactly there are three defective bulbs

Solution

Given, n=number of samples = 20

$$P = 10\% = \frac{10}{100} = 0.1$$

$q = 1-P$

$= 1-0.1$

$= 0.9$

Then

$P(X=x) = n(_xP^x\ q^{n-x},\ x=0,1,2,\ldots n)$

$P(X=x) = 20C_x.(0.1)^x.\ (0.9)^{20-x}$

(i) P (all are good bulbs)

= P (none are defective)

= P (X=0)

$= 20Co.\ (0.1)^6,\ (o.q)^{20-0}$

$= 20Co \times (0.1)^0 \times (0.9)^{20}$

$= 1 \times 1 \times 0.1216$

$= 0.1216$

(ii) P (at most there are 3 defectives)

$= P(X \leq 3)$

$= P(X=0) + P(X=1) + P(X=2) + P(X=3)$

$= 20C_o \quad (0.1)^0 \quad (0.9)^{20} \quad + \quad 20C^1 \quad (0.1)^1 \quad (0.9)^{19} + 20C_2(0.1)^2(0.9)^{18}$

$+ 20C_3(0.1)^3 (0.9)^{17}$

$= 0.1215 + 0.27 + 0.285 + 0.19$

$= 0.8666$

(iii) P(exactly 3 defective bulbs)

$= P(X=3)$

$= 20C_3 \, (0.1)^3 \, (0.9)^{17}$

$= 0.19$

Illustration 28

A machine manufacturing screws is known to produce 5% defective. In a random sample of 15 screws, what is the probability that there are (i) exactly two defectives (ii) at least two defectives?

Solution

Give n = number of samples = 15

$p = 5\% = \frac{5}{100} = 0.05$

$q = 1-p$

$= 1-0.05$

$= 0.95$

Then binomial distribution is

$P(X=x) = n({}_xP^x \, q^{n-x}, x=0,1,2,\ldots n)$

$P(X=x) = 15C_x(0.05)^x \, (0.95)^{15-x}$

(i) P (exactly two detectives)

$= P(X=2)$

$= 15C_2\ (0.05)2\ (0.95)^{13}$

$= 0.1347$

(ii) P (at least two defectives)

$= P(X \geq 2)$

$= 1 - P(X<2)$

$= 1-[P(X=0) + P(X=1)]$

$= 1-[15C_o(0.05)^0\ (0.95)^{15} + 15C_1(0.05)^1\ (0.95)^{14}]$

$= 1-[0.4632+0.3657]$

$= 1-[\quad]$

Illustration 29

Assuming that half the population are consumers of rice, so that the chance of an individual being a consumer is ½ and assume that 100 investigators each take 10 individuals to see whether these are consumers. How many investigators would you expect to report that 3 people or less were consumers.

Solution :

Given, n = 10,

$P = \frac{1}{2} = 0.5$

$q = 1 - P = 1 - \frac{1}{2} = \frac{1}{2} = 0.5$

N=10-0

Thus probability mass function binomial distribution is

$P(X=x) = nC_x p^x\ q^{n-x}$

$P(X=x) = 10C_x(0.5)^x(0.5)^{10-x}$

P (3 people or less were consumers)

$= P(X \leq 3)$

$= P\ (X=0) + P(X=1) + P(X=2) + P(X=3)$

$= 10C_0(0.05)^0\ (0.5)^{10} + 10C_1\ (0.5)^1\ (0.5)$

$+10C_2(0.5)^2(0.5)^8$

$+10C_3(0.5)^3(0.5)^7$

$= 10C_0(0.5)^{10}+10C_1(0.5)^{10}+10C_2(0.5)^{10}+10C_3(0.5)^{10}$

$= (0.5)^{10}\ [10C_0+10C_1+10C_2+10C_3]$

$= (0.5)^{10}[1+10+45+120]$

$= (0.5)^{10} \times 176 = 0.1718$

Expected number of investigators be reported as three are less were consumers

$= N \times P(X\leq3)$

$= 100 \times 0.1718$

$= 171.8$

$= 172$

Illustration 30

6 dice are thrown 729 times. How many times do you expect atleast three dice to show a five or six?

Solution

P = Probability of getting 5 or 6 with one die

$= P\ (\text{getting } 5) + P\ (\text{getting } 6)$

$= \frac{1}{6}+\frac{1}{6}$

$= \frac{1+1}{6} = \frac{2}{6} = \frac{1}{3}$

$q \quad = 1-P$

$= 1 - \frac{1}{3}$

$= \frac{3-1}{3}$

$= \frac{2}{3}$

Thus probability mass function binomial distribution is given by

$P[X=x] = nC_x.P^xq^{n-x}$

$P[X = x)] = 6C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{n-x}$

P (atleast three dice showing five or six)

$= P(X \geq 3)$

$= P(X=3) + P(X=4) + P(X=5) + P(X=6)$

$$= 6C_3\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^3 + 6C_4\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^2 + 6C_5\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right)^1$$

$$+6C_6\left(\frac{1}{3}\right)^6\left(\frac{2}{3}\right)^0$$

$$= 20x\,\frac{2^3}{3^6} + 15x\,\frac{22}{36} + 6x\,\frac{2}{36} + 1x\,\frac{1}{36}$$

$$= \frac{160+60+12+1}{3^6}$$

$$= \frac{233}{3^6}$$

For 729 times, the expected number of times atleast 3 dice showing five or six.

$= NxP(X \geq 3)$

$= 729 \text{ x } \frac{233}{3^6}$

$= 729 \text{ x } \frac{233}{729}$

$= 233$ times

Illustration 31

If the probability of defective bolts is 0.1, find (a) the mean and standard deviation for the distribution of defective bolts in a total of 500 and (6) the moment coefficient of skewness and kurtosis of the distribution.

Solution

Given n=500

P=0.1

Q=1-p=1-0.1 = 0.9

a) Mean of the binomial distribution

= np

$= 500 \times 0.1$

$= 50$

Standard deviation of the binomial distribution

$= \sqrt{npq}$

$= \sqrt{500 \, x \, 0.1 \, x \, 0.9}$

$= 6.71$

b) Moment coefficient of skewness

$$\gamma_1 = \frac{q-p}{\sqrt{npq}} = \frac{0.9-0.1}{\sqrt{500x0.1x0.9}}$$

$$= \frac{0.8}{6.71}$$

$= 0.119$

Since $_1$ is more than zero the distribution is positively skewed. However, skewness is very moderate

Moment coefficient of Kurtosis

$$\gamma_2 = \frac{1-6pq}{npq}$$

$$= \frac{1-6(0.1)(0.9)}{500 \, x \, 0.1 \, x \, 0.9}$$

$$= \frac{0.46}{44.9}$$

$= 0.01$

Since $_2$ is positive the distribution is platykurtic.

POISSON DISRIBUTION

In the previous section we have discussed a discrete probability distribution Binomial distribution. Now we shall discuss another discrete probability distribution called poisson distribution. The characteristic of the poisson distribution are

(i) The occurrence of the events are independent

(ii) The number of trials is large

(iii) The probability of occurrence is very small

Examples of poisson distribution are

(i) The demand for a product

(ii) The occurrence of accident in a factory

(iii) The arrival pattern in a departmental store

(iv) The arrival of calls of a switch board

Definition

A random variable X is said to follow poisson distribution, if it assumes only non-negative values and its probability mass function is given by

$$P(X = x) = \frac{e^{-m}m^{x}}{x!} \quad x = 0,1,2 \ldots.$$

This distribution has only one parameter m and m=mean = np.

Properties of Poisson distribution

1. Mean of the Poisson distribution = np
2. Variance of the Poisson distribution = np
3. Standard deviation of the Poisson distribution = $\sqrt{np}$
4. Skewness of the Poisson distribution = $\frac{1}{np}$
5. Kurtosis of the Poisson distribution = $3 + \frac{1}{np}$

Illustration 32

If 5% of the items produced by a factory are defective, use Poisson distribution to find the probability that in a sample of 100 items

(i) None is defective

(ii) 5 items will defective (Given e^{-5}=0.007)

Solution

Given, P = Probability of producing a defective item

$= 5\%$

$= \frac{5}{100} = 0.05$

and n = 100

m=np = 100 x 0.05 = 5

Thus, probability mass function Poisson distribution is

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

$$P(X = x) = \frac{e^{-5}(5)^x}{x!}$$

(i) P (none is defective) $= P(X=0)$

$$= \frac{e^{-5}(5)^0}{0!}$$

$$= \frac{0.007x1}{1}$$

$= 0.007$

(ii) P (5 items will be defective)

$= P(X=5)$

$$= \frac{e^{-5}(5)^5}{5!}$$

$$= \frac{0.007 x 3125}{120}$$

= 0.1823

Illustration 33

A car hire firm has two cars which is hires out day by day. The number of demands for a car on each day is disbuted as a Poisson distribution with mean 1.5. Calculate (i) the proportion of days on which neither car is used and (ii) the proportion of days on which some demand is refused ($e^{-1.5}$=0.2231)

Solution

Given that M=1.5

Thus the probability mass function of Poisson distribution is

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

(i.e) $P(X = x) = \frac{e^{-1.5}(1.5)^x}{x!}$

(i) P(neither car is used)

$= P(X=0)$

$= \frac{e^{-1.5}(1.5)^0}{0!}$

$= \frac{0.2231 \times 1}{1} = 0.2231$

Hence the proportion of days that neither car is used

$= N.P(X=0)$

$= 365 \times 0.2231$

$= 81.4$ days

(ii) P (some demand is refused)

$= P(X>2)$

$= 1-[P(X\leq 2]$

$= 1-[P(X=0)+P(X=1) + P(X=2)]$

$= 1 - \left[\frac{e^{-1.5}(1.5)^6}{0!} + \frac{e^{-1.5}(1.5)^1}{1!} + \frac{e^{-1.5}(1.5)^2}{2!}\right]$

$= 1 - e^{-1.5}\left(\frac{1}{1} + \frac{1.5}{1} + \frac{2.25}{2}\right)$

$= 1-0.2231\ [3.625]$

$= 1-0.8087$

$= 0.1913$

Hence the proportion of days that some demand is refused

$= NxP\ (X>2)$

$= 365 \times 0.1913$

$= 69.83$ days

Illustration 34

BHEL, Trichy, employing a large number of workers finds that over a period of time, the average absentee rate is three workers per shift. Calculate the probability that on a given shift (i) exactly 3 workers will be absent (ii) more than 2 workers will be absent [$e^{-3}=0.0497$]

Solution

Given that m = average rate = 3

Thus probability mass function of Poisson distribution is

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

$$P(X = x) = \frac{x^{-3}(3)^x}{x!}$$

(i) P (exactly 3 workers) $= P(X=3)$

$$= \frac{e^{-3}x(3)^3}{3!}$$

$$= \frac{0.0497 \, x \, 27}{6} = 0.224$$

(ii) p (more than 2 workers are absent)

$= P(X>2)$

$= 1-P(X\leq 2)$

$= 1-[P(X=0) + P(X=1) + P(X=2)]$

$$= 1 - \left[\frac{e^{-3}(3)^0}{0!} + \frac{e^{-3}(3)^1}{1!} + \frac{e^{-3}(3)^2}{2!}\right]$$

$= 1-[0.045 + 0.149 + 0.223]$

$= 1-0.417$

$= 0.583$

Illustration 36

A book contains 100 misprints randomly throughout its 100 pages. What is the probability that a page observed at random contains at least 2 misprints? $[e^{-1}=0.368]$

Solution

Given that

$$m = \frac{100}{100} = 1 = \textit{The average number of misprints per page}$$

Thus, the probability mass function of Poisson distribution is

$$P(X = x) = \frac{e^{-m}(1)^x}{x!}$$

$$P(X = x) = \frac{e^{-1}(1)^x}{x!}$$

P(at least 2 misprints)

= P(X=2)

= 1-P(X<2)

= 1-[P(x=0) + P(x=1)]

$$= 1 - \left[\frac{e^{-1}(1)^0}{0!} + \frac{e^{-1}(1)^1}{1!}\right]$$

$$= 1 - \left[\frac{e^{-1}x1}{1} + \frac{e^{-1}x1}{1}\right]$$

= 1-[2e^{-1}]

= 1-0.736

= 0.264

Illustration 37

It is given that 3% of electric bulbs manufactured by a company are defective. Using the Poisson approximation, find the probability that a sample of 100 bulbs will contains (i) No defective (ii) Exactly one defective [e^{-3}=0.05]

Solutions

Give that P=3% = 3/100 = 0.03

n=100

m = np = 100 x 0.03 = 3

Thus, probability mass function of Poisson distribution is

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

$$P(X = x) = \frac{e^{-3}(3)^x}{3!}$$

(i) P(no defective) = P(X=0)

$$= \frac{e^{-3}(3)^0}{0!}$$

$$= \frac{e^{-3}x1}{1} = 0.05$$

(ii) P(one defective) = P(X=1)

$$= \frac{e^{-3}(3)^1}{31!}$$

$$= \frac{0.05 \times 3}{1} = 0.15$$

Illustration 38

The mean of the Poisson distribution is 2.25. Find mean, variance, standard deviation, skewness and kurtosis of the Poissoin distribution.

Given that mean m=2.25 = np

(i) Mean of the Poisson distribution m=np=2.25

(ii) Variance of the Poisson distribution = np = 2.25

(iii) Standard deviation of the Poisson distribution

$$= \sqrt{np}$$

$$= \sqrt{2.25}$$

$$= 1.5$$

(iv) Skewness of the Poisson distribution

$$= \frac{1}{nP} = \frac{1}{m} = \frac{1}{2.25} = 0.444$$

(v) Kurtosis of the Poisson distribution

$$= 3 + \frac{1}{m}$$

$$= 3 + \frac{1}{2.25} = 3+0.444 = 3.444$$

NORMAL DISTRIBUTION

Definition: A random variable X is said to follow normal distribution with mean μ and variance ^{2}, if its probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < \infty, \sigma > 0, -\infty < \mu < \infty$$

Where μ=mean of the random variable x

=standard deviation of the normal distribution

=the constant 3.14.

Standard Normal Distribution

Let X is a normal variable with mean μ and S.D. . Define $z = \frac{x-\mu}{\sigma}$ then Z is called standard normal variable then

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Properties of normal distribution :

1. The normal curve is bell shaped
2. In normal distribution, mean = median = mode
3. It is symmetrical around the distribution mean
4. If ranges between -∞ to ∞
5. The total area under normal curve is 1
6. The area under the curve between two points is the probability that a variable which is normally distributes will assure that value between those two points.

Illustration 39

The weekly wages of 1000 workmen are normally distributed around a mean of Rs. 70 with a S.D of Rs. 5. Find the number of workers whose weekly wages will be

(i) Between Rs. 69 and Rs. 72

(ii) Less than Rs. 69

(iii) More than Rs. 72

Solution

Given that,

μ = mean of the normal distribution = 70

= standard deviation of the normal distribution = 5

When X=69, then $z = \frac{x-\mu}{\sigma} = \frac{69-70}{5} = \frac{1}{5} = -0.2$

When X=72 then $z = \frac{x-\mu}{\sigma} = \frac{72-70}{5} = \frac{2}{5} = -0.4$

(i) P (weekly wages between Rs. 69 and Rs. 72)

$= P(69<X<72)$

$= P(-0.2 <Z<0.4)$

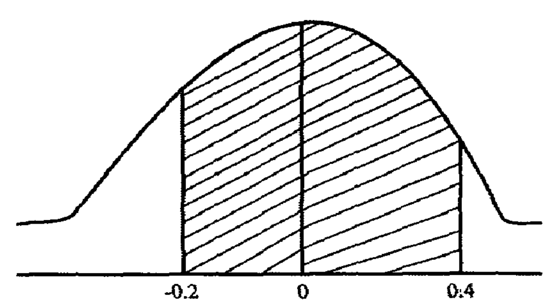
$= P(-0.2<Z<0) + P(0<Z<0.4)$

$= P(0<Z<0.02) + P(0<Z<0.4)$

$= 0.0793 + 0.1554$

$= 0.2347$

Out of 1000 workmen, the number of workers whose wages lies between Rs. 69 and Rs. 72



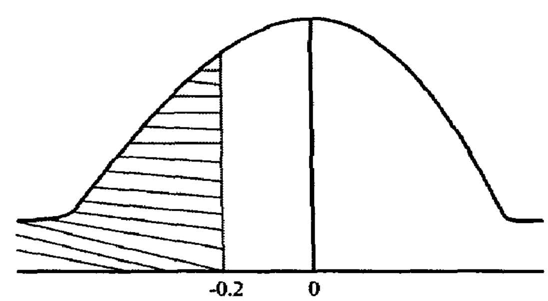
$= N \times P\ (69 < X < 72)$

$= 1000 \times 0.2347$

$= 234.7$

$= 235$

(ii) P (weekly wages less than Rs. 69)



$= P(X<69)$

$= P(Z<-0.2)$

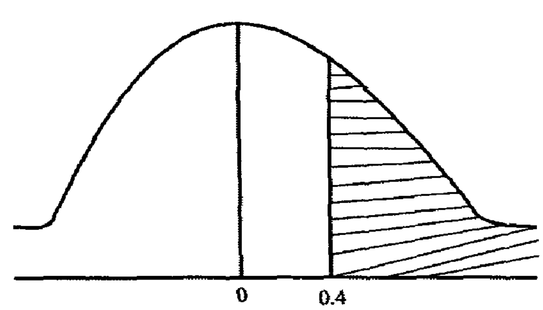
$= 0.5-P(-0.2<Z<0)$

$= 0.5-P(0<Z<0.2)$

$= 0.5-0.0793$

$= 0.4207$

Out of 1000 workmen, the number of workers whose wages less than Rs.69



$= NXP(X<69)$

$= 1000 \times 0.4207$

$= 420.7$

$\simeq 421$

(iii) P(weekly wages more than Rs. 72)

$= P(X>2)$

$= P(Z>0.4)$

$= 0.5 - P(<Z<0.4)$

= 0.5 – 0.1554 = 0.3346

Out of 1000 workmen, the number of workers whose weekly wages are greater than 72

= N.P(Z>4)

= 1000 x 0.3446

= 344.6

= 345

Illustration 40

The customer accounts at a certain departmental store have an average balance of Rs. 480 and a standard deviation of Rs. 160. Assuming that the account balances are normally distributed.

(i) What proportion of the accounts is over Rs. 600?

(ii) What proportion of the accounts is between Rs. 400 and Rs. 600

(iii) What proportion of the accounts is between Rs. 240 and Rs. 360

Solution

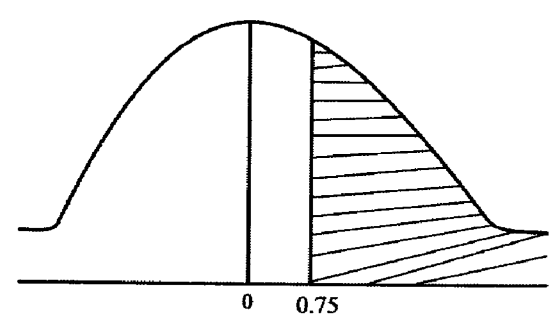
Given that mean μ=480, standard deviation =160

(i) When X=600, the standard normal variate

$$Z = \frac{x-\mu}{\sigma}$$

$$= \frac{600 - 480}{160}$$

= 0.75



P(X>600) = P(Z>0.75)

= 0.5-P(0<Z<0.75)

= 0.5 – 0.2734

= 0.2266

Hence 22.66% of the accounts have a balance in onless of Rs. 600

(ii) Where X=400

Then standard normal variable $Z = \frac{X-\mu}{\sigma}$

$= \frac{400-480}{160} = 0.5$

When X = 600,

Then standard normal variable $Z = \frac{X-\mu}{\sigma}$

$= \frac{600-480}{160} = 0.75$

P(400≤X≤600) = P(-0.5≤Z≤0.75)

= P(-0.5≤Z≤0) + P(0≤Z≤0.75)

= P(0≤Z≤0.5) + P(0≤Z≤0.75)

= 0.1915 + 0.2734 = 0.4649

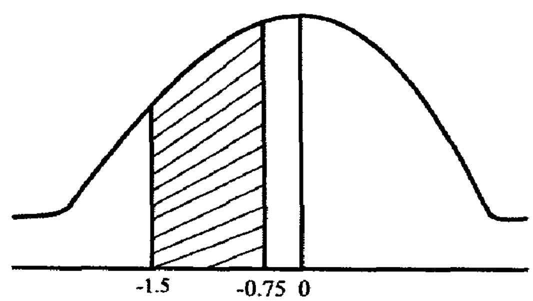
Hence 46.49% of the accounts have an average balance between Rs. 400 and Rs. 600.

(iii) When X=240, then

Standard normal variable $Z = \frac{X-\mu}{\sigma}$

$= \frac{240-480}{160}$

$= \frac{-240}{160} = -0.75$



When X = 260 then

Standard normal variate

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

$= \frac{360-480}{160}$

= -1.5

P(240≤X≤360) = P(-0.75≤Z≤-1.5)

= P(1.5≤Z≤0) – P(-0.75≤Z≤0)

= P(0≤Z<1.5) – P(0<Z<0.75)

$= 0.4332 - 0.2734$

$= 0.1598$

Hence 15.08% of the accounts have an average balance between Rs. 240 and Rs. 360

Illustration 41

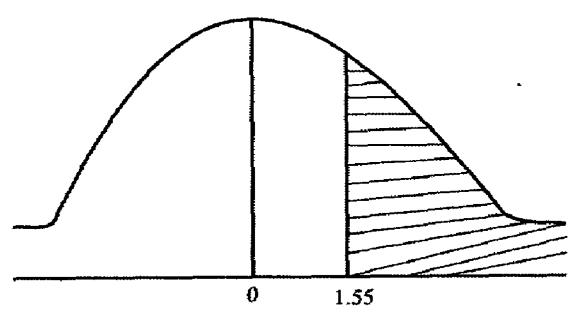
It is known that from the past experience that the number of telephone call made daily in a certain city between 10 am and 11 am has a mean of 352 and standard deviation of 31. What percentage of times will there be more than 400 telephone calls made in this locality between 10 am and 11 am?

