M.A. ECONOMICS - I YEAR

DKN1A : QUANTITATIVE METHODS

SYLLABUS

Unit - I : Differentiation and its Application
Rules of differentiation – Basic rules of derivative – quotient rule- chain rule – logarithmic and exponential rule – conditions for maxima and minima of a function-point of inflexion-simple application of derivatives-calculation of marginal functions from total functions – maximization of profit and revenue-minimization of cost - elasticity of demand – relationship between average cost and marginal cost using derivatives – relation between average revenue, marginal revenue and price elasticity of demand – partial derivatives and its applications – optimization of functions with two independent variables – applications.

Unit - II : Integration and its Application
Concept of integration – definite and indefinite integration – rules of integration – application of integrations to the calculation of total function from marginal function – application of definite integration to calculate area between two curves; consumer's surplus and producer's surplus.

Unit - III : Correlation and Regression Analysis
Meaning, assumptions and limitations of simple correlation; Pearsons correlation coefficient and Spearmen's rank correlation coefficients and their properties; probable error; Regression – methods of estimation of linear equation using ordinary least square method – standard error of regression coefficient – Partial and Multiple correlation and regression (applications only); Methods of estimation of non-linear equation – parabolic and exponential.

Unit - IV : Probability and Theoretical Distributions
Classical and empirical definition of probability; addition and multiplication theorems of probability; conditional probability; Elementary concept of random variable; probability mass and density functions; Theoretical distributions – properties (without derivatives) of Binomial, Poisson, Normal and log normal distributions; Fitting Binomial, Poisson and Normal distribution – Central Limit Theorem.

Unit - V : Sampling Techniques and Hypothesis Testing
Basic concepts and laws of sampling – different types of random and non-random sampling; concept of estimator; sampling distribution of mean and proportion-standard error and its uses in test of hypothesis. Procedure for testing a hypothesis – Null and Alternative hypothesis – confidence interval and level of significance – Type I and Type II errors – Hypothesis testing based on Z", "t", "X^2" (Chi-square) and "F" tests - Techniques of analysis of variance – Non-parametric tests; advantages and disadvantages – sign test; median test – Fisher's Transformation Tests.

References :

1. R.G.D. Allen, Mathematical Analysis for Economists.
2. A.C.Chiang – Fundamental Methods of Mathematical Economics.
3. Metha and Medhani – Mathematics for Economists.
4. Edward T. Dowling – Mathematical Methods for Business and Economics.
5. Verma IP – Quantitative Techniques..
6. *SC. Gupta* and V.K. Kapoor, Fundamentals of Mathematics Statistics.
7. S.P. Gupta – Statistical Methods.
8. M. Des Raj, Sampling Theory.
9. G.W. Cochran, Sampling Techniques.

QUANTITATIVE METHODS

UNIT - I

DIFFERENTIATION AND ITS APPLICATION

Meaning of Differentiation

Differentiation is the process of finding the rate at which a variable quantity is changing. To express the rate of change in any function, we have the concept of the 'Derivative' which involves small change in the dependent variables with reference to a small change in independent variables. The problem is to find a function derived from the given relationship between the two variables so as to express the idea of change. This derived function is called the "Derivative" of a given function. The process of obtaining the derivative is called "Differentiation".

Rules of Differentiation

Rule No. I
Polynomial Function Rule

a) If $y = x^n$, then $dy/dx = nx^{n-1}$

b) If $y = x$, then $dy/dx = x^{1-1} = x^0 = 1$

Examples

1) If $y = x^{10}$, $dy/dx = 10x^{10-1} = 10x^9$

2) If $y = x^{-7}$, $dy/dx = -7x^{-7-1} = -7x^{-8} = -7/x^8$

3) If $y = x^{3/2}$, $dy/dx = 3/2\, x^{\frac{3}{2}-1} = 3/2x^{1/2}$

4) If $y = \sqrt{x}$, $y = x^{1/2}$, $dy/dx = 1/2x^{\frac{1}{2}-1} = 1/2x^{-1/2}$