Given that mean $\mu = 352$

and standard deviation $=31$

When X = 400 then

Standard normal variate



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

$$= \frac{400-352}{35}$$

$= 1.55$

The probability that there will be more than 400 calls

$= P(X>400)$

$= P(Z>1.55)$

$= 0.5 - P(0<Z<1.55)$

$= 0.5 - 0.4394$ (using standard normal table)

$= 0.606$

Hence the percentage of days on which the number of calls will exceed 400 is 60.6%

Illustration 43

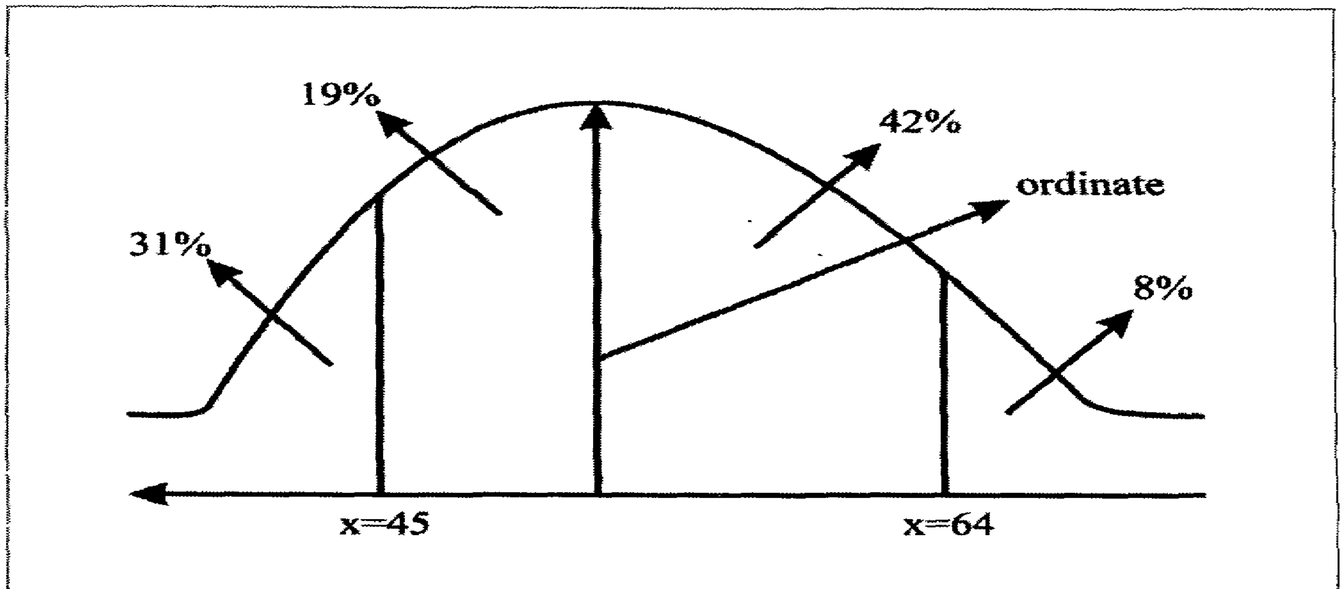
In a normal distribution 31% of the items are under 45 and 8% are over 64. Find the mean and standard deviation of the distribution.

Solution

Let the mean of the normal distribution be μ and standard deviation of the normal distribution be .

The are lying to the left of the ordinate at X=45 is 31%=0.31.

The are lying to the right of the ordinates up to the mean is 0.5-0.31=0.19



The value of Z corresponding to the area 0.19 is 0.5 nearly.

$$\therefore \frac{x-\mu}{\sigma} = -0.5$$

$$\frac{45+\mu}{\sigma} = -0.5$$

$45-\mu = -0.5$

$45=-0.5 \ +\mu$

$-0.5 \ +\mu=45$ (1)

Area to the left of the ordinate at X=64 is 0.5-0.08 = 0.42 and hence value of Z corresponding to this area is 1.4 nearly.

$$\frac{X-\mu}{\sigma} = 1.4$$

$$\frac{64-\mu}{\sigma} = 1.4$$

64-μ = 1.4

64=1.4 +μ

1.4 +μ=64 (2)

Solving (1) and (2)

1.4 +μ=64

-0.5 +μ=4.5

1.9 =19

$$\sigma = \frac{19}{1.9} = 10$$

=10

Substituting =10 in equation (2)

1.4 +μ = 64

1.4 x 10 + μ = 64

μ = 64-14

μ = 50

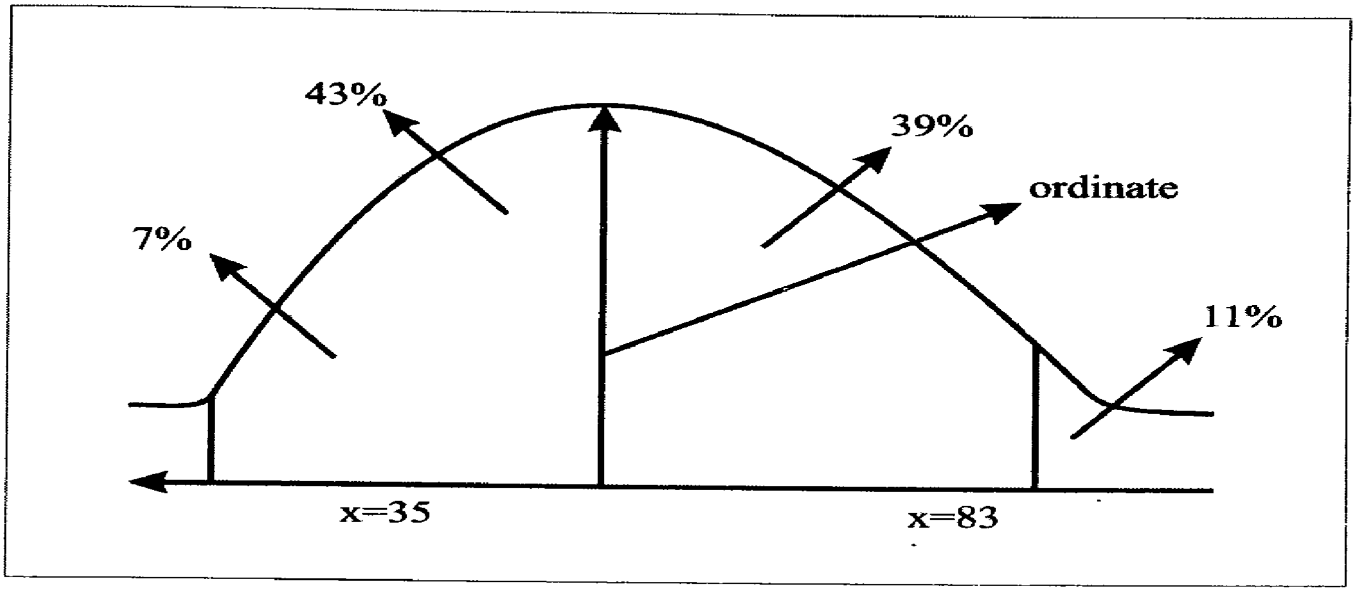
Mean = μ=50

Standard deviation =10

Illustration 44

In a normal distribution, 7% of the items are under 35 and 87% are under 63. What are the mean and standard deviation of the distribution.

Let mean of normal distribution be μ standard deviation of normal distribution be



The are lying to the left of the ordinate at x=35 is 0.07.

The corresponding value of Z is negative.

The area lying to the right of the orinates at x=63 upto mean is

$$0.5\text{-}0.07 = 0.43$$

The value of Z corresponding to the area 0.43 is 1.4757

(ie) $\frac{X-\mu}{\sigma} = -1.4757$

$$\frac{35-\mu}{\sigma} = -1.4757$$

$$35\text{-}\mu = \text{-}1.4757$$

$$35 = \text{-}1.4757 \ \ +\mu$$

The area lying to the left of the ordinate at X=63 upto the mean is 0.39.

The value of Z corresponding to the area 0.39 is 1.2263.

$$\frac{X-\mu}{\sigma} = -1.2263$$

$$\frac{63-\mu}{\sigma} = -1.2263$$

$$63\text{-}\mu = \text{-}1.2263$$

$$63 = \text{-}1.2263 \ \ +\mu \qquad (2)$$

Solving (1) and (2) we get

Mean $\mu = 50.288$

S.D $= 10.36$

Uniform Distribution

A random variable 'x' is said to have a uniform distribution, if its probability density function is given by

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0 & other\ wise \end{cases}$$

Properties of uniform distribution

1. Mean of the uniform distribution $= \frac{a+b}{2}$
2. Variance of the uniform distribution $= \frac{(b-a)^2}{12}$
3. Moment generating function of uniform distribution $= \frac{e^{bt}-e^{at}}{(b-a)t}$

Illustration

If X is uniformly distributed random variables with mean 1 and variance $\frac{4}{3}$, find P(X<0)

Solution

For uniform distribution, we know that

$$\text{Mean} = \frac{a+b}{2} = 1$$

$$a+b=2 \qquad \text{-----} \qquad (1)$$

$$\text{Variance} = \frac{(a-b)^2}{12} = \frac{4}{3}$$

$$\Rightarrow (b-a)^2 = \frac{12x4}{3}$$

$$(b-a)^2 = 16$$

$$b-a = \pm 4 \qquad \text{-----} \qquad (2)$$

Solving (1) and (2) we get

a=-1, b=3

$$\therefore f(x) = \frac{1}{b-a}$$

$$= \frac{1}{3-(-1)}$$

$$= \frac{1}{4} \text{ in } -1<x<3$$

$$\therefore P(X < 0) = \int_{-1}^{0} f(x)dx$$
$$= \int_{-1}^{0} \frac{1}{4} dx$$
$$= \frac{1}{4} [x]_{-1}^{0}$$
$$= \frac{1}{4} [0 - (-1)] = \frac{1}{4}$$

Illustration

Subway trains on a certain line run every half hour between midnight and six in the morning. What is the probability that a man entering the station at a random time during this period will have to wait atleast twenty minutes.

Solution

Let x be the random variable which denotes the waiting time for the next time.

Assume that a man arrives at the station at random, the random variable X is distributed uniformly in (0,30) with p.d.f

$$f(x) = \begin{cases} \frac{1}{b-a}, Q < X < b \\ 0 \quad otherwise \end{cases}$$

$$f(x) = \begin{cases} \frac{1}{30-0}, 0 \le X \le 30 \\ 0 \quad otherwise \end{cases}$$

$$= \begin{cases} \frac{1}{30-0}, 0 \le X < 30 \\ 0 \quad otherwise \end{cases}$$

P(atleast 20 minutes) $= P(X \geq 20)$

$$= \int_{20}^{30} f(x)dx$$
$$= \int_{20}^{30} \frac{1}{30} dx$$
$$= \frac{1}{30} (x)_{20}^{30}$$
$$= \frac{1}{30} (30 - 20)$$
$$= \frac{1}{30} (1)$$
$$= \frac{1}{3}$$

SUMMARY

In this unit we come to know that the calculations of probability of an event, addition theorem of probability, multiplication theorem of probability, and Baye's theorem, binomial distribution, Poisson distribution and normal distribution.

KEYWORDS

1. Event
2. Independent events
3. Probability of an event
4. Addition theorem of probability
5. Multiplication theorem of probability
6. Baye's theorem
7. Probability mas function
8. Probability density function
9. Binomial distribution
10. Poisson distribution
11. Normal distribution

KEY TO CHECK YOUR PROGRESS:

1. From a bag containing three red and two black balls, two balls are drawn at random. Find the probability they are the same colour. $\left[And\ :\ \frac{2}{5}\right]$.
2. If from a pack of cards a single card is drawn. What is the probability that it is either a spade or a king. $\left[And\ :\ \frac{16}{52}\right]$.

3. Find the probability drawing a queen, a king and a Jockey in that order from a pack of cards in three consecutive draws, the cards drawn not been replaced. Ans = 0.00048
4. A class consists of 80 student, 20 of them are girls and 55 boys, 10 of them are rich and remaining poor 20 of them are fair complexioned. What is the probability of a fair complexioned rich girl? [Ans : 0.0098]
5. The probability that a boy will get a scholarship is 0.9 and that a girl will get is 0.8 what is the probability that atleast one of them will get the scholarship? [Ans : 0.98]
6. The incidence of a certain disease is such that on the average 20% of workers suffer from it. If 10 works are selected at random, find the probability that (i) exactly 2 workers suffer from the disease (ii) not more than 2 workers suffer from the disease AM (i) 0.302 (ii) 0.678
7. In a town 10 accidents took place in a span of 50 days. Assuming that the number of accidents per day follows the Poisson distribution. Find the probability that there will be three or more accidents in a day [Ans : 0.001]
8. An aptitude test for selecting officers in a bank was conducted on 1,000 candidates, the average score is 42 and the standard deviation of scores in 24. Assuming normal distribution for the scores find
   a) The number of candidates whose score exceeds 58.
   b) The number of candidates whose score lies between 30 and 66.

   Ans a) 252 b) 533
9. A sample of 100 day dry battery cells tested to find length of the life produced the following results :

Mean = 12 hours, =3 hours

Assuming the data to be normally distributed, what percentage of battery cells are expected to have life

i) More than 15 hours

ii) Less than 6 hours and

iii) Between 10 and 14 hours

Ans : a) 15.87% b) 2.28% c) 49.74%

10. At a certain examination 10% of the students who appeared for the paper in statistics got less than 30 marks and 97% of the student got less than 62 marks. Assuming the distribution is normal. Find the mean and the standard deviation of the distribution.

UNIT IV : INDEX NUMBERS AND TIME SERIES

Introduction

Historically, the first index was constructed in 1764 to compare the Italian price index in 1750 with the level in 1500. Though originally developed for measuring the effect of change in prices, index numbers have today become one of the most widely used statistical devices and there is hardly and field where they are not used. Newspapers headline the fact that prices are going up or down, that industrial production is rising or falling, that imports are increasing or decreasing, that crimes are rising in a particular period as disclosed by index numbers. They are used to feel the pulse of the economy and they have come to be used as indicators of inflationary or deflationary tendencies. In fact, they are described barometers of economic activity. (i.e) if one wants to get an idea as to what is happening to an economy, he should look to important indices like the index number of industrial production, agricultural productions, business activity etc.,

Some prominent definitions of index numbers are given below.

1. Index numbers are devices for measuring differences in the magnitude of a group of related variables". - Croxton & Cowden
2. An index number is a statistical measure designed to show changes in a variable or group of related variables with respect to time, geographic location or other characteristic such as income, profession, etc."

   - Spiegel

3. Generally speaking, index numbers measure the size or magnitude of some object at a particular point in time as percentages of some base or reference object in the past.

- Berensen & Levine

Unit objectives

The objectives of the unit are as follows

To understand the various types of index numbers

To construct consumer price index

To know about the various tests of perfection of index numbers.

Unit structure

Introduction

Definition

Unit objectives

Unit structure

Characteristic of index numbers

Uses of index numbers

Types of index numbers

Business index number

Sensex, Nifty

Problem related to construction of index numbers

Weighted price index(WPI)

Tests of consistency of index numbers

Chain base Index and Fixed base index

Construction of consumer price index(CPI)

Aggregate Expenditure method

Family Budget method

Production index.

Limitations of index numbers

Exercise

Characteristic of Index Numbers

1. Index numbers are specialized average helps comparing the changes in variables which are is different unit.
2. Index numbers are expressed in percentages to show the extent of change without using percentage sign (%)
3. Relative variations are measured with the help of index numbers. Index numbers are for comparison they compare changes taking place over time or between places and like categories.

Uses of Index Numbers

1. The index numbers are used as end results in many situations. For example index numbers are often cited in new reports and acting as general indicators of economic condition.
2. Index numbers are used as part of an intermediate computation to understand other information latter.
3. Sales indices were used to modify and improve the estimates of future.
4. Consumer price index is used to determine the "real" buying power of money.

Types of Index Numbers

There are various types of index numbers.

The most important types are

1. Price index numbers
2. Quantity index numbers
3. Value index numbers.

Price Index Numbers

Price Index is a normalized average of prices for a given class of goods or services in a given region, during a given interval of time. It is a statistic designed to help in compare how these prices, taken as a whole, differ between time periods or geographical locations price index have several potential uses. For particularly broad indices the index can said to measure the economy's price level or a cost of living. More narrow price indices can help producers with business plans and pricing. Sometimes, they can be useful in helping to guide investment".

Quantity index Numbers

Quantity Index numbers study the changes in the volume of goods produced or consumed. For instance industrial production, agricultural production, import, export etc., they are useful and helpful to study the output in an economy.

Value Index Numbers

Value Index numbers compare the total value of a certain period with the total value of the base period. Here the total value is equal to the price of each, multiplied by the quantity. For instance, indices of profits, sales, inventories etc.,

Business Index Number

Today index numbers are one must widely used statistical tools in Business. They are used to the pulse of economy and come to used as indicate as of inflation and deflationary Tendency Application of index numbers in business is called Business Index Number. It is integration of business practices and index numbers for the purpose facilitating forward planning and decision making by the management.

Application index number in business

i) Index numbers act as economic parameters.
ii) Index numbers help in policy formulation.
iii) It reveals trends and Tendencies.
iv) It helps to measure purchasing power.

INDICES USED IN INDIAN CAPITAL MARKET:

Indices are used to give information about the price movements of securities in the stock markets. Stock Market Indices are meant to capture the overall behavior of equity market. A share index is a compilation of several shares prices into a single number. A stock market index is created by selecting a group of stocks that are representatives of the whole market or a specified sector or segment of the market. An Index is calculated with reference to a base period and a base index value. The movement in an index represents the overall increase or decrease in prices of shares included in the index. However, it does not mean that all scrips have moved in a particular direction. The basic idea of an index is that one is able to get an idea or a broad picture about how the market or a particular segment has performed by looking at the index rather entering the complicated process of looking at each scrip. Stock market indices are useful for a variety of reasons. Some of reasons of maintaining share indices are :

They provide a historical comparison of returns on money invested in the stock market against other forms of investments.

They can be used as a standard against which to compare the performance of an equity fund.

It is an indicator of the performance of the overall economy.

Stock indices reflect highly upto date information.

Indices based Index Funds, Index Futures, Index Options play an important role in financial investments and risk management.

Share indices reflect the changing expectations of the stock market about India's corporate sector. When the index goes up, it is because the stock market thinks that the prospective growth in the future will be better than previously thought. When prospects of dividends in the future become pessimistic, the index drops. The index gives, an instant reading about how the stock market perceives the future of India's corporate sector.

Indices are calculated as weighted average and each share (security) is given weight proportional to its market capitalization. Suppose, an index consists of three shares A, B and C and total capitalization of these 3 securities is Rs. 350 crores (A – Rs. 200 crores, B – Rs. 100 crores and C- Rs. 50 crores), then weights of these securities in the index are 4 : 2 : 1.

Types of Indices:

Different types of indices are available in practice. These can be broadly classified into two groups:

i) **Broad Market Index :**

These indices consist of all the large stocks of the capital market. A broad market index works as a benchmark of the entire capital market and also of entire economy. S & P CNX NIFTY, S & P CNX 500 and SENSEX are example of broad market indices.

ii) **Specialised Sector Indices :**

These are also known as Sectoral Indices. These are the indices calculated in respect of shares of companies operating is a particular sector of the economy. BSE – PSU is a specialized index dealing with the price movement in the shares of PSUs. Recently, in July 2007, BSE has launched 'BSE Reality Index', comprising of 11 companies engaged in real estate development business.

In India, there are number of indices being calculated by BSE, NSE, Research Houses, Business Newspapers. NSE is calculating 9 major and 16 other indices on the basis of share prices movement on NSE. Similarly, BSE is also calculation 15 different types of indices. However, among all, the most important has been the SENSEX, the price sensitivity index of BSE Ltd. and the NIFTY.

SENSEX: The Barometer of Indian Capital Market.

SENSEX is based on the closing prices of 30 constituent shares which are periodically selected on the basis of large, liquid and representative companies.

BSE was established as the first stock exchange in India in 1875. Since then, it has occupied the place of prime relevance in the Indian Capital Market. In 1986, BSE came out with a share index SENSEX, which has since become the barometer of Capital Market in India. SENSEX has the base year 1978 – 79 for which the index number is taken as 100. SENSEX is based on the closing prices of 30 constituent shares which are periodically selected on the basis of large, liquid and representative companies. There has been a regular upward and downward movement in the SENSEX to reflect the price behavior of the 30 shares. Initially, SENSEX was calculated on the basis of total market capitalization of a particular company. However, w.e.f. Sept. 1, 2003, SENSEX has been converted into a **free float index**. Now, the weights of different shares are assigned on the basis of market capitalization of freely trading shares. The shares held by promoters are kept outside the index calculation. Now, SENSEX is calculated using the "Free Float Market Capitalization" method in which the level of index at any particular point of time reflects the market value of free floats of 30 shares. The market value of total shares is multiplied by the free float factor to find out the free float market capitalization. During trading hours, the current prices of index scrips are used by the SENSEX mechanism to calculate the value of SENSEX every 15 seconds. The closing SENSEX value for the day is computed on the basis of weighted average of all trades in all the 30 shares during last 30 minutes of the trading session. Weight assigned to a particular share is adjusted for Right, Bonus or Public issue made by the company. As on May 31, 2011, 30 shares included in SENSEX, their weights, free- float factors and the $\boldsymbol{\beta}$ factor are given in Table.

Table: SENSEX composition, Weights, Free – float factor and β as on 31-5-13

	Company	Weightage (%) in SENSEX as on	Beta Values	Co-efficient of Detetmination (R^2)	Free Float Factor
1.	Reliance	11.57	0.93	0.49	0.55
2.	Infosys Tech	9.20	0.82	0.41	0.85
3.	ICICI Bank	8.44	1.44	0.62	1.00
4.	ITC Ltd.	7.07	0.70	0.31	0.70
5.	Larsen & Toubro	6.09	1.13	0.49	0.90
6.	Housing Devpt.	6.07	1.14	0.52	0.90
7.	HDFC Bank Ltd	6.00	1.03	0.52	0.80
8.	State Bank of India	4.43	1.20	0.46	0.45
9.	TCS Ltd.	4.59	0.89	0.37	0.30
10.	ONG Corp. Ltd.	3.25	0.71	0.21	0.20
11.	Bharti Airtel	3.36	0.79	0.18	0.35
12.	Tata Steel	2.67	1.20	0.52	0.70
13.	Tata Motors	2.78	1.32	0.47	0.70
14.	BHEL	2.25	0.82	0.37	0.35
15.	NTPC Ltd.	1.88	0.68	0.34	0.70
16.	Mahindra & Mahindra	2.23	1.18	0.40	0.80

17.	Hindalco Industries	1.79	1.50	0.53	0.70
18.	Hindustan Unilever Ltd.	2.22	0.59	0.20	0.50
19.	Sterilite Industries	1.75	1.31	0.45	0.45
20.	Jindal Steel	1.84	0.97	0.47	0.45
21.	Wipro Ltd.	1.85	0.85	0.34	0.25
22.	Tata Power	1.38	0.55	0.27	0.70
23.	Bajaj Auto	1.31	0.67	0.02	0.70
24.	Maruti Suzuki	1.20	0.85	0.27	0.50
25.	Cipta Ltd.	1.15	0.60	0.13	0.65
26.	Hero Honda	1.25	0.53	0.06	0.50
27.	REL Infrastructure	0.56	1.16	0.24	0.55
28.	Jaiprakash Associates	0.68	1.77	0.49	0.55
29.	DLF Limited	0.68	1.49	0.52	0.25
30.	Reliance Com. Ltd.	0.44	1.16	0.21	0.35
SENSEX			1.00		

Beta = Co-variance (SENSEX, Stock)/Variance (SENSEX)

R^2 = (Correlation)^2

Average Daily Volatility = One standard deviation of daily returns of individual stock price for last one year

Returns = Variation in the stock price over last one year

S & P CNX NIFTY:

It is also called NIFTY. NSE introduced NIFTY in association with CRISIL on November 3, 1995 when it started with base 1,000. It is built out of 50 largest and highly liquid stocks. Over the years, NIFTY had emerged as an important indicator of the movement of prices of important shares on the National Stock Exchange. With effect from July 2009, NIFTY is calculated based on Free – Float Capitalization methodology.