Rule No. II
Constant Function Rule

a) Derivative of a constant

If y = c, where c is constant,

Then $dy/dx = d/dx\,(c) = 0$

Examples

1) If y = 7, $dy/dx = 0$

2) If y = 115, $dy/dx = 0$

b) Derivative of the product of a constant and a function

If $y = ax^n$, where 'a' is constant and $a \neq 0$,

then $dy/dx = d/dx\,(ax^n)$

$= a\, d/dx\,(x^n)$

$dy/dx = a\, nx^{n-1}$

Examples

1) If $y = 10x^{12}$, find dy/dx

$dy/dx = 10\, d/dx\,(x^{12})$

$= 10\,(12x^{12-1})$

$dy/dx = 120\, x^{11}$

2) If $y = 9x^{-4}$, find dy/dx

$dy/dx = 9\, d/dx\,(x^{-4})$

$= 9\,(-4x^{-4-1})$

$dy/dx = -36\, x^{-5}$

3) If $y = -11x^{-9}$, find dy/dx

$dy/dx = -11\, d/dx\, (x^{-9})$

$= -11\, (-9x^{-9-1})$

$dy/dx = 99x^{-10}$

Rule No. III

Linear Function Rule

If $y = mx + c$, where 'm' and 'c' are constants, then

$dy/dx = d/dx\, (mx + c)$

$= d/dx\, (mx) + d/dx\, (c)$

$= m.1 + 0$

$dy/dx = m$

Examples

1) If $y = 3x + 6$, find dy/dx

$dy/dx = d/dx\, (3x + 6)$

$dy/dx = 3$

2) If $y = 9x + 2$, find dy/dx

$dy/dx = d/dx\, (9x + 2)$

$dy/dx = 9$

Rule No. IV

Addition Rule

If y = u + v, where u and v are the differentiable functions of *x*, then

$dy/dx = du/dx + dv/dx$

Examples

1) If $y = x^3 + x^8$, find dy/dx

$dy/dx = d/dx\, (x^3 + x^8)$

$= d/dx\, (x^3) + d/dx\, (x^8)$

$dy/dx = 3x^2 + 8x^7$

2) If $y = x^4 + x^9 + x^{11}$, find dy/dx

$dy/dx = d/dx\, (x^4 + x^9 + x^{11})$

$= d/dx\, (x^4) + d/dx\, (x^9) + d/dx\, (x^{11})$

$dy/dx = 4x^3 + 9x^8 + 11x^{10}$

3) If $y = 11x^{-3} + 4x^{-9} + 3x + 7$, find dy/dx

$dy/dx = d/dx\, (11x^{-3} + 4x^{-9} + 3x + 7)$

$= d/dx\, (11x^{-3})\ d/dx\, (4x^{-9}) + d/dx\, (3x) + d/dx\, (7)$

$dy/dx = -33x^{-4} - 36x^{-10} + 3$

Rule No. V

Subtraction Rule

If y = u - v, where u and v are the differentiable functions of *x*, then

$dy/dx = du/dx - dv/dx$

Examples

1) If y = x^3 - x^8, find dy/dx

$dy/dx = d/dx\,(x^3) - d/dx\,(x^8)$

$dy/dx = 3x^2 - 8x^7$

2) If y = x^4 - x^9 - x^{11}, find dy/dx

$dy/dx = d/dx\,(x^4) - d/dx\,(x^9) - d/dx\,(x^{11})$

$dy/dx = 4x^3 - 9x^8 - 11x^{10}$

3) If y = $3x^{-9} - 7x^{-7} - 8x - 8$, find dy/dx

$dy/dx = d/dx\,(3x^{-9}) - d/dx\,(7x^{-7}) - d/dx\,(8x) - d/dx\,(8)$

$dy/dx = -27x^{-10} + 49x^{-8} - 8$

Rule No. VI
Multiplication Rule

If y = uv, where u and v are the differentiable functions of x, then

$$dy/dx = u\frac{dv}{dx} + v\frac{du}{dx}$$

Examples

1) If y = $(2x^3 + 9)(x^2 + 3x)$, find dy/dx

$$dy/dx = u\frac{dv}{dx} + v\frac{du}{dx}$$

$dy/dx = (2x^3 + 9)\, d/dx\,(x^2 + 3x) + (x^2 + 3x)\, d/dx(2x^3 + 9)$

$= (2x^3 + 9)(2x+3) + (x^2 + 3x)(6x^2)$

$= 4x^4+6x^3+18x+27+6x^4+18x^3$

$dy/dx = 10x^4+24x^3+18x+27$

2) Find dy/dx, if y = $(x^3 + 3)(2x^2 - 3x^3)$

$dy/dx = (x^3 + 3)\, d/dx\,(2x^2 - 3x^3) + (2x^2 - 3x^3)\, d/dx\,(x^3 + 3)$

$= (x^3 + 3)(4x - 9x^2) + (2x^2 - 3x^3)(3x^2)$

$= 4x^4 - 9x^5 + 12x - 27x^2 + 6x^4 - 9x^5$

$dy/dx = -18x^5 + 10x^4 - 27x^2 + 12x$

Rule No. VII
Quotient Rule

If y = $\frac{u}{v}$, where u and v are the differentiable functions of x and v ≠ 0, then

$$\frac{dy}{dx} = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$$

Examples

1) Evaluate dy/dx for y = $\frac{x+1}{x-1}$

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

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

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

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

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

2) If y = $\frac{x^2-1}{x^2+1}$, find dy/dx

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

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

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

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

$$dy/dx = \frac{4x}{(x^2+1)^2}$$

3) Calculate dy/dx for y = $\frac{x^3+2x}{x^2+1}$

$$dy/dx = \frac{(x^2+1)\, d/dx\,(x^3+2x)-(x^3+2x)\, d/dx\,(x^2+1)}{(x^2+1)^2}$$

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

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

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

$$dy/dx = \frac{x^4+x^2+2}{(x^2+1)^2}$$

Rule No. VIII
Chain Rule

If y is the function of u where u is the function of x, then the derivative of y with respect to *x* (i.e., dy/dx) is equal to the product of the derivative of u with respect to u and the derivative of u with respect to x.

If y = F(u), where u = f(x), then

$$\frac{dy}{dx} = \left(\frac{dy}{du}\right)\left(\frac{du}{dx}\right)$$

Examples

1) If z = 7y + 3, where y = 5x^2, find $\frac{dz}{dx}$

$$\frac{dz}{dx} = \left(\frac{dz}{dy}\right)\left(\frac{dy}{dx}\right)$$

$$\frac{dz}{dy} = \frac{d}{dy}(7y+3) = 7$$

$$\frac{dy}{dz} = \frac{d}{dz}(5x^2) = 10x$$

$$\frac{dz}{dx} = (7)(10x) = 70x$$

2) Find dz/dx, if y = x^2 + 3x and z = y^2 + 1

$$\frac{dz}{dx} = \left(\frac{dz}{dy}\right)\left(\frac{dy}{dx}\right)$$

dz/dy=2y, dy/dx=2x+3

dz/dx= 2y(2x+3)= 2(x^2+3x)(2x+3) ……(since y=x^2+3x)

$$\frac{dz}{dx} = (2x^2+6x)(2x+3)$$

$$= 4x^3 + 12x^2 + 6x^2 + 18x$$

$$\frac{dz}{dx} = 4x^3 + 18x^2 + 18x$$

3) Find du/dx by using Chain Rule Technique for the function u = f $(x^2 + 3x)^2$

u = $(x^2 + 3x)$

u = y^2 and y = $(x^2 + 3x)$

$$\frac{du}{dx} = \left(\frac{du}{dy}\right)\left(\frac{dy}{dx}\right)$$

$$\frac{du}{dy} = \frac{d}{dy}(y^2) = 2y$$

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

$$\frac{du}{dx} = (2y)(2x+3) = 2(x^2+3x)(2x+3) \ldots \text{since } y = (x^2+3x)$$

$$= (2x^2 + 6x)(2x + 3)$$

$$= 4x^3 + 6x^2 + 12x^2 + 18x$$

du/dx $= 4x^3+18x^2+18x$

Rule No. IX

Parametric Function Rule

If both x and y are the differentiable function of ‘t’ then the derivative of y with respect to x is obtained by dividing the derivative of y with respect to ‘t’ by the derivative of x with respect to t (t is the parameter)

If x = f(t) and y = g(t) then

$$\frac{dy}{dx} = \left(\frac{dy}{dt}\right)\left(\frac{dt}{dx}\right) = \frac{dy}{dt} \div \frac{dx}{dt}$$

Example

Find dy/dx, if x = at^3 and y = 3at

$$\frac{dy}{dx}=\frac{dy}{dt}\div\frac{dx}{dt}$$

$$\frac{dy}{dt}=\frac{d}{dt}(3at)=3a$$

$$\frac{dx}{dy}=\frac{d}{dt}(at^3)=3at^2$$

$$\frac{dy}{dx}=3a\div 3at^2 \;=\frac{3a}{3at^2}$$

$$\frac{dy}{dx}=\frac{1}{t^2}$$

Rule No X

Implicit Function Rule

An equation y = x^2 + 7x - 8 directly express y in terms of x. Hence, this function is "Explicit function".

An equation x^2y + y +3x = 0 does not directly express y in terms of x. Hence, this function is "Implicit Function".

To find dy/dx for the Implicit Function, every term in both sides of the function is to be differentiated with reference to x.