S & P CNX Defty:

S & P CNX Defty is S & P CNX Nifty, measured in US $ terms. If the S & P CNX Nifty rises by 2%, it means that the Indian stock market rose by 2% measured in rupees. If the S & P CNX Defty rises by 2%, it means that the Indian stock market rose by 2%, measured in US dollars. The S & P CNX Defty is calculated on real time basis. Data for the S & P CNX Nifty and the dollar – rupee rate is absorbed in real time, and used to calculate the S & P CNX Defty in real time. When there is currency volatility, the S & P CNX Defty is an ideal device for a foreign investor to know where he stands, even intraday.

S & P CNX 500:

The S & P CNX 500 is India's first broad based benchmark of the Indian capital market based on the prices of 500 shares. The S & P CNX 500 represents about 90% of total market capitalization. 500 companies are disaggregated into 72 industries. Industry weightages in the index dynamically reflect the industry weightages in the market. For example, if the banking sector has a 5 % weightage among the universe of stocks on the NSE, banking stocks in the index would have an approximate representation of 5% in the index. The S & P CNX 500 is a market capitalization weighted index. The base date for the index is the calendar year 1994 with the base index value being 1000. Companies in the index are selected based on their market capitalization, industry representation, trading interest and financial performance. The index is calculated and disseminated on real – time basis,

S & P CNX NIFTY Junior:

S & P CNX Nifty Junior is an index built out of the next 50 (after those included in NIFTY) large, liquid stocks in India. It may be useful to think of the S & P CNX NIFTY and the CNX NIFTY Junior as making up the 100 most liquid stocks in India. CNX NIFTY Junior can be viewed as an incubator where young growth stocks are found. As with the S & P CNX NIFTY, stocks in the CNX NIFTY Junior are filtered for liquidity, so they are the most liquid of the stocks excluded from the S & P CNX NIFTY. The maintenance of the S & P CNX NIFTY and the CNX NIFTY Junior are synchronized so that the two indices will always be disjoint sets, i.e., a stock will never appear in both indices at the same time.

CNX Midcap 200:

CNX Midcap 200 is computed using market capitalization weighted method, wherein the level of the index reflects the total market value of all the stocks in the index relative to a particular base period. The method also takes into account continuous changes in the index and importantly corporate actions such as stock splits, rights, etc. without affecting the index value. The base period is calendar year 1994, indexed to a value of 1000. Therefore, the value of CNX Midcap 200 at any point of time reflects the aggregate of the market capitalization of the index constituents relative to the base value of 1000. There are 200 companies in this index.

S & P CNX NIFTY Shariah.

There are some investors who want to invest as per Shariah Guidelines (their religious beliefs), under which investment in companies which are engaged in following activities is not allowed : Alcohol, Gambling, Financials, Advertising and Media, Tobacco, Forward trading in Gold and Silver, Pork. So, there was felt a need for Shariah comply index.

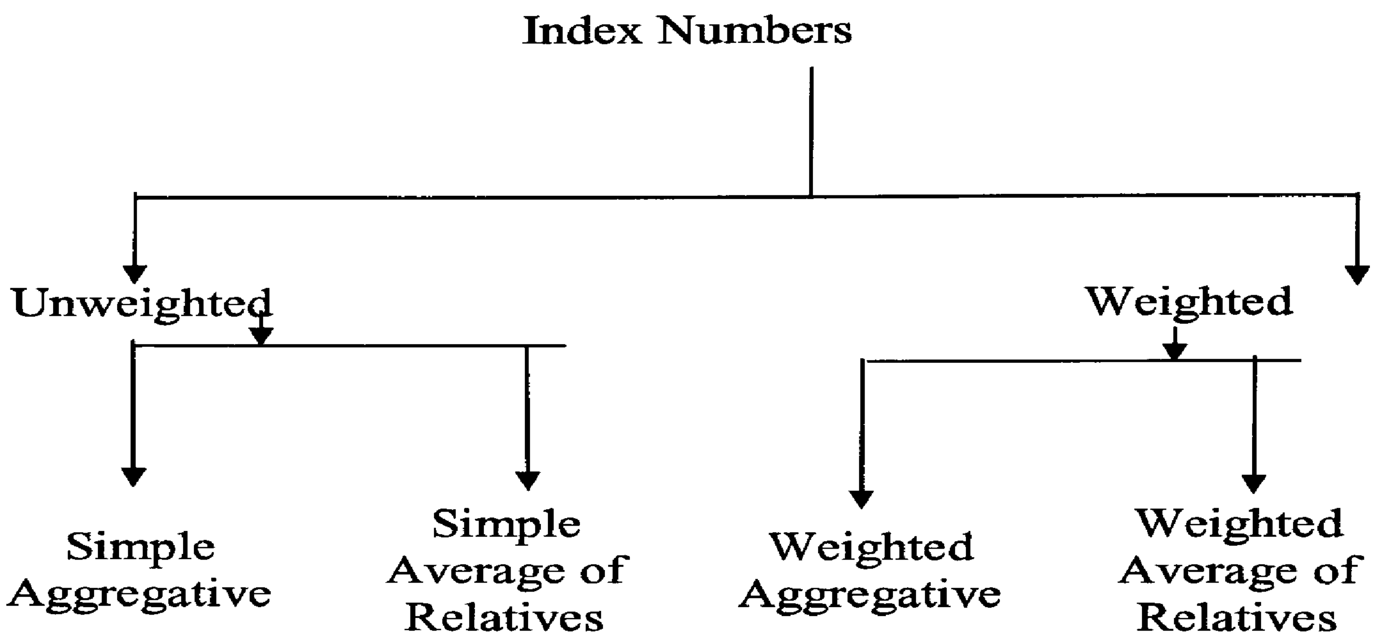
NSE took the lead and w.e.f. February 19,2008, S & P CNX NIFTY Shariah Index was launched. For the purpose of this index, the 50 companies comprising NIFTY are screened for Shariah compliances. Those companies fulfilling Shariah compliances form the S & P CNX NIFTY Shariah Index. Simultaneously, NSE also launched S & P CNX 500 Shariah Index for which Shariah compliant companies are screen from S & P CNX 500 Index.

BSE has also launched BSE TASIS Shariah 50 Index on December 27, 2010. All these Shariah indices are based on Free – Float Market Capitalisation weighted methodology.

Methods of constructing Index Numbers

A large number of formula have been devised for constructing index numbers. Broadly speaking, they can be rouged under two heads.

The following chart illustrates the various methods.



Unweighted Price Index Number

1. Simple Aggregative Method

This is simples method of constructing the index Numbers. The prices of the different commodities of the current year are added and the total is divided by the sum of the prices of the base commodity and multiplied by 100. Symbolically

$$P_{01} = \frac{\Sigma P_1}{\Sigma P_0} x100$$

Where P_1 – Current year price

P_0 – base year price

Illustration 1

From the following data constructed an index for 2012 taking 2011 as base.

Commodity	Price in 2011 (Rs.) P0	Price in 2012 (Rs.) P1
A	50	70
B	40	60
C	80	90
D	110	120
E	20	20
	P_0=300	P_1=360

Simple Aggregative Method

Price index Number $P_{01} = \frac{\Sigma P_1}{\Sigma P_0} x100$

$$= \frac{360}{300} x100 = 120$$

This means that as compared to 2011, in 2012 there is a net increase in the prices of commodities included in the index to the extent of 20%.

2. Simple average of price relative method

In this method, the price relative of each item is calculated separately and then averaged.

A price relative is the price of the current year expressed as a percentage of the price of the base year.

i) When arithmetic mean is used for averaging the price relatives, the formula for computing the index is

$$P_{01} = \frac{\Sigma \frac{P_1}{P_0} x100}{N}$$

Where N refers to the number of items.

ii) When geometric mean is used for overaging the price relatives the formula for computing the index is

$$P_{01} = \text{antilog} \left[\frac{\Sigma \log \frac{P_1}{P_0} x100}{N} \right]$$

Illustration 2

Compute a price index for the following by an average of price relative method by using both arithmetic mean and Geometric mean

Commodity	A	B	C	D	E	F
Price in 2012 (Rs.)	20	30	10	25	40	50
Price in 2013 (Rs.)	25	30	15	35	45	55

Solution

Computation for price index

Commodity	Price in 2012 P_0	Price in 2013 P_1	$P = \frac{P_1}{P_0} x100$	Log p
A	20	25	125	2.0969
B	30	30	100	2.0000
C	10	15	150	2.1716
D	25	35	140	2.1461
E	40	45	112.5	2.0511
F	50	55	110	2.0414
			P=737.5	logP=12.5116

a) Arithmetic mean of price relative index number

$$P_{01} = \frac{1}{N}\Sigma P \quad \text{Where } P = \frac{P_1}{P_0}x100$$

$$= \frac{737.5}{6}$$

$$= 122.92$$

b) Geometric mean of the price relative index number

$$P_{01} = \text{antilog}\left(\frac{1}{N}\Sigma \log P\right)$$

$$= \text{antilog}\left(\frac{12.5116}{6}\right)$$

$$= 121.7$$

Weighted Price Index Number(WPI)

Weighted index number is of two types. They are

(i) Weighted aggregate price index number

(ii) Weighted average of price relatives.

Weighted Aggregative Indices

These indices are of the simple aggregative type with the fundamental difference that weights are assigned to the various items included in various methods of assigning weights and consequently a large number of formulae for constructing index number have been devised of which some of the important methods are

1. Laspeyar's method
2. Paasche's Method
3. Dorbish and Bowley's method
4. Fisher's ideal method
5. Marshall-Edge worth method
6. Kelly's method, and
7. Walseh's method

1. Ley peyre's price index number

In this method, the base year quantities are taker as weights : symbolically,

$$P_{01} = \frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} x100$$

2. Paasche's Price Index Number

In this method, the current year quantities are taken as weights; symbolically

$$P_{01} = \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1} x100$$

3. Dorbish and Bowley's price index number

This is an index number got by arithmetic mean of Laspeyre's and Paache's price index numbers symbolically.

$$P_{01} = \frac{1}{2}\left(\frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} + \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1}\right) x100$$

4. Fisher's ideal price index number

Fisher's ideal index number is given by geometric mean by Laspeyre's and Paache's price index numbers symbolically

$$P_{01} = \sqrt{\frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} + \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1}} x100$$

5. Marshall's – Edgeworth's index number

In this method also both current year as well as base year prices and quantities are considered the formula for constructing the index is

$$P_{01} = \left(\frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} + \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1}\right) x100$$

6. Kelly's price index number

Kelly's index number uses quantities of some period as weights. This weight is kept constant for all periods, denoted by q.

$$P_{01} = \frac{\Sigma P_1 q}{\Sigma P_n q} x100$$

7. Walsch's price index number

Walsch's price index number is given by

$$P_{01} = \left(\frac{\Sigma P_1 (q_0 + q_1)}{\Sigma P_0 (q_0 + q_1)}\right) x100$$

Illustration 3

Construct the index number of price from the following data by applying

a) Laypeyre's method
b) Paasche's method
c) Bowley's method
d) Fisher's ideal method and
e) Marshall – Edgeworth method

Commodity	2011		2012	
	Price	**Quantity**	**Price**	**Quantity**
A	2	8	4	6
B	5	10	6	5
C	4	14	5	10
D	2	19	2	3

Solution

Calculation of various indices

Commodity	2011		2012		p_1q_0	P_0q_0	p_1q_1	P_0q_1
	Price p_0	Quantity q_0	Price P_1	Quantity q_1				
A	2	8	4	6	32	16	24	12
B	5	10	6	5	60	50	30	25
C	4	14	5	10	70	56	50	40
D	2	19	2	3	38	38	26	26
					p_1q_0 =200	p_0q_0 =160	p_1q_1 =130	p_0q_0 =103

a) Laypeyre's Price index

$$P_{01} = \frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} x100$$

$$= \frac{200}{1600} x100$$

$$= 125$$

b) Paache's price index

$$P_{01} = \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1} x100$$

$$= \frac{130}{103} x100$$

$$= 126.21$$

c) Bowley's price index number

$$P_{01} = \frac{\frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} + \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1}}{2} x100$$

$$= \frac{\frac{200}{160} + \frac{130}{103}}{2} x100$$

$$= \frac{1.25+1.262}{2} x100$$

$$= \frac{2.512}{2} x100$$

$$= 125.6$$

d) Fisher's ideal Index number

$$P_{01} = \sqrt{\frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} x \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1}} x100$$

$$= \sqrt{\frac{200}{100} x \frac{130}{103}} x100$$

$$= \sqrt{1.25 \text{ x } 1.262} x \ 100$$

$$= \sqrt{1.578} x100$$

$$= 1.2561 \text{ x } 100$$

$$= 125.61$$

e) Marshall – Edge worth method

$$P_{01} = \left(\frac{\Sigma P_1 q_0 + \Sigma P_1 q_1}{\Sigma P_0 q_0 + \Sigma P_0 q_1}\right) x100$$

$$= \left(\frac{200+130}{160+103}\right) x100$$

$$= \left(\frac{330}{263}\right) x100$$

$$= 1.25.48$$

Weighted average of price relatives

In practice, the Laspeyre's price index is usually calculated using price relatives. For this method we have to use the expenditure in the base year as weights. Symbolically

Weighted average of price relatives

$$P_{01} = \frac{\Sigma PV}{\Sigma V} \qquad \text{Where } P = \frac{P_1}{P_0} \text{x} 100$$

$$V = P_0 q_0$$

Using Geometric mean

$$P_{01} = \text{antilog}\left(\frac{\Sigma V \log P}{\Sigma V}\right)$$

$$\text{Where } P = \frac{P_1}{P_0} \text{x} 100$$

$$V = p_0 q_0$$

Illustration 4

Compute price index by applying weighted average of price relative method

Commodity	2012		2013	
	Price	Quantity	Price	Quantity
Wine	2.50	25	3	30
Beer	4.50	10	6	8
Soft Drink	0.60	10	0.84	15

Construction of price index

Commodity	2012		2013		$P = \frac{P_1}{P_0}$x100	$V=P_0q_0$	PV
	p_0	q_0	P_0	q_0			
Wine	2.5	25	3	30	120	62.50	7500
Bee	4.5	10	6	8	133	45.0	6000
Soft Drink	0.60	10	0.84	15	140	6.0	840
						EV=113.5	EPV=14340

Weighted average of price relative

$$P_{01} = \frac{\Sigma PV}{\Sigma V} \qquad \text{Where } p = \frac{p_1}{p_0} x100$$

$$V = P_0 q_0$$

$$= \frac{143.4C}{113.5C}$$

$$= 126.34$$

Weighted aggregate quantity index numbers

Price index numbers measure and permit comparison of the price of certain goods, the quantity index numbers permit comparison of the physical quantities of goods produced, consumed or distributed. The most common type of quantity index is that of quantity produced.

(i) Laspeyre's quantity index number

$$Q_{01} = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} x100$$

(ii) Paache's quantity index number

$$Q_{10} = \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1} x100$$

(iii) Bowley's quantity index number

$$Q_{01} = \frac{\left(\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}\right) + \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}{2} x100$$

(iv) Fisher's ideal quantity index number

$$Q_{01} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} x \Sigma \frac{q_1 p_1}{q_0 p_1}} x100$$

(v)

$$Q_{01} = \left(\frac{\Sigma q_1 p_0 + \Sigma q_1 p_1}{\Sigma q_0 p_0 + \Sigma q_0 p_1}\right) x100$$

Illustration 5

Compute quantity index by

a) Laspeyre's method

b) Paasche's method

c) Bowley's method

d) Fisher ideal method

e) Marshall Edge worth method

Commodity	Price P_0	2012 Total Value P_0q_0	Price P_1	2013 Total value p_1q_1
A	10	100	12	144
B	12	144	14	196
C	14	196	16	256
D	16	256	18	324
E	18	324	20	400

Solution

Construction of various quantity index

Commodity	2012		2013		q_1p_0	q_0p_0	q_1p_1	q_0p_1
	P_0	$q_0 = \frac{P_0q_0}{P_0}$	P_1	$q_1 = \frac{P_1q_1}{P_1}$				
A	10	10	12	12	120	100	144	120
B	12	12	14	14	168	144	196	168
C	14	14	16	16	224	196	256	224
D	16	16	18	18	288	256	324	288
E	18	18	20	20	36	324	400	360
					$q_1p_0 = 1160$	$q_0p_0 = 1020$	$q_1p_1 = 1320$	$q_0p_1 = 1160$

a) Vayperye's quantity index number

$$Q_{01} = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} x100$$

$$= \frac{1160}{1020} x100 = 113.70$$

b) Paache's quantity index number

$$Q_{01} = \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1} x100$$

$$= \frac{1320}{1160} x100 = 113.80$$

c) Bowley's quantity index number

$$Q_{01} = \frac{1}{2}\left(\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} + \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}\right) x100$$

$$= \frac{1}{2}\left(\frac{1160}{1020} + \frac{1320}{1160}\right) x100$$

d) Fisher ideal quantity Index number

$$Q_{01} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} x \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}} x100$$

$$= \sqrt{\frac{1160}{1020} x \frac{1320}{1160}} x100$$

$$= 11370$$

e) Marshall's – Edgeworth's quantity index number

$$G_{01} = \left(\frac{\Sigma q_1 p_0 + \Sigma q_1 p_1}{\Sigma q_n p_n + \Sigma p_n q}\right) x100$$

$$= \left(\frac{1160+1320}{1020+1160}\right) x100$$

$$= \left(\frac{2480}{2180}\right) x100$$

Value index

Value index number is easy to calculate, here, value is the product of price and quantity. The value of index V is the sum of the value of index V is the sum of the value of the current year is divided by sum of the value of the base year.

$$\text{Value index } V = \left(\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}\right) x100$$

Tests of consistency of index numbers

Several formulae have been suggested for constructing index numbers and the problem is that of selecting the most appropriate one in a given situation. The following tests are suggested for choosing an appropriate index.

1. Unit Test
2. Time Reversal Test
3. Factor Reversal Test
4. Circular Test

Unit Test

The unit test requires that the formula for constructing an index should be independent of the unit in which or for which, prices and quantities are quoted. Except for the simple (unweighted) aggregate index all other formulae discussed in this chapter satisfy this test.

Time Reversal Test

It is a test to determine whether a given method will work both ways in time forward and backward. The test provides that the formula for calculating the index number should be such that two ratios; the current on the base and the base on the current should multiply into unit. It other words, the two should be reciprocals of each other.

Symbolically

$P_{01} \times P_{10} = 1$

Fisher's ideal formula satisfies the time reversal test.

$$P_{01} x P_{10} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} x \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} x \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} x \frac{\Sigma p_0 q_0}{\Sigma p_0 q_1}}$$

$$= \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} x \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} x \frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} x \frac{\Sigma p_0 q_0}{\Sigma p_0 q_1}}$$

$$= \sqrt{1} = 1$$

(ie) $P_{01} \times P_{10} = 1$

The test is not satisfied by Laspeyre's method and the paache's method as can be seen below when Laypeyres method is used

$$P_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$$

$$P_{10} = \frac{\Sigma p_0 q_1}{\Sigma p_1 q_1}$$ changing time (i.e) 0 to 1 and 1 to 0

$$P_{01} \times P_{10} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \neq 1$$

When Paache's method used

$$P_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}$$

$$P_{10} = \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}$$

$$P_{01} \times P_{10} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0} \neq 1$$

Note : $P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$

$$P_{10} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}} \times 100$$

Changing time (i.e) 0 to 1 and 1 to 0.

Factor Reversal test

Another basic test is that the formula for index number ought to permit to interchanging the prices and quantities without giving inconsistent results. (ie) the two results multiplied together should give the true value ratio. A good index number should satisfy not only the time reversal test but also the factor reversal test. A good index number should allow time reversibility interchange of the base year and the current year, without giving inconsistent results.

$$P_{01} \times Q_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

The factor reversal test is satisfied only by the Fiser's ideal index.

For,

$$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$$

Changing P to q and q to p

$$Q_{01} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$

$$P_{01} \times Q_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$

$$= \sqrt{\frac{(\Sigma p_1 q_1)^2}{(\Sigma p_0 q_0)^2}}$$

$$= \frac{\Sigma P_1 q_1}{\Sigma P_0 q_0}$$

$$\therefore P_{01} \times Q_{01} = \frac{\Sigma P_1 q_1}{\Sigma P_0 q_0}$$

4. Circular Test

Another test of the adequacy of the index, number formula is what is known as 'circular test'. If in the use of interest attaches not merely to a comparison of two years but to the measurement of price changes over a period of years, it is frequently desirable to shift the base. A formula is said to meet this test if for example, 2013 index with 2010 as the base is 200 and the 2010 index with 2007 as the base must be 400. Clearly, the desirability of this property is that it enables us to adjust the index values from period to period without referring each time to the original base. A test of this shift ability of base is called the circular test.

Symbolically, if there are three indices P_{01}, P_{12} and P_{20} the circular test will be satisfied if

$P_{01} \times P_{12} \times 2_{01} = 1$

Illustration 6

Construction Fisher ideal index from the following data and show that it satisfies time reversal test and factor reversal test.

Items	2010 Base year		2012 Current year	
	Price	Quantity	Price	Quantity
A	10	40	12	45
B	11	50	11	52
C	14	30	17	30
D	8	28	10	29
E	12	15	13	20

Solution

Items	2010 Base year		2012 Current		p_1q_0	P_0q_0	p_1q_1	P_0q_1
	p_0	q_0	P_1	q_1				
A	10	40	12	45	480	400	540	450
B	11	50	11	52	550	550	572	572
C	14	30	17	30	510	420	510	420
D	8	28	10	29	280	224	290	232
E	12	15	13	20	195	180	260	240
					p_1q_0 =2015	p_0q_0 =1774	p_1q_1 =2172	p_0q_0 =1914

Fisher's ideal index

$$P_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} x \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} x100$$

$$= \sqrt{\frac{2015}{1774} x \frac{2172}{1914}} x100$$

$= 1.135 \ \ x\ 100$

$= 113.5$

Time reversal Test

Time reversal test is satisfied when $P_{01} \times P_{10} = 1$

$$P_{01} x P_{10} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} x \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} x \frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} x \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}}$$

$$= \sqrt{\frac{2015}{1774} x \frac{2172}{1914} x \frac{1914}{2172} x \frac{1774}{2015}}$$

$$= \sqrt{1}$$

$= 1$

Hence time reversal test is satisfied by the given data.

Factor reversal test

Factor reversal test is satisfied when $P_{01} x Q_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$

$$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} x \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} x \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} x \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$

$$= \sqrt{\frac{2015}{1774} x \frac{2172}{1914} x \frac{1914}{1774} x \frac{2172}{2015}}$$

$$= \sqrt{\left(\frac{2172}{1774}\right)^2}$$

$$= \frac{2172}{1774} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

$$\therefore P_{01} x Q_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

Hence factor reversal test is satisfied.

Illustration 7

Calculate fisher's ideal index from the following data and prove that it satisfied both the time reversal test and factor reversal test.

Items	Base year		Current year	
	Price	Expenditure	Price	Expenditure
A	8	80	10	120
B	10	120	12	96
C	5	40	5	50
D	4	56	3	60
E	20	100	25	150

Solution

Commodity	Base year Expenditure			Current year			p_1q_0	p_0q_0	p_1q_1	p_0q_1
	p_0	$V = p_0q_0$	$q_0 = \frac{V}{P_0}$	P_1	$V = p_1q_1$	$q_1 = V/p_1$				
A	8	80	10	10	120	12	100	80	120	96
B	10	120	12	12	96	8	144	120	96	80
C	5	40	8	5	50	10	40	40	50	50
D	4	56	14	3	60	20	42	56	60	80
E	20	100	5	25	150	6	125	100	150	120
							p_1q_0 =451	p_0q_0 =396	p_1q_1 =476	p_0q_0 =426

Fisher ideal index

$$P_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} x \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} x100$$

$$= \sqrt{\frac{451}{396} x \frac{476}{426}} x100$$

$$= \sqrt{1.2726} x100$$

$$= 1.128x100$$

$$= 112.8$$

Time Reversal test

Time reversal test is satisfied when $P_{01}xP_{10} = 1$

$$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} x \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}; P_{10} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} x \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}}$$

$$P_{01}xP_{10} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} x \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} x \frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} x \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}}$$

$$= \sqrt{\frac{451}{396} \times \frac{476}{426} \times \frac{426}{476} \times \frac{396}{451}}$$

$$= 1$$

$P_{01} \times P_{10} = 1$

Factor Reversal test is satisfied when

$$P_{01} \times P_{10} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

$$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}\ ;\ Q_{10} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$

$$P_{01} \times Q_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$

$$= \sqrt{\frac{451}{396} \times \frac{476}{426} \times \frac{\Sigma 426}{396} \times \frac{476}{451}}$$

$$= \sqrt{\frac{(476)^2}{(396)^2}}$$

$$= \sqrt{\left(\frac{476}{396}\right)^2}$$

$$= \frac{476}{396} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

$$P_{01} \times Q_{01} = \frac{\Sigma P_1 q_1}{\Sigma P_0 q_0}$$

Hence factor reversal test is satisfied.

Chain Base Index

The base may be fixed on changing. So far we have used fixed base method in various formulae. In the fixed base method, the base remains constant throughout, ie., the relative for all the year is based on the prices of that single year. On the other hand, in the chain base method, the relative for each year is found out from the prices of the immediately preceding year.

The indices, which we found out by this method, are called link relative index numbers or link relatives. The links are liked together for example, if we are to compute the index number for 2008, the base year is 2008; and index number for 2010, the base year is 2009, for 2011, the base year is 2010 and so on. The following formula is used for finding out the chain index.

$$\text{Chain index} = \frac{\text{Current year link relative x Previous year chain index}}{100}$$

Illustration 8

Construct (a) fixed base and (b) chain base index numbers from the following data relating to production of electricity.

Year	: 2004	2005	2006	2007	2008	2009	2010	2011	2012
Production (1000 kwt)	: 25	27	30	24	28	29	31	35	40

Solution

(a) Calculation of fixed base index number taking 2004 as base

Year	Production (1000 kwt)	Fixed Base index (2004) as Base
2004	25	$\frac{25}{25}$x100 = 100
2005	27	$\frac{27}{25}$x100 = 108
2006	30	$\frac{30}{25}$x100 = 120
2007	24	$\frac{24}{25}$x100 = 96
2008	28	$\frac{28}{25}$x100 = 112
2009	29	$\frac{29}{25}$x100 = 116
2010	31	$\frac{31}{25}$x100 = 124

2011	35	$\frac{35}{25}x100 = 140$
2012	40	$\frac{40}{25}x100 = 160$

In above,

$$\text{Fixed base index Number} = \frac{\text{Current year production}}{\text{Base year production}} x100$$

b) Calculation of Chain Base Index Number

$$\text{Chain base index number} = \frac{\text{(Link relative of curren year)x(Chair index of previous year)}}{100}$$

Year	Production (1000 kwt)	Link Relative $= \frac{\text{current year}}{\text{previous year}} x100$	Chain indices
2004	25	$\frac{25}{25}x100 = 100$	100
2005	27	$\frac{27}{25}x100 = 108$	$\frac{108x100}{100} = 108$
2006	30	$\frac{30}{27}x100 = 111.11$	$\frac{111.11x108}{100} = 120$
2007	24	$\frac{24}{30}x100 = 80$	$\frac{80x120}{100} = 96$
2008	28	$\frac{28}{24}x100 = 116.67$	$\frac{116.67 \text{ x } 96}{100} = 112$
2009	29	$\frac{29}{28}x100 = 103.57$	$\frac{103.5x112}{100} = 116$
2010	31	$\frac{31}{29}x100 = 106.90$	$\frac{106.90x116}{100} = 124$
2011	35	$\frac{35}{31}x100 = 112.90$	$\frac{112.90x124}{100} = 140$
2012	40	$\frac{40}{35}x100 = 114.29$	$\frac{114.29x140}{100} = 160$

Conversion of chain index to fixed index

At times it may be desired to covert the chain base index numbers into fixed base index numbers in such a case the following procedure is followed.

a. For the first year the fixed base index will be taken the same as the chain base index. However, if the index numbers are to be constructed by taking fixed year as the base in that case the index for the first year is taken as 100.

b. For calculating the indices for other years the following formula is used.

$$\text{Current year's FBI} = \frac{\text{Current year's CBI} \times \text{Previous year's FBI}}{100}$$

F.B.I. = Fixed Base Index Number

C.B.I. = Chain Base Index Number

Illustration 9

From the chain base index numbers given below, prepare fixed base index number.

Year :	2008	2009	2010	2011	2012
C.B.I :	80	110	120	90	140

Solution

Computation of Fixing Base Index Numbers

Year	Chain Bad Index Number	$F.B.I = \frac{\text{Current year CBI} \times \text{Previous year FBI}}{100}$
2008	80	80
2009	110	$\frac{110 \times 80}{100} = 88$
2010	120	$\frac{120 \times 88}{100} = 105.60$
2011	90	$\frac{90 \times 105.6}{100} = 95.04$
2012	140	$\frac{140 \times 95.04}{100} = 133.06$

Consumer price index numbers

Consumer price index is also called as the cost of living Index. Statisticians recommend that the term 'cost of living index' or 'Price of living index' or 'Cost of living Price index' or consumer price index can be used in appropriate place. In different countries, cost of living index consumer price index and retail price index are used.

Uses of consumer price index

1. This is very useful in wage negotiations and wage contracts and allowance adjustment in many countries.
2. Government can make use of these indices for wage policy, price policy, taxation general economic policies and rent control.
3. Changes in the purchasing power of money and real income can be measured.
4. We can analyze the market price of particular kinds of goods and services by this index.

Construction of a consumer price index(CPI)

There are some precautions to be taken in the consumer of price index.

1. Determination of the class of people
2. Selection of based period conducting family budget enquiry.

Methods of constructing price index

There are two methods of constructing price index. They are

1. Aggregative expenditure method
2. Family Budget method

1. Aggregative Expenditure method

This method is based upon the Laspeyre's method. It is widely used. The quantities of commodities consumed by a particular group in the base year are the weight.

$$\text{Consumer Price Index Number} = \frac{\sum p_1 q_0}{\sum p_0 q_0} x100$$

2. Family Budget Method

Here an aggregate expenditure of an average family on various item is estimated and its value weighted. Formula is

a) Consumer price index number $= \frac{\sum Pv}{\sum v}$

Where $P = \frac{P_0}{P_1} x100$; P_1 = Current year price

$V = p_0 q_0$ P0 = Base year price

b) Consumer price index number $= \frac{\sum Pw}{\sum w}$

Where $P = \frac{P_1}{P_0} x100$

Illustration 10

Compute consumer price index number from following.

Group	Base year price (Rs.)	Current Year Price (Rs.)	Weight (%)
Food	400	550	35
Rent	250	300	25
Clothing	500	600	15
Fuel	200	350	20
Entertainment	150	225	5

Solution

Construction of consumer price index

Group	Base year P_0	Current Year Price P_1	$P = \frac{P_1}{P_0} x100$	Weight W	PW
Food	400	550	137.5	35	4812.5
Rent	250	300	120.0	25	3000.0
Clothing	500	600	120.0	15	1800.0
Fuel	200	350	175.0	20	3500.0
Entertainment	150	225	150.0	5	750.0
				W=100	PW= 13862.5

Consumer Price Index Number

$$= \frac{\Sigma PW}{\Sigma W}$$

$$= \frac{13862.5}{100} = 138.625$$

Illustration 11

Construct a consumer price index number from the table given below.

Group	Index for 2010	Expenditure
Food	550	46%
Clothing	215	10%
Fuel and Lighting	220	7%
House Rent	150	12%
Misc	275	25%

Solution

Calculation o Cost of Living Index

Group	Index for 2010 I	Expenditure (%) W	IW
Food	550	46	25300
Clothing	215	10	2150
Fuel and Lighting	220	7	1540
House Rent	150	12	1800
Misc	275	25	6875
		W=100	IW=37665

$$\text{Cost of Living Index Number} = \frac{\Sigma IW}{\Sigma W}$$

$$= \frac{37665}{100}$$

$$= 376.65$$

Illustration 12

Compute the cost of living index number using both aggregate Expenditure method and family Budget method from the following information.

Commodity	Unit Consumption in base year	Price in base year	Price in Current year
Wheat	200	1.00	1.20
Rice	50	3.00	3.50
Pulses	50	4.00	5.00
Ghee	20	20.00	30.00
Sugar	40	2.50	5.00
Oil	50	10.00	15.00
Fuel	60	2.00	2.50
Clothing	40	15.00	18.00

Solution

Calculation of cost of living index by Aggregative Expenditure method

Commodity	q_0	p_0	p_1	p_1q_0	p_0q_0
Wheat	200	1.00	1.20	240	200
Rice	50	3.00	3.50	175	150
Pulses	50	4.00	5.00	250	200
Ghee	20	20.00	30.00	600	400
Sugar	40	2.50	5.00	200	100
Oil	50	10.00	15.00	750	500
Fuel	60	2.00	2.50	150	120
Clothing	40	15.00	18.00	720	600
				p_1q_0= 3085	p_0q_0= 2270

Aggregative Expenditure method cost of living index number

$$= \frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} x100$$

$$= \frac{3085}{2270} x100 = 135.9$$

Cost of Living index by family Budget method

Commodity	q_0	p_0	p_1	$\frac{P_1}{P_0}x100 = P$	$V=p_0q_0$	PV
Wheat	200	1.00	1.20	120	200	24000.00
Rice	50	3.00	3.50	116.67	150	17500.00
Pulses	50	4.00	5.00	125	200	25000.00
Ghee	20	20.00	30.00	150	400	60000.00
Sugar	40	2.50	5.00	200	100	20000.00
Oil	50	10.00	15.00	150	500	75000.00
Fuel	60	2.00	2.50	125	120	15000.00
Clothing	40	15.0	18.0	120	600	72000.0

		0	0			0
					V= 2270	PV= 3,08,50 0

Family Budget Method Cost of Living Index Number $= \frac{\Sigma PV}{\Sigma V}$

$$= \frac{3,08,500}{2270}$$

$$= 135.9$$

Production index: Index of industrial production

The index of the industrial production is aimed at reflecting changes (increase or decrease) in the volume of industrial production (i.e., production of non agricultural commodities) in a given period compared to some base period. These indices measure, at regular intervals, the general movement in the quantum of industrial production. Such indices are useful for studying:

i) The progress of general industrialization of a country, and

ii) The effect of tariff on the development of particular industries.

These indices of industrial activity are of great importance n the formulation and implementation of industrial plans. For the construction of indices of industrial production, the data about production of various industries are usually collected under following heads:

i) Textile industries: cotton, silk, woolen, etc.

ii) Metallurgical Industries: Iron and steel, etc.

iii) Mining Industries: coal, pig-iron and ferro-alloys, petrol, kerosene, copper(virgin metal) etc.

iv) Mechanical Industries: Locomotives, sewing machines, aerolanes, etc.

v) Industries subject to excise duty: Tea, sugar, cigarettes and tobacco, distilleries and breweries, etc.

vi) Electricity, gas and steam: Electric lamps, electric fans, electrical apparatus and appliances, etc.

vii) Miscellaneous: glass, paints and varnish, paper and paperboard, cement, chemicals etc.

Usually, the data (figures of output) are obtained for various industries on monthly basis and the indices of industrial production are obtained as the weighted arithmetic mean (or sometimes geometric mean) of the production (quantity) relatives by the formula:

$$I_{oj} \quad \frac{Q_j W_j}{W_j}$$

Where Q_j = production relative= $\frac{q_{ij}}{q_{oj}}$,

And W_j is the weight assigned to jth term (industry)

The weights may be assigned to various industries on the basis of, say, capital invested, net output, production etc. The concept of 'value added by manufacture' is the most commonly used criterion for determining the weights to be assigned to different industries.

Interim Index of Industrial Production (New Series): This official index number of industrial production using 1946 as base year was started (by the office of the Economic Adviser, Ministry of Commerce and industry, Government of India) from January 1949 and consisted of 20 items covering mining and manufacturing. The "value added by manufacture", as obtained from the First Census of Manufacturers 1946was used for determining the weights to be assigned to different items. The index of monthly production was computed by the formula

$$I_{oi} \quad \frac{W_j Q_j}{W_j} \quad 100.$$

Where Q_j = quantity relative for the jth industry for the month in question= $\frac{q_{ij}}{q_{oj}}$

W_j = weight allotted to jth industry

The adjusted index number, after allowing for variation in the number of days in the month for all the industri es except sugar, was obtained by the formula

$$Q_j \quad \frac{Q_j' \times \text{Average no.of days of a month ina year}}{\text{Calender no of days in the month}}$$

Where Q_j and Q_j' are adjusted and un-adjusted quantity relatives respectively.

For sugar, the seasonal variation adjusted indices were obtained by the formula

$$w_j \quad \frac{q_{oj}}{a_{oj}} \quad 3.54$$

Where w_j' is the adjusted weight; q_{oj} is the production in the corresponding month of the base year , a_{oj} is the average monthly production in the base year and 3.54 is the weight given to sugar industry in the overall weighting pattern.

Example

Calculate Index numbers of foodgrains production for the years 1980-81 and 1981-82 with base 1979-80=100 from the following data:

Foodgrains	Weight	Production(million tones)		
		1979-80	1980-81	1981-82
Rice	34	42	53	54
Wheat	12	29	32	35
Jowar	5	11	12	12
Bajra	3	6	5	6
Other cereals	6	11	12	13
Pulses	10	10	11	12

Solution. The indices food of grains production for the years 1980-81 and 1981-1982 with base year 1979-80, using fixed weights are obtained by Kelly's fixed weight formula.

CALCULATIONS FOR FOODGRAINS PRODUCTION INDICES

Food Grains	Weight (W)	Production (million tones)			Weight Production		
		1979-80 q_0	1980-81 q_1	1981-82 q_2	q_0W	q_1W	q_2W
1. Rice	34	42	53	54	1428	1802	1836
2. Wheat	12	29	32	35	348	384	420
3. Jowar	5	11	12	12	55	60	60
4. Bajra	3	6	5	6	18	15	18
5. Other Cereals	6	11	12	13	66	72	78
6. Pulses	10	10	11	12	100	110	120
					2015	2443	2532

$$Q^k_{01} = \frac{q_1W}{q_0W} \quad 100 \quad \frac{2443}{2015} \quad 100 = 121.2$$

$$Q^k_{02} = \frac{q_2W}{q_0W} \quad 100 \quad \frac{2532}{2015} \quad 100 \quad 125.6$$

Example

Index of Industrial Production covers three groups of industries. This index increased from 106.4 to 150.2 from one point of time to another. The index numbers of individual three groups of industries over the same period changed as follows: mining and Quarrying from 102.0 to 144;Manufacturing from 106.5 to 146.6; Electricity from 110.4 to 189.9.Determine the weights for the individual groups of industries.

Solution: Let the weights of three industries Mining and Quarrying, Manufacturing and Electricity be x,y and z respectively such that

$$x+y+z=100 \qquad z=100-x-y$$

CACULATIONS FOR INDEX NUMBERS OF INDUSTRIAL PRODUCTION

Industries	Weights (W)	Index No. (I_1)	I_1W	Index No. (I_2)	I_2W
M and Q	x	102.0	102x	144.1	144.1x
Manufacture	y	106.5	106.5y	146.6	146.6y
Electricity		110.4	110.4x	189.9	189.9x
	(100-x-y)		(100-x-y)		(100-x-y)

Initial index is given by :

$$\frac{I_1W}{W} \quad \frac{102x \quad 16.5y \quad 11.4(100 \quad x \quad y)}{100} \quad 106.4 \qquad \text{(Given)}$$

$$8.2x \quad 3.9y \quad 11040 \quad 100 \quad 106.4 \quad 10640$$

$$8.2x+3.9y-400=0 \qquad \ldots.(1)$$

Final index is given by

$$\frac{I_2W}{W} \quad \frac{144.1x \quad 146.6y \quad 189.9(100 \quad x \quad y)}{100} \quad 150.2 \qquad \text{(Given)}$$

$$45.8x \quad 43.3y \quad 18990 \quad 150.2 \quad 100 \quad 15020$$

$$45.8x+43.3y-3970=0 \qquad \ldots\ldots(2)$$

Multiplying (1) by 45.8 and (2) by 8.2, we get respectively

$$375.56x+178.62y-18320=0 \qquad \ldots(3)$$

$$375.56x+355.06y-32254=0 \qquad \ldots(4)$$

(4)-(3) gives

$176.44y-14234=0$

$$y = \frac{14234}{176.44} = 80.67 \qquad \ldots(5)$$

Substituting in (1) we get

$8.2x+3.9 \times 80.67-400=0$

$8.2x=400-314.61=85.39$

$$x = \frac{85.39}{8.2} = 10.41 \qquad \ldots(6)$$

Substituting in (8) , we have

$Z=100-x-y=100-10.41-80.67=8.92$

Hence the weights of three industries are:

Mining and Quarrying : 10.41 $\approx$ 10

Manufacturing : 80.67 $\approx$ 81

Electricity : 8.92 $\approx$ 9

Limitations of Index Numbers

Even though index numbers are very important in business and economic activities, they have their own limitations, they are

1. There may be error in each stage of the construction of the index number, namely selection of commodities, selection of base period, selection of weight etc.,
2. Index numbers may not present the exact change in price level, because they are based on sampled data.
3. Tastes, habits and customs of people change in course of time and may make the weighting not suitable for the present data
4. In each index there is an index, error, because there is no formula for measuring the price change. So there is the formula error. Hence it will not be a representative one.
5. By selecting a suitable year as the base year, selfish persons may get their desired results.

TIME SERIES

One o the most important tasks before economists and businessman these days is to make estimates for the future. Or example a business man is interested in finding out his likely sales in the year 2008 or as a long-term planning in 2020 or the year 2030 so that he could adjust his production accordingly and avoid the possibility of either unsold stocks or inadequate production to meet the demand similarly, an economist in interested in estimating the likely population in the coming year so that proper planning can be carried out with regard to good supply jobs for the people etc., However, the first step in making estimates for the future consists of gathering information from the past. In this connection one usually deals with statistical data which are collected, observed or recorded at successive intervals of time. Such data are generally referred to as time series.

Definitional of the term time series

A time series is a set of observations arranged in chronological order'.

Uses of time series

Time series analysis is useful in different fields like economics, science, research work, etc., because of the following reasons.

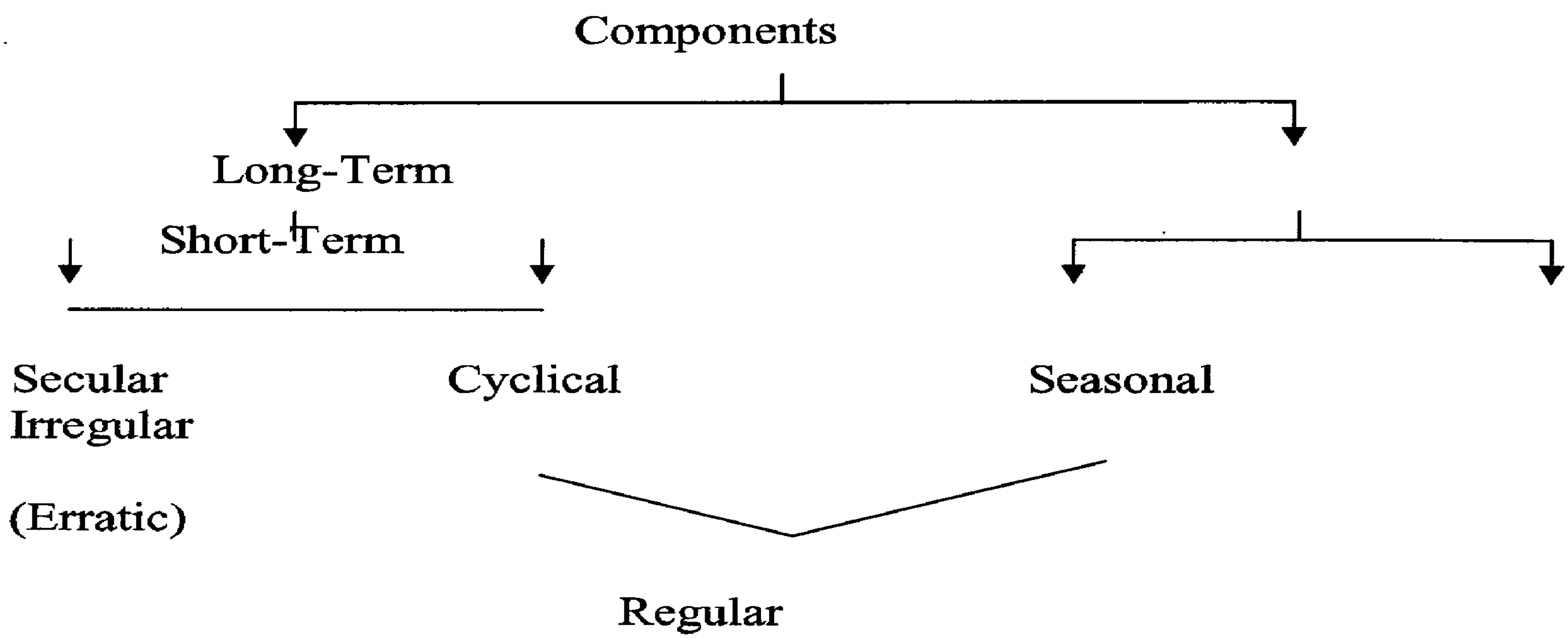
1. It helps in understanding the past behavior.
2. It helps in planning and forecasting
3. Comparison between data of one period with another period is possible.
4. It is useful not only to economists but also to the businessman.
5. It helps in evaluating current accomplishments.