Examples

Find dy/dx of the following functions

1) 2x - 3y = 6

$$d/dx(2x)-d/dx(3y)=d/dx(6)$$

$$2-3\,dy/dx = 0$$

$$-3\,dy/dx = -2$$

$$dy/dx = \frac{-2}{-3} = \frac{2}{3}$$

2) xy = 6

Following product Rule, we get

$$x.d/dx(y)+y.d/dx(x)=\frac{d}{dx}(6)$$

$$(x)\,dy/dx+y=0$$

$$x\frac{dy}{dx}=-y$$

$$\frac{dy}{dx}=\frac{-y}{x}$$

3) 2x^2 + 3xy = 3

$$\frac{d}{dx}(2x^2)+\frac{d}{dx}(3xy)=\frac{d}{dx}(3)$$

$$4x+3\left[x.\frac{d}{dx}(y)+y.\frac{d}{dx}(x)\right]=0$$

$$4x+3x\frac{dy}{dx}+3y=0$$

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

$$\frac{dy}{dx}=\frac{-(4x+3y)}{3x}$$

Rule No. XI

Inverse Function Rule

The derivative of the Inverse Function (i.e, d*x*/dy) is equal to the reciprocal of the derivative of the original function (i.e., $\frac{1}{dy/dx}$)

Problems

1) Find the derivative of the Inverse function y = x^2

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

$$\frac{dy}{dx}=2x$$

Then $\frac{dx}{dy}=\frac{1}{\frac{dy}{dx}}=\frac{1}{2x}$

2) Find the derivative of the Inverse function y = $x^3 + 3x$

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

Then $\frac{dx}{dy}=\frac{1}{\frac{dy}{dx}}=\frac{1}{3x^2+3}$

Rule No. XII

Derivative of a Derivative

$\frac{dy}{dx}$ is the first order derivative of 'y' with respect to '*x*'. The second order derivative or the derivative of the derivative is obtained by differentiating the first order derivative with respect to '*x*' and is written as

$$\frac{d^2y}{dx^2}=\frac{d}{dx}\left(\frac{dy}{dx}\right)$$

Examples

1) If y = $5x^4 + 2x^3$, find

$\frac{dy}{dx}$, $\frac{d^2y}{dx^2}$ *and* $\frac{d^3y}{dx^3}$

$$\frac{dy}{dx} = 20x^3 + 6x^2 \left[First\ Order\ Derivative\right]$$

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d}{dx}(20x^3 + 6x^2)$$

$= 60x^2 + 12x$ [Second Order Derivative]

Rule No. XIII

Exponential Function Rule

The derivative of an Exponential Function is the exponential itself, if the base of the function is ‘e’, the natural base of an exponential

i) If y = e^x , then dy/dx = d/dx (e^x) = e^x

ii) If y = e^{-x}, then dy/dx = d/dx (e^{-x}) = -e^{-x}

iii) If y = e^{ax}, then dy/dx = ae^{ax}

b) If y = e^u, where u = g(x), then

$$\frac{dy}{dx} = \frac{d}{dx}(e^u)\frac{du}{dx} = e^u\left(\frac{du}{dx}\right)$$

Examples

1) If y = e^{x^3+3}, find dy/dx

Let u=x^3+3

then y = e^u

$$\frac{dy}{dx} = e^u\left(\frac{du}{dx}\right) = e^u\,(3x^2)$$

$= 3x^2e^u = 3x^2\,(e^{x^3+3})$

2) If y = e^{2x^2}, find dy/dx

Let u = $2x^2$, then y = e^u

$$\frac{dy}{dx} = e^u\left(\frac{du}{dx}\right) = e^u\,(4x)$$

$= 4x\,(e^u) = 4x(e^{2x^2})$

Rule No. XIV

Logarithmic Function Rule

a) The derivative of a logarithm with natural base such as

y = log x, then $\frac{dy}{dx} = \frac{d}{dx}(\log x)$

$$\frac{dy}{dx} = 1/x$$

b) If y = log u, where u = g(x), then $\frac{dy}{dx} = \frac{dy}{du}.\frac{du}{dx}$

$$\frac{dy}{dx} = \frac{d}{du}(\log u).\frac{du}{dx}$$

(or) $\frac{dy}{dx} = \frac{1}{u}\left(\frac{du}{dx}\right)$

Example

If y = log x^5, find dy/dx

Let u = x^5

y = log u

$$\frac{dy}{dx} = \frac{1}{u}\left(\frac{du}{dx}\right) = {}^{1}\!/_{u}\,(5x^4) = \frac{5x^4}{u}$$

$$= \frac{5x^4}{x^5} = \frac{5}{x}$$

Maxima and Minima

Maxima

A function is said to have attained its "maximum value" or "maxima", if the function stops to increase and begins to decrease. The conditions for maxima is

i) First order differentiation $\frac{dy}{dx} = 0$

ii) Second order differentiation $\frac{d^2y}{dx^2} < 0$

Minima

A function is said to have attained "Minimum value" or "Minima" if the function stops to decrease and begins to increase. The conditions for Minima are

i) First order differentiation $\frac{dy}{dx} = 0$

ii) Second order differentiation $\frac{d^2y}{dx^2} > 0$

Example

1) Find the maxima or minima of the function y = x^2 - 4x - 5.

Solution

y = x^2 - 4x - 5

dy/dx = 2x - 4

First order condition dy/dx = 0

Therefore 2x - 4 = 0

2x = 4

x = 4/2

x = 2

The function may be maximum or minimum at x = 2

Second condition $\frac{d^2y}{dx^2} > 0$ or $\frac{d^2y}{dx^2} < 0$

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d}{dx}(2x - 4)$$

$$\frac{d^2y}{dx^2} = 2 > 0$$

Therefore this function is minimum at x = 2

2) Find the maxima and minima of the function y = $2x^3 - 6x$

Solution

$y = 2x^3 - 6x$

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

$\frac{dy}{dx} = 0$

$6x^2 - 6 = 0$

$6x^2 = 6$

$x^2 = 6/6 = 1$

$x = \sqrt{1} = \pm 1$

$x = -1$ or $x = +1$ give minimum or maximum value

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left[\frac{dy}{dx}\right] = \frac{d}{dx}\left[6x^2 - 6\right]$$

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

when $x = -1$, $\frac{d^2y}{dx^2} = 12 \times -1 = -12 < 0$

when $x = +1$ $\frac{d^2y}{dx^2} = 12 \times 1 = 12 > 0$

when $x = -1$, the function reaches maxima

when $x = +1$, the function reaches minima.

3) Find the maxima and minima of the following function y = $2x^3 - 3x^2 - 36x + 10$

Solution

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

$dy/dx = 6x^2 - 6x - 36$

$dy/dx = 0$

$6x^2 - 6x - 36 = 0$

$6(x^2 - x - 6) = 0$

$x^2 - x - 6 = 0$

Use substitution method or formula method to find x value

formula : $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

substitution method

$x^2 - x - 6 = 0$

$x^2 - 3x + 2x - 6 = 0$

$x(x - 3) + 2(x - 3) = 0$

$(x - 3)(x + 2) = 0$

$x = 3$ or $x = -2$

$x = 3$ or $x = -2$ give maxima or minima

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d}{dx}(6x^2 - 6x - 36)$$

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

when $x = 3$

$$\frac{d^2y}{dx^2} = (12 \times - 3) - 6 = 36 - 6 = 30$$

$$\frac{d^2y}{dx^2} = 30 > 0$$

when x=2

$$\frac{d^2y}{dx^2} = (12 \times - 2) - 6 = -24 - 6 = -30$$

$$\frac{d^2y}{dx^2} = 30 < 0$$

when $x = 3$ the function reaches minima

when $x = -2$ the function reaches maxima.