Types of variations or components of time series

The components of a time series are

1. Secular trend
2. Seasonal variation
3. Cyclical variation and
4. Irregular variations

There are four basic types of variation and these are called the components or elements of time series.



1. SECULAR TREND

The general tendency of the time series data to increase or decrease or stagnate during a long period of time is called secular trend, also known as long term trend. The steady increase in the cost of living recorded by the consumer price index is an example of secular trend. From year, individual year, the cost of living varies a great deal, but we examine a long term period, we see that trend is toward a steady increase. It is an increasing but fluctuating time series.

Uses of Trend

1. The trend describes the basic growth tendency ignoring short term fluctuations.
2. It describes the pattern of behavior, which characterized the series in the past.
3. Future behavior can be forecasted in the assumption that past behavior will continue in the future also.
4. Trend analysis facilitate us to compare two or more time series over different period of time and this helps to draw conclusions about them.

2. SEASONAL VARIATIONS

These are small time fluctuations due to the peculiar conditions caused by seasonal effects. For example the price of rice will go down during the harvest season and again there will be rise after this season. This change occurs periodically. These kinds of variations changing with the seasons are called seasonal variation.

3. CYCLICAL VARIATIONS

This occurs primarily in economic time series and is similar to the seasonal variations. However it occurs periodically at intervals such longer than that found in seasonal variations. For example due to booms and depressions there will be steady upward and downward movements in economic time series.

Irregular variation

Due to war, earth quakes, floods and random factors, time series will change in an irregular fashion. Such changes are called irregular variation.

Methods of estimating trends

There are four methods of estimating secular trend

1. Free hand or Graphic method

2. Semi-average method
3. Moving average method
4. Method of least squares

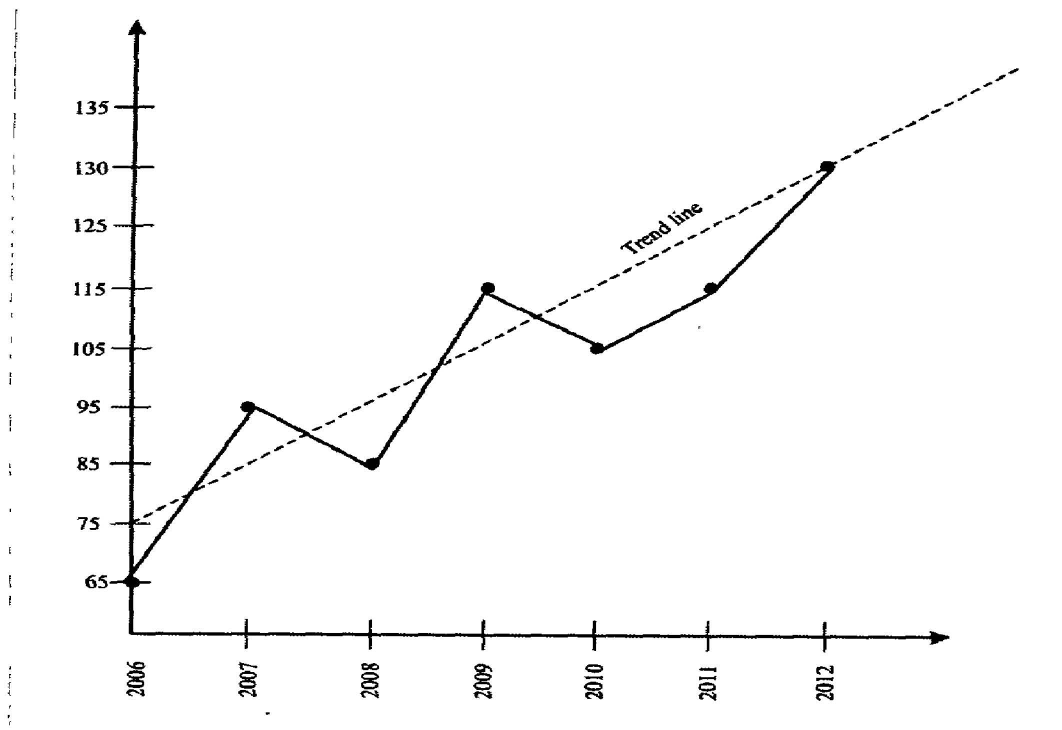
1. Free hand curve (or) Graphic fitting method

This is the easiest, simplest and the most flexible method of estimating secular trend. In this method we must plot the original data on the graphic. Draw a smooth curve carefully which will show the direction of the trend. In this the time is shown on the horizontal axis and the value of the variable is shown on the vertical axis.

Illustration

Fit a trend line to the following data by the tree-hand method.

Year	:	2006	2007	2008	2009	2010	2011	2012
Sales (1000 units)	:	65	95	85	115	110	120	130



Merits :

1. It is the simples, earliest and quickest method. It saves times and labour.
2. It is adaptable and flexible and it can be used to describe all types of trend (ie) linear and non linear.
3. Experienced statistician can draw a free hand line more accurately than a mathematician this can widely be used in applied situations.
4. It will help to understand the character of time series and we can use appropriate mathematical trend.

Demerits

1. It is highly subjective. It is subject to personal bias. The results depends upon the judgment of the person who draws the line. There may be different curves for different persons.
2. It seems to be very simple. But it requires more time for a careful job.
3. If it is not drawn by experienced person, the it is dangerous to use for forecasting purpose.
4. It does not help us to measure trend.

2. SEMI AVERAGE METHOD

In this method the original data are divided into equal parts and averages are calculated for both the parts. These averages are called semi averages.

Merits

1. It is simple and easier to understand than the moving average and least square method.

Illustration

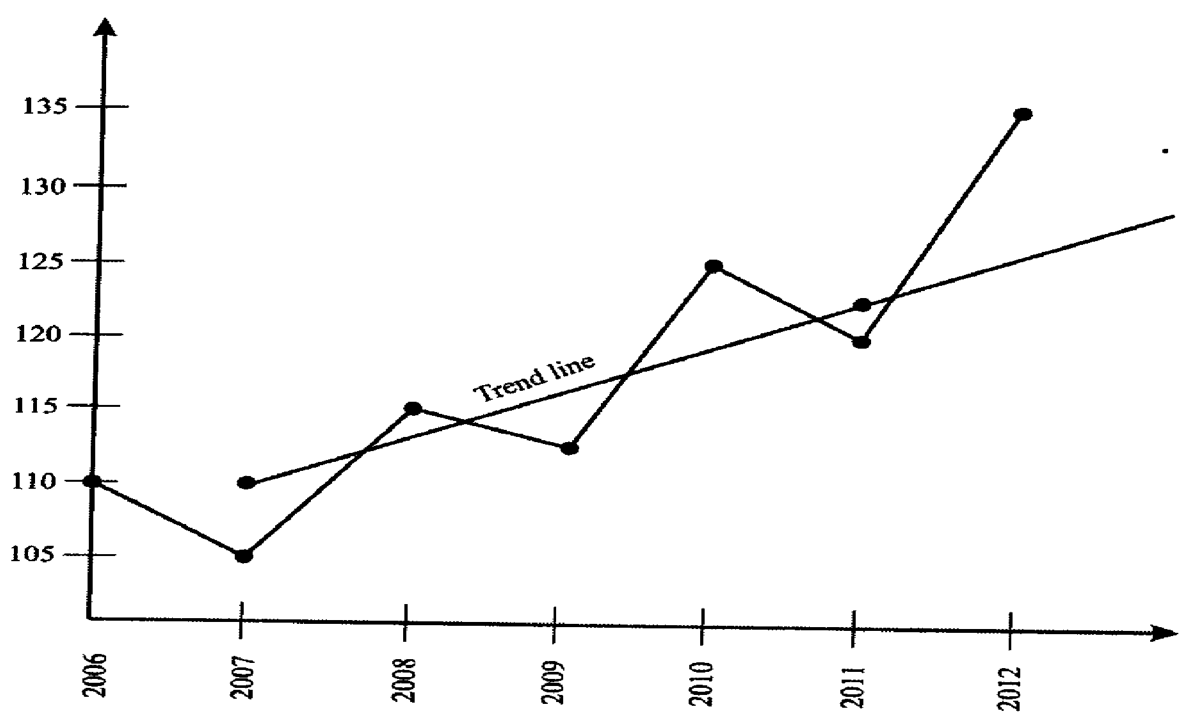
Draw a trend line by the method of semi averages.

Year : 2006 2007 2008 2009 2010 2011 2012

Sales (1000) : 110 105 115 112 120 118 130

Solution

Year	Sales (1000)
2006	110
2007	105 $\left.\right\}$ $\frac{330}{3} = 110$
2008	115
2009	112 middle term left out
2010	120
2011	118 $\left.\right\}$ $\frac{368}{3} = 122.6$
2012	130



3. Moving Average Method

In this method, the average value for a number of years or months or weeks is taken into account and making it at the centre of the tim-span (period of moving average) and it is the normal or tend value for the middle period.

Calculation of moving averages

The formula for calculating 3 yearly morning average is

$\frac{a+b+c}{3}, \frac{b+c+d}{3}, \frac{e+d+e}{3}$

The formula for calculating 4 yearly moving

$\frac{a+b+c+d}{4}, \frac{b+c+d+e}{4}, \frac{e+d+e+f}{4}$

2. As it does not depend upon personal judgment everyone who applies this method will get the same trend line unlike the former method.
3. As the line can be extended both ways, we can get the intermediate values and predict the future values.

Demerits

1. Under this method, it has an assumption of linear trend whether such a relationship exists or not.
2. It is affected by the limitation of arithmetic mean
3. The method is not enough for forecasting the future trend or for removing trend from original data.

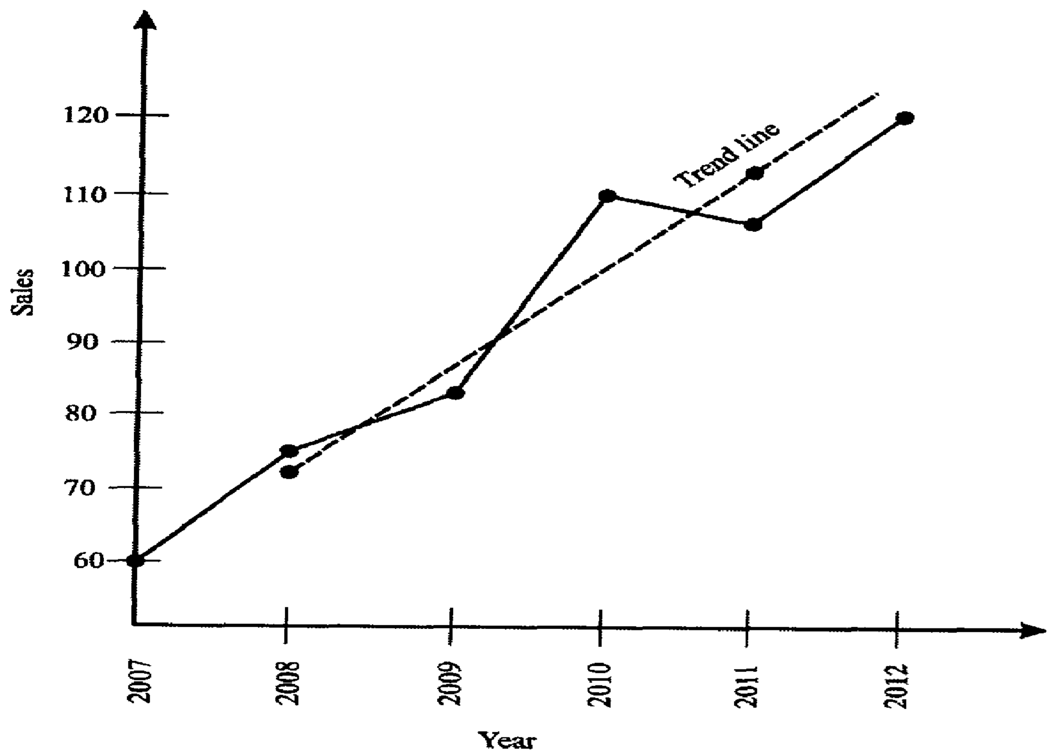
Illustration

Draw a trend line by the method of semi averages.

Year	:	2007	2008	2009	2010	2011	2012
Sales 1000	:	60	75	81	110	106	120

Year	Sales (1000)
2007	60 75 81 $\left.\right\} \frac{216}{3} = 72$
2008	
2009	

2010	110
2011	106 $\left.\right\} \frac{336}{3} = 112$
2012	120



Merits

1. It is sample and easy to understand it is easier to adopt when compared to the method of least square.
2. It is more flexible that other methods. It is elastic in the sense items can be increased or decreased without affecting the moving average, the only snag is that we get trend values or less trend value.
3. It is not only used for the measurement of trend but also for the measurement of seasonal, cyclical and irregular fluctuations.

Demerits

1. The main object of trend value is that it is used for forecasting or predicting future values, because this method is not represented by a mathematical function.

2. There is no rule regarding the choice of the number of the moving average and so the statistician has to use his own judgment.

Illustration

Calculate the 3 yearly moving average of the production figures given below.

Year 2000 2001 2002 2003 2004 2005 2006 2007 2008

Production Tonnes 15 21 30 30 42 46 50 56 63

Solution

Year	Production (in tones)	3 yearly total (m tones)	3 yearly moving average
2001	15	-	
2001	21	66	22.00
2002	30	87	29.00
2003	36	108	41.33
2004	42	124	58.67
2005	46	138	56.33
2006	50	152	63.00
2007	56	169	63.00
2008	63	189	-

Illustration

Assuming a four yearly cycle calculate the trend by the method of moving average from the following data relating to the production of tea in India.

Year 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

Production 464 515 518 467 502 540 557 571 586 612
(in millions)

Solution

Trend method of moving average

Year	Production	4 yearly total	2 figure moving total	Moving average centered
2003	464			
2004	515	1964		
2005	518	2002	3966	495.8
2006	467	2027	4029	503.6
2007	502	2066	4093	511.6
2008	540	2170	4236	529.5
2009	557	2254	4424	553.0
2010	571	2326	4580	572.5
2011	586			
2012	612			

Method of Least Squares

It is a mathematical as well as an analytical method. Under this method a straight line trend can be fitted to the given time series of data. This is a most important and accurate method of measures the long-term trend. In this method, we can fit either a straight line or a curve, which is the lies of best fit for the given data.

By making use of the equation of this line we get the trend values. For all the given year and also we can predict or forecast the future trend values. The equation of the line of best fit can be taken as $y=a+bx$

Where y denotes for the variable, x denotes the time period (which is deviation from a particular year) and a,b are constants to be foundout using the following tow normal equations

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

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

If we take x as the deviation from the middle year then $x=0$. In this case the normal eqn. becomes

$y=na$ and $xy = b\ \ x^2$

$a = \frac{\Sigma y}{n}$ $b = \frac{\Sigma xy}{\Sigma x^2}$

Substituting the value a and b we get the equation of the straight – line trend.

Illustration

Fit a straight line trend for the following data

Year	2005	2006	2007	2008	2009	2010	2011
Production in tonnes	62	70	75	814	89	93	104

Also find the tend values for the given years and estimates the trend value for the year 2014.

Solution

Let the equation of the straight line by $y=a+bx$. Where y denotes the production in tonnes. Here x is the deviation of the middle year.

The normal equations are

$$y=na+b\ \ x$$
$$xy=a\ \ x+b\ \ x^2$$

x year	y	X=X-2008	xy	x^2	Y=82+6.643x Trend values
2005	62	-3	-186	9	62.071
2006	70	-2	-140	4	68.714
2007	75	-1	-75	1	75.354
2008	81	0	0	0	82.000
2009	89	1	89	1	82.643
2010	93	2	186	4	95.286
2011	104	3	312	9	101.929
n=7	y=574	x=0	xy=186	x^2=28	

If $_x$=0, then

$$a = \frac{\Sigma y}{n} = \frac{574}{7} = 82$$

$$b = \frac{\Sigma xy}{\Sigma x^2} = \frac{186}{28} = 6.643$$

The eqn. of the straight line trend is y=82+6.643x

To find the trend values

For the year 2005, y=82+6.643(-3) = 62.071

For the year 2006, y=82+6.643(-2) = 68.714

For the year 2007, y=82+6.643(-1) = 75.357

For the year 2008, y=82+6.643(0) = 82.000

For the year 2009, y=82+6.643(1) = 88.643

For the year 2010, y=82+6.643(2) = 95.286

For the year 20011, y=82+6.643(3) = 101.929

To find the trend value for the year 2014,

For 2012, x=4

For 2013, x=5

For 2014, x=6

Then y = 82+6.643(6)

= 82+39.858

$= 121.858$

Estimated production for the year 2014 = 121.858 tonnes

Illustration

Using the method of least squares obtain the trend value for the following data.

Year	2005	2006	2007	2008	2009	2010	2011	2012
Earning (Rs. Lakhs)	38	40	65	72	69	60	87	95

Assuming the same rate of change continuous, what would be the predicted earning for the year 2014?

Solution

Let the equation of the straight line be

$$y=a+bx$$

The normal equations are

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

$$xy = a \quad x+b \quad x_2$$

FITTING OF STRAIGHT LINE TREND BY THE METHOD OF LEAST SQUARES

Year X	Earning (Rs. Lakhs) y	x=X-2008.5	Xy	x^2
2005	38	-3.5	-133.00	12.25
2006	40	-2.5	-100.00	6.25
2007	65	-1.5	-97.50	2.25
2008	72	-0.5	-36.00	0.25
2009	69	0.5	+34.50	0.25
2010	60	1.5	+90.00	2.25
2011	87	2.5	+217.50	6.25
2012	95	3.5	+332.50	12.25
	y=526	x=0	xy=308	x^2=42

If x=0 then

$$a = \frac{\Sigma y}{n} = \frac{526}{8} = 65.75$$

$$b = \frac{\Sigma xy}{\Sigma x^2} = \frac{308}{42} = 7.333$$

The equation of the straight line trend is

y=a+bx

y=65.75+7.333 x

For 2014, x will be 5.5

When x=5.5, y=65.75+7.333 x 5.5

= 65.75+40.336

= 106.086

Thus the estimated earning for the year 2014 are Rs. 106.086 Lakhs

Parabolic curve

Parabolas are non-linear, they from into smooth curves. The shape of these curves depends upon the value of the constants a,b,c etc. The general form of the equation of the power series $y=a+bx+cx^2+dx^3+\ldots$. The equation of this type does not represent a curve of strictly parable type but in common usage, the term parabolic curve is used to indicate curves obtained by equations of the type. The trend equation in this case is $y_c = a + bx + cx^2$

Where a is the trend value at the time origin

b is the slope at the origin and

c establishes whether the curve is up or down and how much.

The value of a,b and c can be determined by solving three normal equations simultaneously.

$y=na+b \ x+c \ x^2$

$xy = a \;\; x + b \;\; x^2 + c \;\; x^3$

$x^2y = a \;\; x^2 + b \;\; x^3 + c \;\; x^4$

In solving the above second degree equations much time and labour can be save by taking the time origin in the middle of the series so that x=0. But if x=0 then the sum of any add powers of x such as x^3 is also zero.

Therefore, the above three equations are reduced to

$y = na + c \;\; x^2$ (1)

$xy = b \;\; x2$ (2)

$x^2y = a \;\; x^2 + c \;\; x^4$ (3)

Solving equations (1) and (3) we obtain the value of a and c and the value of b is obtained by solving equation (2)

$$a = \frac{\Sigma y - c\Sigma x^2}{n}$$

$$b = \frac{\Sigma xy}{\Sigma x^2}$$

$$c = \frac{\Sigma x^2 y - a\Sigma x^2}{\Sigma x^4}$$

Illustration

Fit a parabola of the second degree to the data given below.

Year	1996	1997	1998	1999	2000
Sales (1000)	16	18	19	20	24

Solution

Year X	y (Sales)	X -x-1998	x^2	x^3	x^4	xy	x^2y	Trend Values
1996	16	-2	4	-8	16	-32	64	16.8
1997	18	-1	1	-1	1	-18	18	17.46
1998	19	0	0	0	0	0	0	19.12

1999	20	1	1	1	1	20	20	21.06
2000	24	2	4	8	16	48	96	23.06
N=5	y= 97	x=0	x^2=10	x^3=0	x^4 =34	xy =18	x^2y =198	

Let the equation of the second degree parabola be $y=a+bx+cx^2$

The normal equations are

$$y=na+b\ x+c\ x^2$$

$$xy = a\ x+b\ x^2+c\ x^3$$

$$x^2y=a\ x^2+b\ x^3+c\ x^4$$

If x=0 then

$$a = \frac{\Sigma y - c\Sigma x^2}{n}$$

$$= \frac{97-c(10)}{5}$$

$$= \frac{97-10c}{5}$$

5a = 97-10c

5a+10c = 97 (1)

$$b = \frac{\Sigma xy}{\Sigma x^2} = \frac{18}{10} = 1.8$$

$$c = \frac{\Sigma x^2 y - a\Sigma x^2}{\Sigma x^4} = \frac{198-a(10)}{34}$$

34c = 198-10a

10a+34c=198 (2)

Solving (1) and (2)

(1)x(2) 10a+20c= 194

(2) $\frac{10a+34c=198}{14c=4}$

Subtracting,

$$c = \frac{4}{14} = 0.29$$

C=0.29

Substituting the value of C in equation (1)

$5a+10(0.29) = 97$

$5a+2.9 = 97$

$5a=97-2.9$

$5a=94.1$

$$a = \frac{94.1}{5} = 18.82$$

$a=18.82;\ b=1.8;\ c=0.29$

$y=18.82+1.8x+0.29x^2$ is the required trend equation required trend equation.