Point of Inflexion

Point of Inflexion is a point at which a curve is changing from concave upward to concave downward, or vice versa. Here both first and second order derivatives are zero.

Example

Given the function y = $x^3 - 3x^2 + 7$, find the point of inflexion.

Solution

$$y = x^3 - 3x^2 + 7$$
$$dy/dx = 3x^2 - 6x$$
$$dy/dx = 0$$
$$3x^2 - 6x = 0$$
$$x(3x - 6) = 0, \quad x = 0$$
$$3x - 6 = 0$$
$$3x = 6$$
$$x = 6/3 = 2$$

when $x = 0$, y = $0^3 - 3\,(0)^2 + 7 = 7$

when x = 2, y = $2^3 - 3\,(2)^2 + 7 = 8 - 12 + 7 = 3$

Second condition $\frac{d^2y}{dx^2} = 0$

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = d/_{dx}(3x^2 - 6x)$$

$$\frac{d^2y}{dx^2} = 6x - 6 = 0$$

$$6x = 0$$
$$x = 6/6 = 1 > 0$$

The point of inflexion is $x = 0$ and y = 7 (or) $x = 2$ and y = 3

Simple application of derivatives
Calculation of Marginal function from Total function

Example

1) For the Total Utility Function U = 20 $x^4 + 7x^3 + 13x^2 + 12\ x + 9$, Compute Marginal Utility.

Solution

$$\text{Marginal Utility} = \frac{du}{dx} = \frac{d}{dx}(u)$$

$$= \frac{d}{dx}\ (20x^4 + 7x^3 + 13x^2 + 12x + 9)$$

$$\text{MU} = 80x^3 + 21x^2 + 26x + 12$$

2) For the Total Utility Function U = $3x^3 - 3x + 4$, compute Marginal Utility

Solution

$$\text{Marginal Utility} = \frac{du}{dx} = \frac{d}{dx}(u)$$

$$= d/dx\ (3x^3 - 3x + 4)$$

$$\text{MU} = 9x^2 - 3$$

3) Compute Marginal Production Function for the production function P = $x^3 + 7x^2 - 12x + 46$

Solution

Production Function P = $x^3 + 7x^2 - 12x + 46$

Marginal Production Function = d/dx (p)

$$= d/dx\ (x^3 + 7x^2 - 12x + 46)$$

$$\text{MP} = 3x^2 + 14x - 12$$

4) For the production function Q = 40f + $3f^2 - f^{3/3}$, Calculate Marginal and Average Production Function

Solution

Production Function Q = 40f + $3f^2 - f^{3/3}$

Marginal Production Function (MP)= d/df (Q)

$$= d/df\ (40f + 3f^2 - f^{3/3})$$

$$= 40 + 6f - 3/3f^0$$

$$= 40 + 6f - 1$$

$$\text{MP} = 39 + 6f$$

Average Production Function (AP)= Q/f

$$= \frac{40f+3f^2-f^{3/3}}{f} = \frac{40f+3f^2-f}{f}$$

$$= 40+3f-1$$

$$\text{AP} = 39+3f$$

5) Find out the Marginal cost Function for the Total cost function, C = $x + 7x^2 + 2x^3 - 9x^4$.

Solution

$$C = x + 7x^2 + 2x^3 - 9x^4$$

$$MC = d/dx\ (x + 7x^2 + 2x^3 - 9x^4)$$

MC = $1 + 14x + 6x^2 - 36x^3$

6) Compute Average Cost and Marginal Cost for the Total Cost, C = $8x^3 + 3x^2 - 6x + 3$

Solution

C = $8x^3 + 3x^2 - 6x + 3$

AC = $\frac{c}{x} = \frac{8x^3 + 3x^2 - 6x + 3}{x}$

= $8x^2 + 3x - 6 + 3/x$

MC = $d/dx\,(8x^3 + 3x^2 - 6x + 3)$

MC = $24x^2 + 6x - 6$

7) Given the Total Cost Function, C = $50 - 2Q + 7Q^2 + Q^3$, find the Marginal cost when Q = 5.

Solution

C = $50 - 2Q + 7Q^2 + Q^3$

MC = $d/dQ\,(50 - 2Q + 7Q^2 + Q^3)$

MC = $-2 + 14Q + 3Q^2$

when Q = 5

MC = $-2 + 14(5) + 3(5)^2$

= $-2 + 70 + 75$

MC = 143

8) Given the Revenue Function R = $80Q - 2Q^2 - 15$, find out the Average and Marginal Revenue Functions

Solution

R = $80Q - 2Q^2 - 15$

AR = $\frac{80Q - 2Q^2 - 15}{Q} = 80 - 2Q - \frac{15}{Q}$

MR = $\frac{dR}{dQ} = d/dQ\,(80Q - 2Q^2 - 15)$

MR = $80 - 4Q$

9) Find out the Marginal Revenue for the Demand Function p = $30 - 2x$

Solution

p = $30 - 2x$

The Revenue (R) = px

R = $(30 - 2x^2)\,x = 30x - 2x^2$

MR = $d/dx\,(R) = d/dx\,(30x - 2x^2)$

MR = $30 - 4x$

10) Find R, AR and MR for the Demand Function q = 100 - 2p, where q is the quantity demanded and ρ price.

Solution

q = 100 - 2p

-2p = q - 100

$$P = \frac{q-100}{-2} = \frac{-q}{2} + \frac{100}{2}$$

$$P = \frac{-q}{2} + 50$$

i) R = $P = \left[\frac{-q}{2} + 50\right] q = \frac{-q2}{2} + 50q$

ii) AR = $\frac{R}{q} = \frac{-q^2/2 + 50q}{q} = \frac{-q}{2} + 50$

iii) MR = $\frac{d}{dq}(R) = \frac{d}{dq}\left(-q/2 + 50q\right) = \frac{-2q}{2} + 50$

MR = - q + 50

Minimization of Cost

Example

1) Given the Total Cost Function C = 1/3 Q^3 - 3Q^2 + 9Q, find Q when Average Cost is minimum. Find also the Marginal Cost at the level of Q.