Trend values are

$y1996 = 18.82+1.8(-2)+0.29(-2)^2=16.38$

$y1997 = 18.82+1.8(-1)+0.29(-1)^2=17.31$

$y1998 = 18.82+1.8(0)+0.29(0)^2=18.82$

$y1999 = 18.82+1.8(1)+0.29(1)^2=20.91$

$y2000 = 18.82+1.8(2)+0.29(2)^2=23.58$

Measurement of seasonal variations

In a time series seasonal variations come into force in regular periods. There is monthly or quarterly seasonal variations in the economic and business phenomena.

The objective of studying seasonal variation is to determine the effect of seasonal variation on the value of given phenomenal and to eliminate them (ie) determining the size of the value of variables.

The study of seasonal variation is important in deciding the business policy of various firms. The time series data are recorded monthly, quarterly, weekly, daily or hourly. There will be difference in them due to seasonal variations.

There are four methods. They are

1. Methods of simple average
2. Ratio to trend method

3. Ratio to moving average method
4. Link relative method

Method of Simple Average

This is the simplest and easiest method of calculating a seasonal index. The steps are

1. Average the data for each month or quarters for all the years.
2. Find the total of each month or quarter.
3. Divided each total by the number of years for which data are given. If we are given monthly data for 4 years, we must first get the total for each month for 4 years and divide each total 4 to get an average.
4. We can get an average of monthly averages by dividing the total of monthly average by 12.
5. We must taken the average of month averages as 100 and get the seasonal index as follows.

$$Seasonal\ Index\ for\ January = \frac{Monthly\ average\ of\ January}{Average\ of\ monthly\ average} x\mathbf{100}$$

Illustration

Compute the average seasonal movement seasonal index for the following data.

Year	1st quarter	2nd quarter	3rd quarter	4th quarter
2008	3.7	4.1	3.3	3.5
2009	3.7	3.9	3.6	3.6
2010	4.0	4.1	3.3	3.1
2011	3.3	4.4	4.0	4.0

Solution : Computation of Seasonal Indices

Year	1^{st} quarter	2^{nd} quarter	3^{rd} quarter	4^{th} quarter
2008	3.7	4.1	3.3	3.5
2009	3.7	3.9	3.6	3.6
2010	4.0	4.1	3.3	3.1
2011	3.3	4.4	4.0	4.0
Total	14.7	16.5	14.2	14.2
Average	3.675	4.125	3.55	3.55
Seasonal Index	98.66	110.74	95.30	95.30

The average of averages $= \frac{3.675+4.125+3.55+3.55}{4}$

$= \frac{14.9}{4}$

$= 3.25$

$$Seasonal\ Index = \frac{Quarterly\ average}{General\ average} x100$$

$$Seasonal\ Index\ for\ the\ first\ quarter = \frac{3.675}{3.725} x100$$

$= 98.66$

$$Seasonal\ Index\ for\ the\ second\ quarter = \frac{4.125}{3.725} x100$$

$= 110.74$

$$Seasonal\ Index\ for\ the\ third\ quarter = \frac{3.55}{3.725} x10$$

$= 95.30$

SUMMARY

From this unit we are able to understand what is index number, how it is constructed, how it is tested for perfection, and how to construct consumer price index.

KEYWORDS

1. Weighted index numbers
2. Aggregate index numbers

3. Value index numbers
4. Family based budget
5. Consumer price index

FURTHER READING

1. R.S.N. Pillai, Business Statistics
2. S.P. Gupta, Statistical Methods
3. Elhance, Statistics.

KEY TO CYP QUESTIONS

1. From the following data, compute price index by Simple Average of Price Relatives method based on (a) Arithmetic Mean, and (B) Geometric Mean.

Commodity	Price in 2004	Price in 2005
Butter	110	120
Cheese	75	80
Milk	13	13
Bread	9	9
Eggs	18	20
Ghee	850	860

Answer:
(a) Arithmetic mean=104.67
(b) Geometric Mean = 104.57

2. Construct Index Numbers of price from the following data by applying:
   1. Laspeyre's method
   2. Paasche's method
   3. Dorbish and Bowley's method

4. Fisher's Ideal nethod, and
5. Marshall- Edgeworth method

commodity	2004		2005	
	Price	quantity	Price	quantity
A	2	8	4	6
B	5	10	6	5
C	4	14	5	10
D	2	19	2	13

Answer:

1. Laspeyre's method = 125
2. Paasche's method = 126.21
3. Dorbish and Bowley's method = 125.61
4. Fisher's Ideal nethod = 125.60, and
5. Marshall- Edgeworth method = 125.48

EXERCISE

1. From the following data construct on index number for 1996 taking 1995 as base

Commodity and Unit	Price in 1995	Price in 1996
Butter(kg)	20.00	21.00
Cheese	15.00	14.00
Milk(lt)	3.00	3.00
Bread(i)	2.80	2.80
Eggs(doz)	6.00	8.00
Ghee(I tin)	250.00	260.00

2. Construct index number of price from the following data by applying.

   1. Laspeyre's method

2. Paasche's method
3. Dorbish and Bowley's method
4. Fisher's Ideal nethod, and
5. Marshall- Edgeworth method

Commodity	Price	quantity	Price	quantity
A	2	8	4	6
B	5	10	6	5
C	4	14	5	10
D	2	19	2	13

UNIT 5: TESTING OF HYPOSTHESIS

Introduction

A hypothesis is an assumption to the tested. The statistical testing of hypothesis is the most important technique in statistical inference. Hypothesis tests are widely used in business and industry of making decisions. It is here that probability and sampling theory plays an every increasing role in constructing the criteria on which business decisions are made. Very often in practice we are called upon to make decisions about population on the basis of sample information. For example, we may wish to decide on the basis of sample data whether a new medicine is really effective in curing a disease, whether one training procedure is better than another etc., such decisions are called statistical decisions.

Unit objectives

This unit addresses to the following objectives.

To understand the hypothesis testing procedure.

To know the concepts of parametric tests.

Unit Structure

Introduction

Unit Objectives

Unit structure

Hypothesis testing

Test of significance of large samples

t-test

f-test

Chi-square test

Exercise

Hypothesis Testing

In an statistical investigation the interest usually lies in the assessment of the general magnitude and the study of variation with respect to one or more characteristics relating to individual belonging to a group the group of individual under study is called population or universe. Population may be finite or infinite.

A finite subset of statistical individuals in a population is called sample.

The number of individuals in a sample is called the sample size.

Parameters and Statistic

In order to avoid verbal confusion with the statistical constants o the population, namely mean μ, variance ^{2} which are usually referred to as parameters. Statistical measures computed from sample observations alone. E.g mean $(\bar{x})$, variance (S^2) etc are usually referred to as statistic.

Null Hypothesis

For applying the test of significance, we first set up of a hypothesis – a definite statement about the population parameter. Such a hypothesis is usually a hypothesis of no-efficient and it is denoted by Ho.

In case of a single statistic, Ho will be that the sample statistic does not differ significantly from the hypothetical parameter value and into the case of two statistics (Ho) will be that the sample statistic do not differ significantly.

Alternative Hypothesis (H1)

Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis, usually denoted by H1.

For example, if we want to test the null hypothesis that the population has a specified mean μ_0 (say) (i.e) $Ho: \mu = \mu_0$, then alternative hypothesis would be

(i) $H_1 : \mu \neq \mu_0$

(ii) $H_1 : \mu > \mu_0$

(iii) $H_1 : \mu < \mu_0$

The alternative hypothesis (i) is known as a two tailed alternative (ii) is know as a right tailed alternative and (iii) is known as left tailed.

The setting of alternative hypothesis is very important to decide whether we have to use a single-tailed (right or left) or two tailed test.

Errors in Sampling

The main objective in sampling theory is to draw valid inferences about the population parameter on the basis of the sample results. In practice we decide to accept or to reject the lot after examing a sample from it. As such we have two types of error.

(i) Type I error : Reject H_0 when it is true

(ii) Type II error : Accept H_0 when it is wrong

Critical Region

A region corresponding to a statistic in the ample space S which lead to the rejection of H_0 is called critical region. Those region which lead to the acceptance of H_0 give us a region called acceptance region.

Level of Significance

The probability ' ' that a random value of the statistic't' belongs to the critical region is known as the level of significance. In other words, level of significance is the size of the Type I error. The levels of significance usually employed in testing of hypothesis are 5% and 1%.

Procedure for testing of Hypothesis

We now summarize below the various steps in testing of a statistical hypothesis in a systematic manner.

(i) Set up the null hypothesis

(ii) Choose the appropriate level of significance (either 5% or 1% level) this is to be decided before sample is drawn.

(iii) Compute the test statistic

$$Z = \frac{t-E(t)}{SE(t)} \text{ under the null hypothesis}$$

(iv) We compare the computed valve of z in step (iii) with the significance value at given level of significance.

If $|Z|<1.96$ H_0 may be accepted at 5% level of significance

If $|Z|>1.96$, H_0 may be rejected at 5% level of significance.

If $|Z|<2.58$, H_0 may be accepted at 1% level of significance

If $|Z|>2.58$ H_0 may be rejected at 1% level of significance

For a single tailed (Right or Left Tail) we compare the computed value of $|Z|$ with 1.645 (at 5% level) and 2.33 at (1% level) and accept or reject H_0 accordingly.

Large Samples

Test the significance of Large samples.

If the size o the sample n>30, then that sample is called large sample.

There are four important test to test the significance of large sample.

1. Test of significance for single proportion
2. Test of significance for difference of proportion
3. Test of significance for single mean
4. Test of significance for difference of mean

Test of Significance for Single Proportion

Suppose a large sample of size n is taken from a normal population.

To test the significant difference between the sample proportion p and the population proportion P, we use the statistic

$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$ Where n is the sample size.

Notes: Limits for population proportion P are given by $P \pm 3\sqrt{\frac{PQ}{n}}$ where q=1-p.

Illustration 1

In a sample of 1000 people in Karnataka 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat eaters are equally popular in this state at 1% level of significance?

Solution

Given n = 1000

p = sample proportion of rice eaters

$= \frac{540}{1000}$

= 054

P = population proportion of rice eaters

$$= \frac{1}{2} = 0.5$$

Q=1-P=1-0.5=0.5

Null hypothesis

Both rice and wheat eaters are equally popular in the state

(ie) H_0 : P=05

Alternative hypothesis H_1 : P≠0.5 (two tailed test)

$$Test\ Statistic\ Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.54-0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}}$$

$$= 2.532$$

Calculated value of Z=2.532

The tabulated value of Z at 1% level of significance for two tailed test is 2.58.

Calculated value of Z<Tabulated value of Z

Accept the Null Hypothesis H_0

Both rice and wheat eaters are equally popular in the state.

Illustration 2

In a big city 325 men out of 600 men were found to be smokers. Does this information the support the conclusion that the majority of men in this city are smokers?

Solution : Given n=600

p = simple portion of smokers

$$= \frac{325}{600}$$

$$= 0.5417$$

P = Population proportion of smokers in the city

$$= \frac{1}{2} = 0.5$$

Null Hypothesis H_0

The number of smokers and non-smokers are equal in city

Alternative Hypothesis

H_1:P>0.5 (Right-tailed)

Test Statistic

$$Z = \frac{P-Q}{\sqrt{\frac{PQ}{n}}} = \frac{0.5417-0.5}{\sqrt{\frac{0.5 \times 0.5}{600}}} = 2.04$$

Calculated value of Z = 2.04

Tabulated value of Z at 5% level of significance for right tailed test is 1.645.

Since calculated value of Z > Tabulated value of Z

We reject the null hypothesis.

(ie) The majority of men in the city are smokers

Illustration 3

A random sample of 500 pineapples was taken from a large consignment and 65were found to be bad. Find the percentage of bad pineapples in the consignment.

Solution

Given n=500

P = Proportion of bad pineapples in the sample

$$= \frac{65}{500} = 0.13$$

q = 1-P = 0.1-0.13

= 0.87

We know that the limits for population proportion P are given by $P \pm 3\sqrt{\frac{pq}{n}}$

$$= 0.13 \pm \sqrt{\frac{0.13 \times 0.87}{500}}$$

$$= 0.13 \pm 0.45$$

= (0.13 + 0.045, 0.13-0.45)

$= (0.175, 0.085)$

The percentage of bad pineapples in the consignment lies between 17.5% and 8.5%.

Test of Significance for difference of proportion

Suppose two large samples of sizes n_1 and n_2 taken respectly from 2 different populations to test the significant difference between the sample proportion P_1 and P_2 and

$$Z = \frac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Where $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ $and\ q = 1 - p$

Illustration 4

Random sample of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200 men and 325 women were in favour of the proposal. Test the hypothesis that proportions of men and women in favour of the proposal are same at 5% level.

Solution

Given sample size n_1=400

n_2=600

Proportion of me $P_1 = \frac{200}{400} = 0.5$

Proportion of women $P_2 = \frac{325}{600}$

$= 0.541$

Null hypothesis H_0 : Assume that there is no significant difference between the opinion of men and women as far as proposal of flyover is concerned

$H_0 : P_1 = P_2$

Alternative hypothesis $H_0 : P_1 \neq P_2$ (two tailed test)

The test statistic

$$Z = \frac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$\text{Where } p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$= \frac{400x0.5+600x0.541}{400+600}$$

$$= \frac{525}{1000}$$

$$= 0.525$$

$$q = 1\text{-}P$$

$$= 1\text{-}0.525 = 0.472$$

$$\therefore Z = \frac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.5 - 0.541}{\sqrt{(0.525)(0.735)\left(\frac{1}{400}+\frac{1}{600}\right)}}$$

$$= \frac{-0.041}{0.012}$$

$$= -1.28$$

$$|Z| \quad = 1.28$$

Calculated value of Z=1.28

The tabulated value of Z at 5% level of significance is 1.96

Calculate value of Z < tabulated value of Z

We accept null hypothesis

(ie) there is no significant difference of opinion between men and women as for as proposal of flyover is emeerned.

Illustration 5

Before an increase in excise duty on tea, 800 persons out of a sample 1000 persons were found to be a tea drinkers. After an increase in duty, 800 people were tea drinkers in a sample of 1200 people. Using standard error of proportion, state whether there is a significant decrease in the consumption of tea after the increase in excise duty.

Solution : By data,

$n_1 = 1000$ $\quad\quad\quad\quad$ $n_2 = 1,200$

$P_1 = \frac{800}{1000}$ $\quad\quad\quad\quad$ $P_2 = \frac{800}{1200}$

$= 0.8$ $\quad\quad\quad\quad$ $= 0.667$

$$Now\ P = \frac{n_1P_1+n_2P_2}{n_1+n_2}$$

$$= \frac{1000 x 0.8 x 1200 x 0.667}{1000+1200}$$

$$= \frac{1600}{2200} = 0.727$$

$$q = 1\text{-}p\text{=}1\text{-}0.727 = 0.273$$

Null hypothesis H_0 : Assume that there is no significant difference in the consumption of tea before and after the increase in excise duty.

$H_0 : P_1=P_2$

Alternative hypothesis $H_1:P_1>P_2$ (right tailed test)

The test statistic is $Z = \frac{P_1-P_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

$$= \frac{0.8-0.667}{\sqrt{0.727 x 0.273 \left(\frac{1}{1000}+\frac{1}{1200}\right)}}$$

$$= \frac{0.8-0.667}{0.019}$$

$$= 6.972$$

Calculated value of Z=6.972

Tabulated value of Z at 5% level of significance (right tailed test) is 1.64.

Calculated value of Z>Tabulated value of Z.

We reject the null hypothesis.

(ie) There is a difference in the consumption of tea before and after the increase in excise duty.

Note : If we want to test the significance of the difference between P_1 and P then

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

Then the statistic is $Z = \frac{P_1 - P}{\sqrt{\frac{n_2 pq}{n_1(n_1+n_2)}}}$

Illustration 6

In a random sample of 400 students of the university teaching department, it was found that 300 students failed in the Examination. In another sample of 500 students of the affiliated colleges the number of failures is the same examination was found to be 300. Find out whether the proportion of failures in the university teaching department significantly greater than the proportion of failures in the university teaching departments and affiliated colleges taken together.

Solution : By data,

Given $n_1 = 400$ $\qquad$ $n_2 = 500$

$P_1 = \frac{300}{400} = 0.75$ $\qquad$ $P_3 = \frac{300}{500} = 0.6$

$$p = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

$$= \frac{400\left(\frac{300}{400}\right) + 500\left(\frac{300}{500}\right)}{400+500}$$

$= 0.667$

q $= 1\text{-}P$

$= 1\text{-}0.667 = 0.333$

Null Hypothesis H_0 : Assume that there is no significant difference between P_1 and P

The test statistic is, $Z = \frac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$= \frac{0.75-0.667}{\sqrt{\frac{300x0.667 \; x \; 0.333}{400\;(400+500)}}}$$

$$= \frac{0.083}{0.0175} = 4.74$$

Calculate value of Z=4.74

Tabulated value of Z at 5% level of significance (Two tailed test) is 1.96.

Calculated value o Z>Tabulated value of Z.

Reject the null hypothesis

The proportion of failures in the affiliated colleges is greater than the proportion of failures in university departments and affiliated colleges taken together.

Test of significance for single mean

Suppose we want to test whether the given sample of size n has been drawn from a population with mean μ. We set up null hypothesis that there is no difference between $\bar{x}$ and $\bar{\mu}$ where $\bar{x}$ is the sample mean.

The test statistic is

$$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

Where is the S.D of the population?

If the population SD is not known, then use the statistic

$$\bar{x} = \frac{\bar{x}-H}{\frac{S}{\sqrt{n}}}$$

Where S is the sample S.D

Note : The values $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$ are called 95% financial limits or confidence limits for the mean of the population corresponding to the given sample.

$\bar{x} \pm 2.5\frac{\sigma}{\sqrt{n}}$ are called 99% confidence limits.

Illustration 7

A sample of 900 members has a mean of 3.4 cms and S.D. 2.61 cms. Is the sample from a large population of mean 3.25 cm and S.D. 2.61 cms. If the population is normal and its mean is unknown, find the 95% financial limits of the mean.

Solution

Given n=900 $\mu=3.25$

$\bar{x} = 3.4\,cm$ =2.61

S=2.61

Null Hypothesis

H_0 : Assume that the sample has been drawn from the population with mean $\mu=3.25$.

Alternative Hypothesis

$H_1 : \mu \neq 3.25$

The test statistic is

$$Z = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}$$

$$= \frac{3.4-3.25}{\frac{2.61}{\sqrt{900}}}$$

Calculated value of Z=1.724

Tabulated value of Z at 5% level of significance (two-tailed test) is 1.96.

Calculated value of Z<tabulated value of

We accept the null hypothesis H_0

(ie) The sample has been drawn from the population with mean $\mu=3.25$.

95% confidence limits are

$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}} = 3.4 \pm 1.96x\frac{2.61}{\sqrt{900}}$$

$= 3.4 \pm 0.1705$

$= (3.4+0.1705, 3.4-0.1705)$

$= 3.57$ and 3.2295

Illustration

An insurance agent has claimed that the average age of policy holders who issue through him is less than the average for all agents which is 30.5 years. A random sample of 100 policy holders who had issued through him gave to following age distribution.

Age	16-20	21-25	26-30	31-35	36-40
No. of persons	12	22	20	30	16

Calculate the Arithmetic mean and standard deviation of this distribution and use these values to test his claim at 5% level of significance.

Solution

Computation of mean and standard deviation.

Age X	**F**	**x=Midx**	**d=x-A =x-28**	**fd**	**d^2**	**fd^2**
15.5-20.5	12	18	-2	-24	4	48
20.5-25.5	22	23	1-	-22	1	22
25.5-30.5	20	28	0	0	0	0
30.5-35.5	30	33	1	30	1	30
35.5-40.5	16	38	2	32	4	64
	N=100			fd=16		fd^2=164

Mean $\bar{x} = A + \frac{\Sigma f d}{N} x c$

$$= 28 + \frac{16 x 5}{100} = 28.8$$

Standard deviation $S = \sqrt{\frac{\Sigma f d^2}{N} - \left(\frac{\Sigma f d}{N}\right)^2} x c$

$$= \sqrt{\frac{164}{100} - \left(\frac{16}{100}\right)^2} x\ 5$$

$$= 6.35$$

Null Hypothesis H_0 : The sample is drawn from the population with mean μ(i.e) $\bar{x}$ and μ do not differ significantly.

(ie) $H_0 : \mu=30.5$ years

Alternative hypothesis $H_1:\mu<30.5$ year (left-tailed test)

Now $\bar{x} = 28.8$, $S=6.35$

$\mu=30.5$ $n=100$

The test statistic is

$$Z = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}$$

$$= \frac{28.8-30.5}{\frac{6.35}{\sqrt{100}}} = -2.677$$

$|Z| = 2.677$

Calculated value of Z=2.677

Tabulated value of Z at 5% level of significance is 1.645 (left tailed test).

Calculated value of Z>Tabulated value of Z.

The null hypothesis H_0 is rejected.

(ie) $\bar{x}$ and μ differ significantly

(ie) the sample is not drawn from population with mean μ

Test the significance for difference of mean.

Let $\bar{x_1}$ be the mean of a sample of size n_1, from a population with mean μ_1 and variance σ_1^2.

Let $\overline{x_2}$ be the mean of a sample of size n_2 from a population with mean μ_2 and variance σ_2^2.

To test whether there is any significant difference between $\overline{x_1}$ and $\overline{x_2}$ we have to use the statistic

$$Z = \frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

Note : If the samples have been drawn from the same population then $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$Z = \frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}}$$

If is not known, we can use a ^{2} given to

$$\sigma^2 = \frac{n_1 s_1^2+n_2 s^2}{n_1+n_2}$$

Illustration

The mean of large samples 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples the regarded as drawn from the same population of S.D 2.5 inches.

Solution

Given $n_1 = 1000$ $n_2 = 2000$

$\bar{x} = 67.5$ Inches $\bar{x} = 65$ inches

Population standard deviation =2.5 inches Null Hypothesis H_0 : The samples have been drawn from the same population of S.D. 2.5 inches

$H_0 : \mu_1=\mu_2$

Alternative hypothesis $H_1:\mu_1\neq\mu_2$

To test statistic is

$$Z = \frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}}$$

$$= \frac{67.5-68}{\sqrt{\frac{(2.5)^2}{1000}+\frac{(2.5)^2}{2000}}}$$

$$= \frac{-0.5}{0.0968}$$

$$= -5.16$$

$$|Z| = 5.16$$

Calculated value of Z=5.16

Calculated value of Z>Tabulated value of Z.

We reject null hypothesis to

(ie) the samples are not drawn from the same population of S.D 2.5 inches.

Illustration

The mean yield of wheat from a district A was 210 pounds with standard deviation 10 pounds per acre from a sample of 100 plots. In another district the mean yield was 220 pounds with S.D. 20 pounds from a sample of 150 plots. Assuming that the S.D of yield in the entire state was 11 pounds. Test whether there is any significant difference between the mean yields of crops in the two districts.