Solution

C = 1/3 Q^3 - 3Q^2 + 9Q

AC = $\frac{C}{Q} = \frac{1/3Q^3 - 3Q^2 + 9Q}{Q} = 1/3Q^2 - 3Q + 9$

when AC is minimum

i) First derivative of AC = 0

ii) Second derivative of AC > 0

d/dQ (AC) = $\frac{d}{dQ}(1/3Q^2 - 3Q + 9)$

= 2/3 Q – 3

When AC is minimum d/dx(AC) = 0

2/3 Q – 3 = 0

2/3 Q = 3

Q = 3 (3/2)

Q = 9/2

when AC is minimum $\frac{d^2}{dQ^2}(AC) > 0$

$\frac{d^2}{dQ^2}(AC) = \frac{d}{dQ}\left(2/3Q - 3\right)$

= 2/3 > 0

MC = $\frac{d}{dQ}(C) = \frac{d}{dQ}\left(1/3Q^3 - 3Q^2 + 9Q\right)$

MC = Q^2 - 6Q + 9

$$MC \text{ at } Q = 9/2 = (9/2)^2 - 6 \times 9/2 + 9 = \frac{81}{4} - \frac{54}{2} + 9$$

$$= \frac{81 - 108 + 36}{4} = \frac{117 - 108}{4} = \frac{9}{4}$$

Maximization of profit and Revenue

Given the following Revenue (R) and Cost (C) functions for a firm $R = 20q - q^2$ and $C = q^2 + 8q + 2$, find the equilibrium level of output, price, total revenue, total cost and profit.

Solution

For profit (Π) maximization, the following two conditions must be satisfied.

i) $\frac{d\Pi}{dq} = 0$ ii) $\frac{d^2\Pi}{dq^2} < 0$

$$\text{Profit } (\pi) = \text{Revenue (R)} - \text{Cost (C)}$$
$$= (20q - q^2) - (q^2 + 8q + 2)$$
$$= 20q - q^2 - q^2 - 8q - 2$$
$$= -2q^2 + 12q - 2$$
$$d\pi/dq = d/dq(-2q^2 + 12q - 2)$$
$$= -4q + 12$$

since $d\pi/dq = 0$

$$-4q + 12 = 0$$
$$-4q = -12$$
$$q = -12/-4 = 3$$

$$\frac{d^2q}{dq^2} = \frac{d}{dq}\left(\frac{d\Pi}{dq}\right) = \frac{d}{dq}(-4q + 12) = -4$$

$d^2\pi/d^2q = -4 < 0$, Hence profit is maximum.

Price

Putting q = 3 in the AR function, we have

$$AR = \frac{R}{q} = \frac{20q - q^2}{q} = \frac{20q}{q} - \frac{q^2}{q} = 20 - q$$

$p = 20 - q = 20 - 3 = 17$

Revenue

Putting q = 3 in Revenue function, we have $R = 20q - q^2 = 20(3) - (3)^2 = 60 - 9 = 51$,

Cost

Putting q = 3 in the cost function, we have, $C = q^2 + 8q + 2 = 3^2 + 8(3) + 2 = 9 + 24 + 2 = 35$

Profit

putting q = 3 in the profit function, we have,

$$\pi = 12q - 2q^2 - 2$$
$$= 12(3) - 2(3)^2 - 2 = 36 - 18 - 2$$
$$= 36 - 20 = 16$$

Output (q) = 3
price (p) = Rs.17
Revenue (R) = Rs.51
Cost (C) = Rs.35
Profit Π = Rs.16

2) State the conditions for firm's equilibrium and derive level of Output, Price, Total Revenue, Total Cost and profit for $R = 12x - 4x^2$ and $AC = 8 - x$

Solution

For profit maximization, the following two conditions must be satisfied.

i) $d\pi/dx = 0$ ii) $d^2\pi/dx^2 < 0$

R = $12x - 4x^2$

AC = $8 - x$

C = $AC \times x = (8 - x)\,x = 8x - x^2$

π = R - C = $(12x - 4x^2) - (8x - x^2)$

= $12x - 4x^2 - 8x + x^2$

= $4x - 3x^2$

$d\pi/dx = 4 - 6x$

since $d\pi/dx = 0$

$4 - 6x = 0$

$-6x = -4$

$x = 4/6 = 2/3$

$d^2\pi/dx^2$ $d/dx\,(d\pi/dx) = d/dx\,(4-6x)$

$d^2\pi/dx^2 = -6 < 0$

Profit is maximum.

Price

Putting $x = 2/3$ in the AR function, we have,

R = $12x - 4x^2$

$$AR = \frac{12x - 4x^2}{x} = \frac{12x}{x} - \frac{4x^2}{x} = 12 - 4x$$

P = $12 - 4x$ = $12 - 4\,(2/3) = 12 - 8/3$

$$= \frac{36-8}{3} = \frac{28}{3} = 9.33$$

Revenue

Putting x = 2/3 in the profit function, we have

$\pi = 4x - 3x^2$

$$= 4\,(2/3) - 3(2/3)^2 = \frac{8}{3}\ \frac{-12}{9} = \frac{8-4}{3}$$

= 4/3 = 1.3

Output (x) = 2/3

Price = Rs.9.33

Revenue = Rs.6.22

Cost = Rs.4.90

Profit = Rs.1.30

Elasticity of Demand

Elasticity of Demand can be calculated if the demand function is given. The formula used to find elasticity of demand is

$$\text{Elasticity of Demand } \eta = \frac{-dQ}{dp} \times \frac{p}{Q}$$

Example

1) Find the elasticity of demand at p = 2, if the demand function $q = 30 - 5p - p^2$

Solution

Demand function $q = 30 - 5p - p^2$

Substitute the value of p in the demand function, we can get the value of q.

Therefore, $q = 30 - 5(2) - (2)^2 = 30 - 10 - 4$

$= 30 - 14 = 16$

$q = 30 - 5p - p^2$

$dq/dp = -5 - 2p$

Substitute the value of p

$dq/dp = -5 - 2(2) = -5 - 4 = -9$

$$\eta = -(-9) \times \frac{2}{16} = 9 \times \frac{2}{16} = \frac{18}{16} = 1.125$$

2) Find the elasticity of demand, if the demand function p = 100 - Q

Solution

$p = 100 - Q$

$100 - Q = p$

$-Q = p - 100$

$Q = -p + 100$

$$\eta = -\frac{dQ}{dp} \times \frac{p}{Q}$$

$$\frac{dQ}{dp} = d/dp(-p + 100) = -1$$

$$\eta = -(-1)\frac{p}{Q} = p/Q$$

$$\eta = \frac{100 - Q}{Q} = \frac{100}{Q} - \frac{Q}{Q} = \frac{100}{Q} - 1$$

Relation between average revenue, marginal revenue and price elasticity of demand

Elasticity of Demand can be easily determined if Average Revenue and Marginal Revenue are given. Likewise, Marginal Revenue and Elasticity of Demand are known to us. Average Revenue can also be found if Marginal Revenue and Elasticity of demand are given. The formulas are given below.

1) Formula to find η, when AR and MR are given.

$$\eta = \frac{AR}{AR - MR}$$

2) Formula to determine MR, when AR and η are given

$$MR = AR\left[\frac{\eta - 1}{\eta}\right]$$

Example

1) If MR is Rs.50 and the price elasticity of demand is 2, find the AR

Solution

$$AR = MR\left[\frac{\eta}{\eta - 1}\right] = 50\left[\frac{2}{2 - 1}\right]$$

$$= 50\left[\frac{2}{1}\right] = 100$$

AR = Rs100.

2) If AR is Rs.30 and the price elasticity is 4, find MR

Solution

$$MR = AR\left[\frac{\eta-1}{\eta}\right] = 30\left[\frac{4-1}{4}\right]$$

$$= 30\left(\frac{3}{4}\right) = \frac{90}{4} = 22.50$$

MR = Rs.22.50

3) Find elasticity of demand if AR is Rs.20 and MR is Rs.10.

Solution

$$\eta = \frac{AR}{AR-MR} = \frac{20}{20-10} = \frac{20}{10} = 2$$

Partial derivatives

The change in 'u' with respect to *x*, treating 'y' as a constant is called the "partial Derivative" of 'u' with respect to x and is denoted by ∂u/∂*x* . Similarly, the change in 'u' with respect to 'y' treating x as a constant is called the "partial Derivatives" of 'u' with respect to y and is denoted by ∂u/∂y.

Example

Find $\frac{\partial u}{\partial x}$ *and* $\frac{\partial u}{\partial y}$ for the following functions.

1) u = *x*y

$$\frac{\partial u}{\partial x} = y \text{ ….. y is constant}$$

$$\frac{\partial u}{\partial y} = x \text{ ….. } x \text{ is constant}$$

2) u = *x* + y

$$\frac{\partial u}{\partial x} = 1$$

$$\frac{\partial u}{\partial y} = 1$$

3) u = $4xy^2$

$$\frac{\partial u}{\partial x} = 4y^2$$

$$\frac{\partial u}{\partial x} = 8xy$$

5) If u = *x*y + yz + z*x*, find $\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}$ *and* $\frac{\partial u}{\partial z}$

$$\frac{\partial u}{\partial x} = y + z \qquad (y \text{ and } z \text{ are constants})$$

$$\frac{\partial u}{\partial y} = x + z \qquad (x \text{ and } z \text{ are constants})$$

$$\frac{\partial u}{\partial z} = y + x \qquad (x \, and \;\; y \, are \;\; constants)$$

Partial Differentiations of second order or Higher Order

We can also find the Partial Derivatives of second order or higher order.

a) The Direct Partial Differentiation

i) $\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right)$

ii) $\frac{\partial^2 u}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial y}\right)$

b) The Cross Partial Differentiations

i) $\frac{\partial^2 u}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial y}\right)$

ii) $\frac{\partial^2 u}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x}\right)$

$$\frac{\partial^2 u}{\partial x \partial y} = \frac{\partial^2 u}{\partial y \partial x}$$

Example

1) For u = $x^3 + y^2$, find all the partial derivatives.

Solution

i) $\frac{\partial u}{\partial x} = 3x^2$ ii) $\frac{\partial u}{\partial y} = 2y$

iii) $\frac{\partial^2 u}{\partial x^2} = 6x$ iv) $\frac{\partial^2 u}{\partial y^2} = 2$

v) $\frac{\partial^2 u}{\partial x \partial y} = 0$ vi) $\frac{\partial^2 u}{\partial y^2} = 0$

$$\frac{\partial^2 u}{\partial x \partial y} = \frac{\partial^2 u}{\partial y \partial x} = 0$$

2) Find all the partial derivatives of second order of the function u = $x^3 + 3x^2y + y^3$

Solution

i) $\frac{\partial u}{\partial x} = 3x^2 + 6xy$ ii) $\frac{\partial u}{\partial y} = 3x^2 + 3y^2$

iii) $\frac{\partial^2 u}{\partial x^2} = 6x + 6y$ iv) $\frac{\partial^2 y}{\partial y^2} = 6y$

v) $\frac{\partial^2 u}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial y}\right) = 6x$ vi) vi) $\frac{\partial^2 u}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x}\right) = 6x$

$$\frac{\partial^2 u}{\partial x \partial y} = \frac{\partial^2 u}{\partial y \partial x}$$

Total Differentials

We have seen in Partial Differentiation, $\partial u/\partial x$ denotes the change in 'u' when there is a small change in x, treating y as a constant. Then we write as $\frac{\partial u}{\partial x}.dx$

Similarly, $\partial u/\partial y$ denotes the change in 'u', when there is a small change in y, treating x as a constant. Then we, can write this as $\frac{\partial u}{\partial y}.dy$

Therefore the total change in u due to the small change both in x and y will be

$$du = \frac{\partial u}{\partial x}.dx + \frac{\partial u}{\partial y}.\, dy$$

Example

1) Find the total differential of $u = 4x^2 + 3y^2$

Solution

$$du = \frac{\partial u}{\partial x}.dx + \frac{\partial u}{\partial y}.dy$$

$$\frac{\partial u}{\partial x} = 8x, \quad \frac{\partial u}{\partial y} = 6y$$

$du = 8x.dx + 6y.dy$

2) Find the total differential of $u = 3x^2 + xy - 2y^3$

Solution

$$du = \frac{\partial u}{\partial x} dx + \frac{\partial u}{\partial y} dy$$

$$\frac{\partial u}{\partial x} = 6x + y, \qquad \frac{\partial u}{\partial y} = x - 6y^2$$

$du = (6x + y)dx + (x - 6y^2)\, dy$

2) Compute Marginal Utilities of x and y for the Utility Function $U = 5xy - y^2$

Solution

a) Marginal Utility of x = $\frac{\partial u}{\partial x} = 5y$

b) Marginal Utility of y = $\frac{\partial u}{\partial y} = 5x - 2y$

4) Compute Marginal Utilities of x and y at $x = 1$ and $y = 3$ for the Total Utility Function $U=2xy^2 + 3x^2y - 2x + 7y$

Solution

Marginal Utility of x = $\partial u/\partial x = 2y^2+6xy-2$

Marginal Utility of x at x = 1 and y = 3

$= 2(3)^2 + 9\,(1)^2\,(3) - 2$

$= 18 + 18 - 2 = 36-2 = 32$

Marginal Utility of y = $\partial u/\partial y = 4xy + 3x^2 + 7$

Marginal Utility of y at x = 1 and y = 3

$= 4\,(1)\,(3) + 3\,(1)^2 + 7$

$= 12 + 3 + 7$

$= 22$

UNIT - II

INTEGRATION AND ITS APPLICATION

Concept of Integration

The reverse or inverse process of "Differentitation" is called "Indefinite Integration" or "Anti Differentiation". The process of Integration is finding the function whose Derivatives or Differential is given.

Indefinite Integral

If d/dx f(x) = f(x), then the usual notation for the indefinite integral is, $\int f(x)\,dx \text{ or} \int y\,dx = F(x)$.

where

i) The symbol ∫ is called Integral sign.

ii) f(x) is called the "Integral" or the function to be integrated.

iii) dx suggests that the operation of integration is to be with respect to the variable x. The statement, "Evaluate $\int f(x)\,dx$ means "Find the antiderivatives or integral of f(x)".