Solution:

Given $\overline{x_1} = 210, \overline{x_2} = 200$

$n_1 = 100, n_2 = 150$

Population S.D. =11

Null hypothesis H_0 : There is no difference between $\overline{x_1}$ and $\overline{x_2}$

(ie) $H_0 : \overline{x_1} = \overline{x_2}$

Alternative Hypothesis H1 : $\overline{x_1} \neq \overline{x_2}$

The test statistic is

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

$$= \frac{210-200}{\sqrt{\frac{11^2}{100} + \frac{11^2}{150}}}$$

$$= \frac{10}{\sqrt{\frac{121}{100}+\frac{121}{150}}} = 7.04178$$

Calculated value of Z=7.04178

Tabulated value of Z at 5% level of significance is 1.96.

Calculated value of Z>Tabulated value of Z.

We reject the null hypothesis H_0.

(ie) There is a difference between $\overline{x_1}$ *and* $\overline{x_2}$

(ie) There is a significant difference between the mean yield of crops in the two district.

Test of Significance of small samples.

Definition

Small sample

When the size of the sample (ii) is less than 30 than that sample is called a small sample. The following are some important tests for small samples.

(i) Student's t-test

(ii) F-test

(iii) ^{2}-Test

Student's t-test

The student's t-test is defined by statistic

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

Where $\bar{x} = \frac{\Sigma x}{n}$; μ=population mean

$S^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$; n = sample size

Note :

1. If the standard deviation of a sample is given directly, than the statistic is given by

$$t = \frac{\bar{x} - \mu}{\frac{S.D}{\sqrt{n-1}}}$$

Confidence or fiducial limits for μ

(i) If $t_{0.05}$ is the table value of t for (n_{-1}) degrees of freedom at 5 level of significance then 95% confidence limit for μ is given by $\bar{x} \pm t_{0.05} \frac{S}{\sqrt{n}}$

(ii) 99% confidence limit for μ is

$\bar{x} \pm t_{0.01} \frac{S}{\sqrt{n}}$ where $t_{0.01}$ is the tabulated value of t for (n-1) degrees of freedom at 1% level of significance.

Illustration

The mean life of a sample of 20 fluorescent light bulbs produced y a company is computed to be 1570 hours with a S.D. of 120 hours. The company claims that the average life of the bulbs produced by the company is 1600 hours. Using to level of significance of 0.05, Is the claim acceptable?

Solution

Given,

n=sample size=25

$\bar{x}$=sample mean=1570

μ=population mean=1600

S=standard deviation=120

Null Hypothesis H_0

The claim is acceptable

(ie) H_0 : μ=11600 hours

Alternative hypothesis

$H_1 : \mu \neq 1600$ hours

The test statistic is $t = \dfrac{\bar{x}-\mu}{\frac{s}{\sqrt{n-1}}}$

$$= \frac{1570-1600}{\frac{120}{\sqrt{24}}}$$

$$= \frac{-30}{24.49}$$

$$= -1.22$$

$|t| = 1.22$

Calculated value of t=1.22

Degrees of freedom = n-l=25

The tabulated value of t at 5% level of significance with 24 degrees of freedom two tailed test is 2.06.

Calculated value of t<Tabulated value of t

We accept the null hypothesis H_0

(ie) the claim that the average life of the bulbs produced by the company is 1600 hours is acceptable.

Illustration

The mean weekly sales of soap bars in departmental stores were 146.3 bars per store. After an advertising campaign the mean weekly sales in 22 stores for a typical week increased to 153.7 and showed a S.D. of 17.2 was the advertising campaign successful.

Solution

n=number of stores = 22<40

The sample is small

Also given, sample mean $\bar{x} = 153.7$

Population mean μ=146.3

Sample S.D S=17.3

Since we known $\bar{x}$,μ S.D and n, we use student's t-test.

Null Hypothesis H_0 : the advertising campaign was not successful.

Alternative hypothesis H_1:μ>146.3 (Right tailed test)

$$\text{The test statistic } t = \frac{\bar{x}-\mu}{\frac{S.D}{\sqrt{n-1}}}$$

$$= \frac{153.7-146.3}{\frac{17.2}{\sqrt{22-1}}}$$

$$= 1.97$$

Calculated value of t-test = 1.97

Tabulated value of t at 5% level of significance with 21 degrees of freedom for right tailed test is 1.72.

Calculated value of t > tabulated value of t

We reject null hypothesis

Advertising campaign was successful

Illustration

A random sample of size 26 values from a normal population showed a mean of 53 and a sum of squares of deviation from the mean equals to 150. Can this sample be regarded as taken from the population having 56 as mean? Obtain 95% confidence limits of the mean of the population.

Given, n=sample size = 16

$$\bar{x} = sample\ mean = 53$$

$$\Sigma(x-\bar{x})^2 = 150$$

$$\therefore S^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{150}{16-1} = \frac{150}{15} = 10$$

S^2=10

S=$\sqrt{10}$

Null hypothesis H_0 : The sample is taken from the population having 56 as mean

(ie) H_0 : μ=56

Alternative hypothesis $H_1 : \mu \neq 56$.

The test statistic $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

$$= \frac{53 - 56}{\frac{\sqrt{10}}{\sqrt{15}}} = -3.79$$

$$|t| = 3.79$$

Calculated value of t-test = 3.79

Degrees of freedom = n-1=16-1=15

The tabulated value off t at 5% level of significance for 15 degrees of freedom for two tailed test is 2.13.

Calculated value of t-test > Tabulated value of t-test.

We reject the null hypothesis H_0

(i.e) the sample cannot be regarded as taken from the population

The 95% confidence limit of the mean of the population is given by $\bar{x} \pm t_{0.05} \frac{s}{\sqrt{n}}$

$$= 53 \pm 2.13x \frac{\sqrt{1C}}{\sqrt{1\epsilon}}$$

$$= 53 \pm 2.13x \frac{\sqrt{1C}}{4}$$

$= 53 \pm 1.6827$

$= (53+1.6827, 53-1.6827)$

= 54.68 and 51.31

Hence 95% confidence limit is [54.68, 51.31]

Students t-test (when S.D. of the samples is not given directly)

Illustration

A random sample of 10 boys had the following I.Q.'8 70,120,110,101,88,83,95,98,107,100. Do these data support the assumption of a population mean I.Q of 100? Find the reasonable range in which most of the mean I.Q values of samples of 10 boys lie?

Solution

Here S.D. and mean of the sample are not given directly. We have to determine these S.D. and mean as follows.

X	$x-\bar{x}=x-97.2$	$(x-\bar{x})^2$
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
x=972		$\Sigma(x-\bar{x})^2 = 1833.6$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{972}{10} = 97.2$$

$$S^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

$$= \frac{1833.60}{9}$$

$$= 203.73$$

$$S^2 = 203.73$$

$$S = \sqrt{203.73}$$

$$= 14.27$$

Null Hypothesis H_0 : The data support the assumption of a population mean I.Q of 100 in the population. (ie.) $H_0:\mu=100$

Alternative hypothesis H_1 : $\mu \neq 100$.

The test statistic $t = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}$

$$= \frac{97.2-100}{\frac{14.27}{\sqrt{10}}}$$

$$= \frac{-2.8}{4.514}$$

$$= -0.62$$

$$|t| = 0.62$$

Calculated value of t = 0.62

Degrees of freedom = n-1=10-1=9.

Tabulated value of t for 9 degrees of freedom at 5% level of significance is 2.26 (Two tailed test).

Calculate value of t-test < tabulated value of t

We accept the null hypothesis is H_0.

The data support the assumption of

Meem I.Q of 100 in the population.

The 95% confidence limits are given b

$$\bar{x} \pm t_{0.05} \frac{s}{\sqrt{n}} = 97.2 \pm 2.26\ x\ \frac{14.27}{\sqrt{10}}$$

$$= 97.2 \pm 2.26 \text{ x } 4.514$$

$$= 97.2 \pm 10.20$$

$$= 97.2 + 10.20,\ 97.2-10.2$$

$$= 107.41 \text{ and } 86.99$$

The 95% confidence limits within which the mean I.Q values of samples of 10 boys will be (86.99, 107.41).

Illustration

The height of 10 males of given locality are found to be 70,67,62,68,61,68,70,64,64,66 inches. Is it reasonable to believe that the average height is greater than 64 inches? Test at 5% level assuming that for 9 degrees of freedom P(t>1.83)=0.05.

Solution

Calculation for sample mean and SD

X	$x-\bar{x}=x-66$	$(x-\bar{x})^2$
70	4	16
67	1	1
62	-4	16
68	2	4
61	-5	25
68	2	4
70	4	16
64	-2	4
64	-2	4
66	0	0
$x=660$		$\Sigma(x-\bar{x})^2=90$

Null Hypothesis H_0

The average height is not greater than 64 inches.

(ie) H_0:μ=64 inches.

Alternative hypothesis H_1 : μ>64 inches

Test statistic $t=\frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}$

$=\frac{66-64}{\frac{3.16}{\sqrt{10}}}$

$=2$

Calculated value of t=2

Degrees of freedom = n-1

= 10-1=9

Tabulated value of t for 9 degrees of freedom at 5% level of significance for single tailed test is 1.83.

Calculated value of t>Tabulated value of t

We reject the null hypothesis is H_0

Student's t-test for difference of means

To test the significant difference between two means $\overline{x_1}$ and $\overline{x_2}$ of samples of sizes n_1 and n_2, use the statistic.

$$t = \frac{\overline{x_1}-\overline{x_2}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

When $S^2 = \frac{\Sigma(x_1-\overline{x_1})^2+\Sigma(x_2-\overline{x_2})^2}{n_1+n_1-2}$

(or) $S^2 = \frac{n_1 s_1^2+n_2 s^2}{n_1+n_2-2}$

S_1, S_2 are sample standard deviation, this case, degrees of freedom = $n_1+n_2=2$.

Illustration

Samples of two types of electric lights bulbs were tested for length of life and following data were obtained

	Type 1	**Type 2**
Samples Number	$n_1=8$	$n_2=7$
Samples mean	$\overline{x_1} = 1234\ hrs$	$\overline{x_2} = 1036\ hrs$
Sample SD	$S_1 = 30$ hrs	$S_2 = 40$ hrs

Is the difference in the mean sufficient to warrant that type I is superior to type II regarding length of life.

Solution :-

Given, $n_1=8$ $\quad n_2=7$

$\overline{x_1} = 1234$ $\quad \overline{x_2} = 1036$

$S_1=36$ $\quad S_2=40$

$$S^2 = \frac{n_1 S_1^2+n_2 S_2^2}{n_1+n_2-2}$$

$$= \frac{8(36)^2+7(40)^2}{8+7-2}$$

$$= 1659.08$$

$$S = \sqrt{1658.08}$$

Null Hypothesis H_0 : The two types I and II of electric bulbs are identical

Alternative hypothesis H_1 : $\mu_1 > \mu_2$

The test statistic $t = \frac{\overline{x_1} - \overline{x_2}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

$$= \frac{1234-1036}{\sqrt{1658.08} x \sqrt{\frac{1}{8}+\frac{1}{7}}}$$

$= 9.39$

Calculated value of t = 9.39

Degrees of freedom $= n_1+n_2-2$

$= 8+7-2$

$= 13$

Tabulated value of t for 13 degrees of freedom at 5% level of significance (one tailed test) is 1.77.

Calculated value of t>tabulated value of t

We reject the null hypothesis H0

(i.e) The two types I and II of electric bulbs are not identical.

Illustration

The means of two random samples of size 9 and 7 are 196.42 and 198.82 respectively. The sum of the squares of the deviation from the mean are 26.94 and 18.73 respectively can the sample be considered to have been drawn from the same normal population.

Solution

Given $n_1=9$ $\quad n_2=7$

$\overline{x_1} = 196.42$ $\quad \overline{x_2} = 198.82$

$\Sigma(x_1 - \overline{x_2})^2 = 26.94$ $\quad \Sigma(x_1 - \overline{x_2})^2 = 18.73$

$$S^2 = \frac{\Sigma(x_1-\overline{x_2})^2 + (x_1-\overline{x_2})^2}{n_1+n_{21-2}}$$

$$= \frac{26.94+18.73}{9+7-2} = 3.26$$

$$S^2 = 3.26$$

S=1.81

Null Hypothesis H_0

The two samples are drawn from the same population.

(ie) H0 : $\mu_1 = \mu_2$

Alternative hypothesis H_1: $\mu_1 \neq \mu_2$

The test statistic $t = \frac{\overline{x_1}-\overline{x_2}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

$$= \frac{196.42-198.82}{1.81\sqrt{\frac{1}{9}+\frac{1}{7}}}$$

$$= \frac{-2.4}{0.912} = -2.63$$

$|t| = 2.63$

Calculated value of t=2.6

Degrees of freedom = n_1+n_2-2

= 9+7-2

= 14

Tabulated value of t for 14 degrees of freedom at 5% level of significance is

2.15.

Calculated value of t> tabulated value of t.

We reject null hypothesis H_0.

(ie) The two samples are not drawn from the same population

Illustration

The nicotine content in milligrams of two samples of tobacco were found to be as follows.

Sample A	24	27	26	21	25	
Sample B	27	30	28	31	22	36

Can it be said that two samples come from normal population having the same mean.

Solution

Calculation for sample mean and S.D's

x_1	$x_1 - \overline{x_1}$	$(x_1 - \overline{x_2})^2$	x_2	$x_1 - \overline{x_1}$	$(x_2 - \overline{x_2})^2$
24	-0.6	0.36	27	-2	4
27	2.4	5.76	30	1	1
26	1.4	1.96	28	-1	1
21	-3.6	12.96	31	2	4
25	0.4	0.16	22	-7	49
$\Sigma x_1 = 12$		$\Sigma(x_1 - \overline{x_2})^2 = 21.2$			$\Sigma(x_2 - \overline{x_2})^2 = 21.2$

Mean $\overline{x_1} = \frac{\Sigma x_1}{n_1} = \frac{123}{5} = 24.6$

$\overline{x_2} = \frac{\Sigma x_2}{n_2}$

$= \frac{174}{6}$

$= 29$

$\Sigma(x_1 - x_2)^2 = 21.2$

$\Sigma(x_2 - x_2)^2 = 108$

$S^2 = \frac{\Sigma(x_1-\overline{x_2})^2 + (x_1-\overline{x_2})^2}{n_1+n_{21-2}}$

$= \frac{21.2+108}{5+6-2}$

$= 14.35$

$S^2 = 14.35$

$S = \sqrt{14.35} = 3.78$

Null hypothesis H_0

The two samples have been drawn from the normal population with the same mean

(ie) $H_0 : \mu_1 = \mu_2$

Alternative hypothesis $H_0 : \mu_1 \neq \mu_2$

The test statistic $t = \frac{\overline{x_1} - \overline{x_2}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$$= \frac{24.6-29}{3.78\sqrt{\frac{1}{5}+\frac{1}{n^2}}}$$

$$= 1.92$$

$$|t| = 1.92$$

Calculated value of t = 1.92

Degrees of freedom = n_1+n_2-2

$= 5+6-2=9$

Tabulated value of t for 9 degrees of 5% level of significance is 2.262.

Calculated value of t<Tabulated value of t

We accept the null hypothesis H_0

(ie) The samples come from normal population with the same mean.

F-Test

To test whether if there is any significant difference between two estimates of population variance (or)

To test if the two samples have come from the same population, we use F-Test.

In this case, we set up null hypothesis.

$H_0: \sigma_1^2 = \sigma_2^2$ (i.e) Population variances are same

Under H_0, the test statistic is

$$F = \frac{S_1^2}{S_2^2} \text{ if } \quad \text{if} \quad S_1^2 > S_2^2$$

Where $S_1^2 = \frac{\Sigma(x_1-\overline{x_1})^2}{n_1-1}$

$$S_2^2 = \frac{\Sigma(x_2-\overline{x_2})^2}{n_2-1}$$

The degrees of Freedom $\gamma_1 = n_1 - 1, \gamma_2 = n_2 - 1$

(or)

$$F = \frac{S_1^2}{S_2^2} \text{ if } \quad \text{if} \quad S_1^2 > S_2^2$$

Illustration

In one sample of 8 observation the sum of the squares of deviation o the sample values from the sample mean was 84.4 and in the other sample of 10 obervations it was 102.6 Test whether this difference is significant at 5% level.

Solution

Here $n_1=8$ $n_2=10$

$\Sigma(x_1 - \overline{x_1})^2 = 84.4$ $\Sigma(x_2 - \overline{x_2})^2 = 102.6$

$$S_1^2 = \frac{\Sigma(x_1-\overline{x_1})^2}{n_1-1} = \frac{84.4}{7} = 12.057$$

$$S_2^2 = \frac{\Sigma(x_2-\overline{x_2})^2}{n_2-1} = \frac{102.6}{9} = 11.4$$

Null hypothesis H_0

$\sigma_1^2 = \sigma_2^2$ Population variances are same.

Alternative hypothesis $H_1 = \sigma_1^2 \neq \sigma_2^2$

The test statistic

$$F = \frac{S_1^2}{S_2^2} = \frac{12.057}{11.4} = 1.057$$

Calculated value of F=1.067

Degrees of freedom, $\gamma_1 = n_1 - 1, = 8 - 1 = 7$

$\gamma_2 = n_2 - 1, = 10 - 1 = 9$

Tabulated value of F at 5% level of significance for (7,9) degrees of freedom is 3.29.

$F_{0.05}(7,9) = 3.29$

Calculated value of F < Tabulated value of F

We accept the null hypothesis H_0

There is no significant difference between population variances.

Illustration

It is known that the mean diameter of rivets produced by two firms A and B are practically the same but the standard deviation differ. For 22 riverts. Produced by firm A, the standard deviation is 2.99 mm and for 16 rivers produced by firm B and standard deviation is 3.8mm. Copute the statistic you would use to test whether the products of firm A have same variability at those of firm B and test its significance.

Solution

Given $n_1=22$ $n_2=16$

$S_1=2.9$ $S_2=3.8$

Here the S.D.'s of the samples S1 and S2 are given

The population variance S_1^2 and S_2^2 and obtained by using the relations

$$n_1S_1^2 = (n_1 - 1)S_1^2$$

$$= \frac{22x(2.9)^2}{n_1-1}$$

$$= 8.805$$

$$n_2S_2^2 = (n_2 - 1)S_2^2$$

$$= \frac{16x(3.8)^2}{16-1}$$

$$= 15.393$$

Null Hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$

(i.e) The product of both the firms A and B have the same variability.

Since $S_2^2 > S_1^2$,

The best statistic

$$F = \frac{S_2^2}{S_1^2}$$

$$= \frac{15.393}{8.805}$$

$$= 1.74$$

Calculated value of F=1.74

Degrees of freedom = (r_2, r_1)

$$= (n_{2-1}, n_{1-1})$$

Tabulated value of F with (15,21) degrees of freedom at 5% level of significance is 2.20.

Calculated value of F< Tabulated value of F

We accept the null hypothesis H_0

(ie) the products of both the firm A and B have the same variability.

Illustration

The time taken by workers in performing a job by method I and Method II is given below.

Method I	20	16	26	27	23	22	
Method II	27	33	42	35	32	31	38

Do the data show that the variances of time distribution from population from which these samples are drawn do not differ significantly?

Solution

Calculation of sample variances

x_1	$x_1 - \overline{x_1} =$	$(x_1 - \overline{x_2})^2$	x_2	$x_1 - \overline{x_1} = x$	$(x_2 - \overline{x_2})^2$
20	-2.3	5.29	27	-7.4	54.7
16	-6.3	39.69	33	-1.4	1.96
26	3.7	13.69	42	7.6	57.76
27	4.7	22.09	35	0.6	0.36
23	0.7	0.49	32	-2.4	5.76
22	-0.3	0.09	34	-0.4	0.16

			38	3.6	12.72
$\Sigma x_1 = 1$		$\Sigma(x_1 - \overline{x_2})^2 =$	$\Sigma x_2 = 34$		$\Sigma(x_2 - \overline{x_2})^2$

Given $n_1=6$, $n_2=7$

$$\overline{x_1} = \frac{\Sigma x_1}{n_1} = \frac{134}{6} = 22.3$$

$$S_1^2 = \frac{\Sigma(x-\overline{x_1})^2}{n_1-1}$$

$$= \frac{81.34}{5}$$

$$= 16.26$$

$$\overline{x_2} = \frac{\Sigma x_2}{n_2} = \frac{241}{7} = 34.4$$

$$\Sigma(x_1 - \overline{x_1})^2 = 81.34; \quad \Sigma(x_2 - \overline{x_2})^2 = 133.72$$

$$S_1^2 = \frac{\Sigma(x_1-\overline{x_1})^2}{n_1-1}$$

$$= \frac{81.34}{5}$$

$$= 16.26$$

$$S_2^2 = \frac{\Sigma(\overline{x_2}-\overline{x_2})^2}{n_2-1}$$

$$= \frac{133.72}{6}$$

$$= 22.29$$

Null Hypothesis H_0 : There is no significant difference between population variances.

(ie) $H_0: \sigma_1^2 = \sigma_2^2$

Alternative hypothesis $H_1: \sigma_1^2 \neq \sigma_2^2$

Since $S_2^2 > S_1^2$, the test statistic

$$F = \frac{S_2^2}{S_1^2}$$

$$= \frac{22.29}{16.26E}$$

$$= 1.37$$

Calculated value of F=1.37

Degrees of freedom $= (n_2 - 1, n_1 - 1)$

$= (7-1, 6-1)$

$= (6,5)$

Tabulated value of F for (6,5) degrees of freedom at 5% level of significance is 4.95.

Calculated value of F test < Tabulated value

We accept the null hypothesis H_0

(ie) There is no significant difference between the variances of the time distribution by the workers.

Illustration

The nicotine contents in milligrams in two samples of tobacco were found to be as follows.

Sample A	24	27	26	21	25	
Sample B	27	30	28	31	22	36

Can it be said that two samples come from same normal population.

Solution

To test whether the two samples come from the same normal population we have to test

(i) The equality of variance by using F-Test

(ii) The equality of means by using t-test

Given $n_1=5$, $n_2=6$

Calculations for mean's and S.D's of the samples

x_1	$x_1 - \overline{x_1} =$	$(x_1 - \overline{x_2})^2$	x_2	$x_1 - \overline{x_1} =$	$(x_2 - \overline{x_2})^2$
24	-0.6	0.36	27	-2	4
27	2.4	5.76	30	1	1
26	1.4	1.96	28	-1	1
21	-3.6	12.96	31	2	4
25	0.4	0.16	32	-7	49
			36	7	49
$\Sigma x_1 =$		$\Sigma(x_1 - \overline{x_2})^2 = 21.2$	$\Sigma x_2 = 174$		$\Sigma(x_2 - \overline{x_2})^2 =$

$$\overline{x_1} = \frac{\Sigma x_1}{n_1} = \frac{123}{5} = 24.6$$

$$\Sigma(x_1 - \overline{x_1})^2 = 21.2$$

$$S_1^2 = \frac{\Sigma(x_1 - \overline{x_1})^2}{n_1 - 1} = \frac{21.4}{4} = 5.3$$

$$\overline{x_2} = \frac{\Sigma x_2}{n_2} = \frac{108}{5} = 21.6$$

$$\overline{S_2} = \frac{\Sigma(x_2 - \overline{x_2})^2}{n_2 - 1} = \frac{108}{5} = 21.6$$

(i) F-Test

Null Hypothesis $H_0: \sigma_1^2 = \sigma_2^2$

(ie) The population variances are equal

If $S_2^2 > S_1^2$, the test statistic

$$F = \frac{S_2^2}{S_1^2}$$

$$= \frac{21.6}{5.3}$$

$= 4.07$

Calculated value of F = 4.07

Degrees of freedom = $(n_2\text{-}1, n_1\text{-}1)$

= (5,4)

Tabulated value of F for (5,4) degrees of freedom at 5% level of significance is 6.26 calculated value of F< Tabulated value of F.