Rules of Integration

Rule I

$\int dx = x + c$

Rule II

If a function is multiplied by a constant, the integral of that function is also multiplied by that same constant.

$\int k\,dx = k\int dx = kx + c$, where k is constant .

Example

1) $\int 5\,dx = 5\int dx$

$= 5x + c$

2) $\int 7\,dx = 7\int dx$

$= 7x + c$

Rule III

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + c,\ n \neq -1$$

Example

1) $$\int x^5\,dx = \frac{x^{5+1}}{5+1} + c$$

$$= \frac{x^6}{6} + c$$

2) $$\int \frac{1}{x^7}\,dx = \int x^{-7}dx$$

$$= \frac{x^{-7+1}}{-7+1} + c$$

$$= \frac{x^{-6}}{-6} + c$$

$$= \quad -\frac{1}{6x^6} + c$$

Rule IV

Integral of sum or Difference

Integral of sum or difference of a number of functions is equal to the sum or difference of the Integrals of the separate functions.

Examples

1) $\int (x^3 - x + 1)\, dx \quad = \quad \int x^3\, dx - \int x\, dx + 1 \int dx$

$$= \frac{x^4}{4} + \frac{x^2}{2} + x + c$$

2) $\int (x^5 + 1/x^7 - x - 1)\, dx \quad = \quad \int x^5\, dx + \int x^{-7} - \int x\, dx - 1\int dx$

$$= \frac{x^6}{6} + \frac{x^{-6}}{-6} - \frac{x^2}{2} - x + c$$

3) $\int (8x^3 - 3x^2 + x - 1)\, dx \quad = \quad \int 8x^3\, dx - \int 3x^2\, dx + \int x\, dx - \int 1dx$

$$= 8 \int x^3\, dx - 3 \int x^2\, dx + \int x\, dx - 1 \int dx$$

$$= 8\frac{x^4}{4} - 3\frac{x^3}{3} + \frac{x^2}{2} - x + c$$

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

Rule V

Integral of a multiple by a constant

If a function is multiplied by a constant number, this number will remain a multiple of the integral of the function.

Example

1) $\int 4x^8\, dx \quad = \quad 4\frac{x^{8+1}}{8+1} + c$

$$= 4\frac{x^9}{9} + c$$

2) $\int 7\, x^6\, dx \quad = \quad 7 \int x^6\, dx$

$$= 7\frac{x^{6+1}}{6+1} + c$$

$$= 7\frac{x^7}{7} + c$$

$$= x^7 + c$$

Rule VI

Integration by substitution

In the case of a product or quotient of two diffentiable function of x, it may be possible to express them as a constant multiple of another function f(u) and its derivative is du/dx.

a. Product

Example

1) Evaluate $\int 4x^2 (x^3 + 5)^3 \, dx$

Solution

Let u = $x^3 + 5$

du/dx = $3x^2$

$3x^2$ dx = du

dx = du/$3x^2$

On substitution we have,

$$\int 4x^2 (x^3 + 5)^3 \, dx = \int 4x^2.u^3 \, dx \;\ldots\; (\text{Since } U = x^3 + 5)$$

$$= \int 4x^2.u^3.\frac{du}{3x^2}$$

$$= \int \frac{4}{3}u^3 \, du$$

$$= \frac{4}{3}\int u^3 \, du$$

$$= \frac{4}{3}\frac{u^{3+1}}{3+1} + c$$

$$= \frac{4}{3}\frac{u^4}{4} + c$$

$$= \frac{1}{3}u^4 + c$$

$$= 1/3\,(x^3 + 5)^4 + c \;\ldots\; (\text{since } u = x^3 + 5)$$

$$= \frac{(x^3+5)^4}{3} + c$$

2) Evaluate $\int 9x^4 (x^5 + 7)^8 \, dx$

Solution

Let u = $x^5 + 7$

du/dx = $5x^4$

$5x^4$ dx = du

dx = du/$5x^4$

On substitution we have,

$$\int 9x^4 (x^5 + 7)^8 \, dx = \int 9x^4.u^8 \, dx \;\ldots\; (\text{ since } u = x^5 + 7$$

$$= \int 9x^4.u^8 \frac{du}{5x^4} \;\ldots \left(\sin ce\, dx = \frac{du}{5x^4}\right)$$

$$= \int \frac{9}{5}u^8 \, du$$

$$= \frac{9}{5}\int u^8 \, du$$

$$= \frac{9}{5}\frac{u^9}{9} + c$$

$$= \frac{u^9}{5} + c$$

$$= \frac{(x^5+7)^9}{5} + c \;\ldots(\sin ceu = x^5 + 7)$$

b) Quotient

1) Evaluate $\int \frac{3x}{(x^2-2)^2}\, du$

Solution

Let u = x^2 - 2

du/dx = 2x

2x dx = du

dx = du/2x

On substitution we have,

$$\int \frac{3x}{(x^2-2)^2}\, dx = \int \frac{3x}{u^2}\, dx \;... \;(\sin ce\;\; u = x^2 - 2$$

= ∫ 3x . u^{-2} dx

$$= \int 3x.u^{-2} \frac{du}{2x} \;\; ...(\sin ce\;\; dx = \frac{du}{2x})$$

$$= \int \frac{3}{2} u^{-2}\, du$$

$$= \frac{3}{2}\int u^{-2}\, du$$

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

$$= \frac{3}{2} . \frac{u^{-1}}{(-1)} + c$$

$$= -\frac{3}{2}(x^2-2)^{-1} + c ...(\sin ce\, u = x^2 - 2)$$

2) Evaluate $\int \frac{40x^3}{(20x^4+2)^4}\, dx$

Solution

u = 20 x^4 + 2

du/dx = 80 x^3

80 x^3 dx = du

dx = du/80 x^3

On substitution we have,

$$\int \frac{40x^3}{(20x^4+2}\; dx = \int \frac{40x^3}{u^4} . dx$$

$$= \frac{40x^3}{u^4} \times \frac{du}{80x^3}$$

= ∫1/2u^{-4}du = 1/2∫u^{-4}du

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

$$= -\frac{1}{6} u^{-3} + c$$

$$= -\frac{1}{6} . \frac{1}{u^3} + c$$

$$= \quad -\frac{1}{6u^3}+c$$

$$= \quad -\frac{1}{6(20x^4+2)^3}+c \quad (Since\ u=20x^4+2)$$

Rule VII

Logarithmic Function Rule

In differentiation, we have seen that, if y = log*x*, then,

$$\frac{dy}{dx}=\frac{1}{x}$$

Thus, in Integration, we shall define that,

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

where, C is a constant.