We accept the null hypothesis H_0

Population variances are equal

(ii) t-test,

Refer illustration

CHI-SQUARE TEST OF GOODNESS OF FIT

Suppose we are given a set of observed frequencies obtained under some experiment and we want to test if the experimental results support a particulars theory or hypothesis. Karal pearson developed a test for testing the significance of discrepancy between experimental values and the theoretical values obtained under some theory or hypothesis. This test is known as ^{2} test of goodness of fit. Karl Pearson proved that the statistic.

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

Where O-Observed frequency

E – Expected frequency

^{2} is used to test whether differences between observed and expected frequencies are significant.

If the data is given in a series of 'n' number then degrees of freedom = n-1

In the case of Binomial distribution d.f = n-1

In the case of Poisson distribution d.f=n-2

In the case of Normal distribution d.f=n-3

Illustration

The number of automobile accidents per week in a certain community are as follows 12,8,20,2,14,10,15,6,9,4 Are these frequencies in agreement with the belief that accident conditions were the same during this 10 week period.

Solution

Expected frequencies of accident

$$Each\ week = \frac{12+8+20+2+14+10+15+6+9+4}{10}$$

$$= \frac{100}{10} = 10$$

Null hypothesis H_0 : The accident conditions were the same during the 10 week period.

Observed frequency	Expected Frequency E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
12	10	2	4	0.4
8	10	-2	4	0.4
20	10	10	100	10.0
2	10	-8	64	6.4
14	10	4	16	1.6
10	10	0	0	0
15	10	5	25	2.5
6	10	-4	16	1.6
9	10	-1	1	0.1
4	10	-0	36	3.6
	100			$\Sigma\frac{(O-E)^2}{E} = 26.6$

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

$$= 26.6$$

Calculated value of ^{2}-Test = 26.6

Degrees of freedom = n-1

= 10-1

'= 9

Tabulated value of ^{2} Test for 9 degrees of freedom at 5% level of significance is 16.9.

Calculated value of ^{2} test > Tabulated value of ^{2}=test

We reject the null hypothesis H_0

(ie) The accident conditions were not the same during the 10 week period.

Illustration

The theory predicts the proportion of beans in four groups A,B,C and D should be 9:3:3:1. In an experiment with 1600 beans the numbers in the four groups were 382, 313, 287 and 118. Does the experimental result support the theory.

Solution: If we divide (382+313+287+118=1600) in the ratio 9:3:3:1 we get the expected frequencies are 900, 300, 300, 100.

$$\frac{Total}{9+3+3+1} = \frac{1600}{16} = 100$$

9 Parts = 900

3 Parts = 300

3 Parts = 300

1 Part = 100

Observed frequency O	Expected Frequency E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
882	900	18	324	0.360
313	300	13	169	0.563
287	300	13	169	0.563
118	100	18	324	3.240
				$\Sigma\frac{(O-E)^2}{E} = 4.726$

The test statistic $\Psi^2 Test = \sum \frac{(O-E)^2}{E}$

$= 26.6$

Calculated value of ^{2}-Test = 4.726

Degree of freedom = n-1 = 4-1 = 3

Tabulated value of ^{2} Test for 3 degrees of freedom at 5% level of significance is 7.81.

Calculated value of ^{2}-test < Tabulated value of ^{2}-test

We accept the null hypothesis H_0.

To the experiment support the theory

CHI SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES

Definition : Literally, an attribute means a quality or characteristic. Example of attributes are drinking, smoking, blindness, honesty etc.,

An attribute may be marked by its presence (position) or absence in a number of a given population.

Let us consider two attributes A and B. A is divided into two classes and B is divided into two classes. The various cell frequencies can be expected in the following table known as 2x2 contingency table.

A	a	b
B	c	d

			Total
	a	b	a+b
	c	d	c+d
Total	a+c	b+d	N=a+b+c+d

The expected frequencies are given by

$E(a)=\frac{(a+c)(a+b)}{N}$	$E(b)=\frac{(b+d)(a+b)}{N}$	a+b
$E(c)=\frac{(a+c)(a+d)}{N}$	$E(d)=\frac{(b+d)(c+b)}{N}$	c+d
a+c	b+d	N=Total frequencies

Null hypothesis H_0 : Attributes are independent

Degrees of freedom = (r-1) (s-1)

Where r = number of rows

S = number of columns

Illustration

The following table gives the classification of 100 workers according to sex and nature of work. Test whether the nature of work is independent of the sex of the workers.

Males	Stable	Unstable	Total
Males	40	20	60
Females	10	30	40
Total	50	50	100

Solution

Null hypothesis H_0 : Attributes are independent

(ii) The nature of work is independent of the sex of the workers.

Expected frequencies are given in the table

$\frac{50x60}{100}=30$	$\frac{50x60}{100}=30$	60
$\frac{50x40}{100}=20$	$\frac{50x40}{100}=20$	40
50	50	100

Calculation of ^{2}

O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
40	30	10	100	3.333
20	30	-10	100	3.333
10	20	-10	100	5.000
30	20	10	100	5.000
				$\Sigma\frac{(O-E)^2}{E} = 16.666$

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

$$= 16.666$$

Calculated value of ^{2}-Test = 16.666

Degrees of freedom = (r-1) (S-1)

= (2-1) (2-1)

= 1x1=1

Tabulated of ^{2}-test for 1 degree of freedom at 5% level of significance is 3.84

Calculated value of ^{2}>tabulated value of ^{2}

We reject the null hypothesis H_0

Attributes are not independent

(ie) The nature of work is not independent of the sex of the workers.

Illustration

In a certain sample of 2000 families 1400 families are consumers of tea. Out of 1800 Hindu families, 1236 families are consumers of tea. Use ^{2}-test and state whether there is any significant difference between consumption of tea among Hindu and non-Hindu families.

Solution: By Data,

	Hindu	Non Hindu	Total
Families consuming tea	1236	164	1400
Families not consuming tea	564	36	600
Total	**1800**	**200**	**2000**

Null hypothesis H_0 : There is no significant difference between the consumption of tea among Hindu and Non Hindu families the expected frequencies are

	Hindu	Non-Hindu	Total
Consuming tea	$\frac{1800x1400}{2000} = 1260$	$\frac{200x400}{2000} = 140$	1400
Not consuming tea	$\frac{1800x600}{2000} = 540$	$\frac{200x600}{2000} = 60$	600
Total	**1800**	**200**	**2000**

O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
1236	1260	-14	576	0.4571
164	140	14	576	1.0666
564	540	14	576	4.1143
36	60	-14	576	9.6
				$\Sigma\frac{(O-E)^2}{E} = 15.238$

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

$$= 15.238$$

Calculated value of $^2 = 15.238$

Degrees of freedom $= (r-1)(s-1)$

$= (2-1)(2-1)$

$= 1$

Tabulated value of ^{2}-test for 1 degree of freedom at 5% level of significance is 3.841.

Calculated value of ^{2}>Tabulated value of ^{2}

We reject the null hypothesis H_0

(ie) Both Hindu and Non-Hindu families differ significantly as regard the consumption of tea among them.

SUMMARY

This chapter summarizes the t-test, F-test, Chi-square test are commonly used in business research.

KEYWORDS

Null Hypothesis

Level of significance

Degrees of freedom

Table values

Sample test value

CHECK YOUR PROGRESS QUESTIONS

1. A machine which produces mica insulating washers for use in elastic devices is set to turn out washers having a thickness of 10mm. A sample of 10 washers has an average thickness 9.522mm with a S.D. of 0.6mm calculate student's t-test.
2. A random sample of size 20 from a normal population gives a sample mean of 42 and sample S.D 6-Test the hypothesis that the population mean is 44.
3. The average number of articles produced by two machines per day is 200 and 250 with S.D. 20 and 25 respectively on the basis of regards of 25 days production. Can you regard both the machines equally efficient of 1% level of significance?

4. Memory capacity of 10 students was tested before and after training state whether the training was effective or not from the following scores.

Before training	12	14	11	8	7	10	3	0	5	6
After training	15	16	10	7	5	12	10	2	3	8

5. In one sample of 10 observations, the sum of the squares of the deviation o the sample values from the sample mean was 120 and in the other sample of 12 observations it was 314. Test whether the difference is significant at 5% level.
6. Two samples are drawn from the normal populations are

Sample 1	20	16	26	27	23	22	18	24	25	19		
Sample 2	27	33	42	35	22	34	38	28	41	43	39	37

7. A die is thrown 120 times and the frequencies of various face are as follows

Face No	1	2	3	4	5	6
Frequency	10	15	25	25	18	27

Test whether the die was fair
8. In a set of random numbers, the digits 0,1,2....9, were found to have the following frequencies.

Digits	0	1	2	3	4	5	6	7	8	9
f	43	32	38	27	38	52	36	31	39	24

Test whether they are significantly different from those expected on the hypothesis of uniform distribution.
9. 1000 students at college level were graded according to their I.Q and the economic conditions of their homes use

^{2}-test to find out whether there is any association between economic conditions at home and I.Q.

Economic conditions	High	Low	Total
Rich	460	140	600
Poor	240	160	400
Total	**700**	**300**	**1000**

10. A machine puts out 16 in perfect articles in a sample 500 after machine is over hauled; it pulls out 3 imperfect articles in a batch of 100. Has the machine improved?
11. A machine produced 20 defective articles in a batch of 400. After over hauling it produced 10 defectives in a batch of 300. Has the machine improved?
12. In a random sample of 400 persons from a large 120 are females can it said that males and females are in the ratio 5:3 in the population use 1% level of significance.
13. 4 coin is tossed 900 times and head appears 490 times. Does this result support the hypothesis that the coin is unbiased.
14. A wholesale apples claim 4% of the apples supplied by him are defective a random sample of 600 apples contained 36 defective apples. Test the claim of the wholesaler.

F - Distribution ($\alpha = 0.05$ in the Right Tail)

df_2 \ df_1	1	2	3	4	5	6	7	8	9
	Numerator Degrees of Freedom								
Denominator Degrees of Freedom									
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123
4	7.7086	9.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	6.9988
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799

F - Distribution ($\alpha = 0.05$ in the Right Tail)

Denominator Degrees of Freedom df_2 \ df_1 (Numerator Degrees of Freedom)	10	12	15	20	24	30	40	60	120	∞
1	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	19.396	19.413	19.429	19.446	19.454	19.462	19.471	19.479	19.487	19.496
3	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720	8.5494	8.5264
4	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6877	5.6581	5.6281
5	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.3650
6	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6689
7	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4045
12	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2962
13	2.6710	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	2.6022	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	2.5437	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9604
18	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	2.3779	2.3080	2.2341	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.8780
20	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	2.3210	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165	1.8657	1.8117
22	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.9380	1.8894	1.8380	1.7831
23	2.2747	2.2036	2.1282	2.0476	2.0050	1.9605	1.9139	1.8648	1.8128	1.7570
24	2.2547	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424	1.7896	1.7330
25	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.7110
26	2.2197	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027	1.7488	1.6906
27	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6717
28	2.1900	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6376
30	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	1.9926	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290	1.3519	1.2539
∞	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.3940	1.3180	1.2214	1.0000

F - Distribution ($\alpha = 0.01$ in the Right Tail)

Numerator Degrees of Freedom (df_1); Denominator Degrees of Freedom (df_2)

df_2 \ df_1	1	2	3	4	5	6	7	8	9
1	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158
6	13.745	10.925	9.7795	9.1483	8.7459	8.4661	8.2600	8.1017	7.9761
7	12.246	9.5466	8.4513	7.8466	7.4604	7.1914	6.9928	6.8400	6.7188
8	11.259	8.6491	7.5910	7.0061	6.6318	6.3707	6.1776	6.0289	5.9106
9	10.561	8.0215	6.9919	6.4221	6.0569	5.8018	5.6129	5.4671	5.3511
10	10.044	7.5594	6.5523	5.9943	5.6363	5.3858	5.2001	5.0567	4.9424
11	9.6460	7.2057	6.2167	5.6683	5.3160	5.0692	4.8861	4.7445	4.6315
12	9.3302	6.9266	5.9525	5.4120	5.0643	4.8206	4.6395	4.4994	4.3875
13	9.0738	6.7010	5.7394	5.2053	4.8616	4.6204	4.4410	4.3021	4.1911
14	8.8616	6.5149	5.5639	5.0354	4.6950	4.4558	4.2779	4.1399	4.0297
15	8.6831	6.3589	5.4170	4.8932	4.5556	4.3183	4.1415	4.0045	3.8948
16	8.5310	6.2262	5.2922	4.7726	4.4374	4.2016	4.0259	3.8896	3.7804
17	8.3997	6.1121	5.1850	4.6690	4.3359	4.1015	3.9267	3.7910	3.6822
18	8.2854	6.0129	5.0919	4.5790	4.2479	4.0146	3.8406	3.7054	3.5971
19	8.1849	5.9259	5.0103	4.5003	4.1708	3.9386	3.7653	3.6305	3.5225
20	8.0960	5.8489	4.9382	4.4307	4.1027	3.8714	3.6987	3.5644	3.4567
21	8.0166	5.7804	4.8740	4.3688	4.0421	3.8117	3.6396	3.5056	3.3981
22	7.9454	5.7190	4.8166	4.3134	3.9880	3.7583	3.5867	3.4530	3.3458
23	7.8811	5.6637	4.7649	4.2636	3.9392	3.7102	3.5390	3.4057	3.2986
24	7.8229	5.6136	4.7181	4.2184	3.8951	3.6667	3.4959	3.3629	3.2560
25	7.7698	5.5680	4.6755	4.1774	3.8550	3.6272	3.4568	3.3239	3.2172
26	7.7213	5.5263	4.6366	4.1400	3.8183	3.5911	3.4210	3.2884	3.1818
27	7.6767	5.4881	4.6009	4.1056	3.7848	3.5580	3.3882	3.2558	3.1494
28	7.6356	5.4529	4.5681	4.0740	3.7539	3.5276	3.3581	3.2259	3.1195
29	7.5977	5.4204	4.5378	4.0449	3.7254	3.4995	3.3303	3.1982	3.0920
30	7.5625	5.3903	4.5097	4.0179	3.6990	3.4735	3.3045	3.1726	3.0665
40	7.3141	5.1785	4.3126	3.8283	3.5138	3.2910	3.1238	2.9930	2.8876
60	7.0771	4.9774	4.1259	3.6490	3.3389	3.1187	2.9530	2.8233	2.7185
120	6.8509	4.7865	3.9491	3.4795	3.1735	2.9559	2.7918	2.6629	2.5586
∞	6.6349	4.6052	3.7816	3.3192	3.0173	2.8020	2.6393	2.5113	2.4073

F - Distribution ($\alpha = 0.01$ in the Right Tail)

$df_2 \backslash df_1$	10	12	15	20	24	30	40	60	120	∞
1	6055.8	6106.3	6157.3	6208.7	6234.6	6260.6	6286.8	6313.0	6339.4	6365.9
2	99.399	99.416	99.433	99.449	99.458	99.466	99.474	99.482	99.491	99.499
3	27.229	27.052	26.872	26.690	26.598	26.505	26.411	26.316	26.221	26.125
4	14.546	14.374	14.198	14.020	13.929	13.838	13.745	13.652	13.558	13.463
5	10.051	9.8883	9.7222	9.5526	9.4665	9.3793	9.2912	9.2020	9.1118	9.0204
6	7.8741	7.7183	7.5590	7.3958	7.3127	7.2285	7.1432	7.0567	6.9690	6.8800
7	6.6201	6.4691	6.3143	6.1554	6.0743	5.9920	5.9084	5.8236	5.7373	5.6495
8	5.8143	5.6667	5.5151	5.3591	5.2793	5.1981	5.1156	5.0316	4.9461	4.8588
9	5.2565	5.1114	4.9621	4.8080	4.7290	4.6486	4.5666	4.4831	4.3978	4.3105
10	4.8491	4.7059	4.5581	4.4054	4.3269	4.2469	4.1653	4.0819	3.9965	3.9090
11	4.5393	4.3974	4.2509	4.0990	4.0209	3.9411	3.8596	3.7761	3.6904	3.6024
12	4.2961	4.1553	4.0096	3.8584	3.7805	3.7008	3.6192	3.5355	3.4494	3.3608
13	4.1003	3.9603	3.8154	3.6646	3.5868	3.5070	3.4253	3.3413	3.2548	3.1654
14	3.9394	3.8001	3.6557	3.5052	3.4274	3.3476	3.2656	3.1813	3.0942	3.0040
15	3.8049	3.6662	3.5222	3.3719	3.2940	3.2141	3.1319	3.0471	2.9595	2.8684
16	3.6909	3.5527	3.4089	3.2587	3.1808	3.1007	3.0182	2.9330	2.8447	2.7528
17	3.5931	3.4552	3.3117	3.1615	3.0835	3.0032	2.9205	2.8348	2.7459	2.6530
18	3.5082	3.3706	3.2273	3.0771	2.9990	2.9185	2.8354	2.7493	2.6597	2.5660
19	3.4338	3.2965	3.1533	3.0031	2.9249	2.8442	2.7608	2.6742	2.5839	2.4893
20	3.3682	3.2311	3.0880	2.9377	2.8594	2.7785	2.6947	2.6077	2.5168	2.4212
21	3.3098	3.1730	3.0300	2.8796	2.8010	2.7200	2.6359	2.5484	2.4568	2.3603
22	3.2576	3.1209	2.9779	2.8274	2.7488	2.6675	2.5831	2.4951	2.4029	2.3055
23	3.2106	3.0740	2.9311	2.7805	2.7017	2.6202	2.5355	2.4471	2.3542	2.2558
24	3.1681	3.0316	2.8887	2.7380	2.6591	2.5773	2.4923	2.4035	2.3100	2.2107
25	3.1294	2.9931	2.8502	2.6993	2.6203	2.5383	2.4530	2.3637	2.2696	2.1694
26	3.0941	2.9578	2.8150	2.6640	2.5848	2.5026	2.4170	2.3273	2.2325	2.1315
27	3.0618	2.9256	2.7827	2.6316	2.5522	2.4699	2.3840	2.2938	2.1985	2.0965
28	3.0320	2.8959	2.7530	2.6017	2.5223	2.4397	2.3535	2.2629	2.1670	2.0642
29	3.0045	2.8685	2.7256	2.5742	2.4946	2.4118	2.3253	2.2344	2.1379	2.0342
30	2.9791	2.8431	2.7002	2.5487	2.4689	2.3860	2.2992	2.2079	2.1108	2.0062
40	2.8005	2.6648	2.5216	2.3689	2.2880	2.2034	2.1142	2.0194	1.9172	1.8047
60	2.6318	2.4961	2.3523	2.1978	2.1154	2.0285	1.9360	1.8363	1.7263	1.6006
120	2.4721	2.3363	2.1915	2.0346	1.9500	1.8600	1.7628	1.6557	1.5330	1.3805
∞	2.3209	2.1847	2.0385	1.8783	1.7908	1.6964	1.5923	1.4730	1.3246	1.0000

Column heading: Numerator Degrees of Freedom; row heading: Denominator Degrees of Freedom

Table of Chi-square statistics

df	P = 0.05	P = 0.01	P = 0.001
1	3.84	6.64	10.83
2	5.99	9.21	13.82
3	7.82	11.35	16.27
4	9.49	13.28	18.47
5	11.07	15.09	20.52
6	12.59	16.81	22.46
7	14.07	18.48	24.32
8	15.51	20.09	26.13
9	16.92	21.67	27.88
10	18.31	23.21	29.59
11	19.68	24.73	31.26
12	21.03	26.22	32.91
13	22.36	27.69	34.53
14	23.69	29.14	36.12
15	25.00	30.58	37.70
16	26.30	32.00	39.25
17	27.59	33.41	40.79
18	28.87	34.81	42.31
19	30.14	36.19	43.82
20	31.41	37.57	45.32
21	32.67	38.93	46.80
22	33.92	40.29	48.27
23	35.17	41.64	49.73
24	36.42	42.98	51.18
25	37.65	44.31	52.62
26	38.89	45.64	54.05
27	40.11	46.96	55.48
28	41.34	48.28	56.89
29	42.56	49.59	58.30
30	43.77	50.89	59.70

PERCENTAGE POINTS OF THE T DISTRIBUTION

Tail Probabilities								
One Tail		0.10	0.05	0.025	0.01	0.005	0.001	0.0005
Two Tails		0.20	0.10	0.05	0.02	0.01	0.002	0.001
D	1	3.078	6.314	12.71	31.82	63.66	318.3	637
E	2	1.886	2.920	4.303	6.965	9.925	22.330	31.6
G	3	1.638	2.353	3.182	4.541	5.841	10.210	12.92
R	4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
E	5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
E	6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
S	7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
	8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
O	9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
F	10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
	11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
F	12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
R	13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
E	14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
E	15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
D	16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
O	17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
M	18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
	19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
	20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
	21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
	22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
	23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
	24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
	25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
	26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
	27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
	28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
	29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
	30	1.310	1.697	2.042	2.457	2.750	3.385	3.646

MADURAI KAMARAJ UNIVERSITY

(University with Potential for Excellence)

DIRECTORATE OF DISTANCE EDUCATION

Palkalai Nagar, Madurai - 625 021, India

Ph : 0452-2458471 (30 Lines)   Fax : 0452-2458265

E-mail	: mkudde@mkudde.org
General grievances	: mkuddegrievance@gmail.com
UG Courses	: mkuddeug@gmail.com
PG Course	: mkuddepg@gmail.com
MBA Course	: mkuddembag@gmail.com
MCA Course	: mkuddemcag@gmail.com
Education Courses	: mkuddeedu@gmail.com
Website	: www.mkudde.org
IVRS	: 0452 - 2459990
	0452 - 2459596

Student Support Service: 0452 - 2458419

DDE - Examinations

Fax No. : 0452 - 2458261

E-mail : mkuace@yahoo.com

Examn., Grievance Redress Cell: 0452-2458471-Extn.515