Example

1) $\int \frac{2}{x}dx$

$$= \quad 2\int \frac{1}{x}dx$$

$$= \quad 2 \log x + c$$

2) $\int \frac{1}{x+1}\ dx$

Let z = *x* + 1

dz/d*x* = 1

dz = d*x*

$$\int \frac{1}{z}dx = \quad \log z+c$$

$$= \quad \log (x + 1) + c$$

3) $\int \frac{x}{x^2+1}dx$

Let z = x^2 + 1

dz/d*x* = *2x*

2*x* d*x* = dz

x d*x* = dz/2

$$\int \frac{x}{x^2+1}dx = \int \frac{dz}{2z}=\frac{1}{2}\int \frac{dz}{z} \quad ... (\sin ce\ x\,dx=dz/2\ and\ x^2+1=z)$$

$$= \quad \frac{1}{2}\int \frac{1}{z}dz$$

$$= \quad \frac{1}{2}\log z+c \quad ... \left(\int \frac{1}{x} = \log x+c\right)$$

$$= \quad \frac{1}{2}\log (x^2+1)+c$$

Rule VIII

Exponential Function Rule

In differentiation, we have seen that if y = e^x, then

$$\frac{dy}{dx}=e^x$$

Thus, in integration, we shall define that,

$$y = \int e^x\,dx$$
$$= e^x + c$$

Example

1) $\int e^{x+3}\,dx = e^{x+3} + c$

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

$$= \int e^x dx+\int\frac{1}{x^3}dx$$
$$= \int e^x\,dx+\int x^{-3}dx$$
$$= \int e^x+\frac{x^{-3+1}}{-3+1}+c$$
$$= e^x-\frac{x^{-2}}{2}+c$$
$$= e^x-\frac{1}{2x^2}+c$$

Rule IX

$$\int e^{ax}\,dx = \frac{1}{a}e^{ax}+c$$

Example

1) $\int e^{3x}\,dx = \frac{1}{3}e^{3x}+c$

2) $\int e^{7x}\,dx = \frac{1}{7}e^{7x}+c$

Rule X

$$\int a^{kx}\,dx = \int e^{k(\log a)x}.dx$$
$$= \frac{a^{kx}}{k\log a}+c$$

Example

1) $\int 5^{7x}\,dx = \frac{5^{7x}}{7\log 5}+c$

2) $\int 127^{3x}\,dx=\frac{127^{3x}}{3\log 127}+c$

Rule XI

$$\int (ax^n + b)^n \, dx = \frac{1}{a} \frac{(ax+b)^{n-1}}{n+1} + c$$

1) $\int (5x+2)^7 \, dx$

$$= \frac{1}{5} \frac{(5x+2)^8}{8} + c$$

2) $\int (32x+16)^{15} dx$

$$= \frac{1}{32} \frac{(32x+16)^{16}}{16} + c$$

Rule XII

$$\int x^{n-1} (ax^n + b)^m \, dx = \frac{1}{na} \frac{(ax^n + b)^{m+1}}{m+1} + c$$

Example

1) $\int x^5 (8x^6 + 7)^{15} \, dx$

$$= \frac{1}{6 \times 8} \frac{(8x^6 + 7)^{16}}{16} + c$$

$$= \frac{1}{48} \frac{(8x^6 + 7)^{16}}{16} + c$$

2) $\int x^8 (7x^9 + 15)^{32} \, dx$

$$= \frac{1}{63} \frac{(7x^9 + 15)^{33}}{33} + c$$

Definite Integration

We found that finding the antiderivative of a function achieved and indefinite result (no definite numerical values) and so we called this process as "Indefinite Integration". Now, we have to use the other property of the integral to find the area between two curves and the area between the curve and the x-axis, we shall achieve a definite numerical result i.e., a number or a value independent of the constant 'c' and not a function as for the Indefinite Integral. Therefore, we call this process as "Definite Integration".

Notation for Definite Integration

$$\int_a^b y \, dx \ or \int_a^b f(x) \, dx$$

It is to be read as "the Definite Integral" of y or f(x) from x = a to x = b. Then the value of Definite Integral from a to b is written as,

$$\int_a^b y \, dx \ \ or \ \int_a^b f(x) \, dx = \left[F(x)\right]_a^b = F(b) - F(a) \ldots \ b > a$$

Example

1) $\int_1^2 x^2 \, dx$

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

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

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

$$= \frac{8}{3} - \frac{1}{3}$$

$$= \frac{7}{3}$$

2) $\int_2^3 3x\,dx$

$$= \left[3\frac{x^{1+1}}{1+1}\right]_2^3$$

$$= \left[3\frac{x^2}{2}\right]_2^3$$

$$= \left[3\,\frac{3^2}{2} - 3\frac{2^2}{2}\right]$$

$$= \frac{27}{2} - \frac{12}{2}$$

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

3) $\int_2^3 (x^2 + 5x + 7)\,dx$

$$= \left[\frac{x^{2+1}}{2+1} + 5\frac{x^{1+1}}{1+1} + 7x\right]_2^3$$

$$= \left[\frac{x^3}{3} + \frac{5x^2}{2} + 7x\right]_2^3$$

$$= \left[\frac{3^3}{3} + 5\,\frac{3^2}{2} + 7(3)\right] - \left[\frac{2^3}{3} + 5\frac{2^2}{2} + 7(2)\right]$$

$$= \left[\frac{27}{3} + \frac{45}{2} + 21\right] - \left[\frac{8}{3} + \frac{20}{2} + 14\right]$$

$$= \left[\frac{54+135+126}{6}\right] - \left[\frac{16+60+84}{6}\right]$$

$$= \frac{315}{6} - \frac{160}{6}$$

$$= \frac{155}{6}$$

Application of integration to the calculation of total function from marginal function
Cost Function

In differentiation, we have seen that if total cost (c) of producing an output x is given, we have to find out the Marginal Cost (MC), that is first order derivative of the Total Cost (C).

$$MC = \frac{dc}{dx}$$

Thus in integration, if MC is given we have to find out the total cost (C). That is Total cost (C) is the integral of MC, (i.e. dc/dx) with reference to x.

i.e) $C = \int MC\, dx$

$= \int \frac{dc}{dx} dx$

Example

1) If MC = 3 - 2x - x^2, find the Total Cost.

Solution

Total Cost $= \int MC\, dx$

$= \int (3 - 2x - x^2)\, dx$

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

2) Compute Total Cost for the Marginal Cost Function C = 2 + 6x - 4x^2, if total fixed cost is 50.

Solution

Marginal Cost $= 2 + 6x - 4x^2$

Total Cost $= \int MC\, dx$

$= \int (2 + 6x - 4x^2)\, dx$

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

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

Since total fixed cost is 50,

Total Cost $= 2x + 3x^2 - \frac{4}{3}x^3 + 50$

3) Compute Total, Average cost for the Marginal Cost function C = 4 + 7x - 5x^2, if the total fixed cost is 40.

Solution

Marginal Cost $= 4 + 7x - 5x^2$

Total Cost $= \int MC\, dx$

$= \int (4 + 7x - 5x^2)\, dx$

$= 4x + \frac{7}{2}x^2 - \frac{5}{3}x^3 + 40$

Average Cost = TC/x

$$= \frac{4x + \frac{7}{2}x^2 - \frac{5}{3}x^3 + 40}{x}$$

$$= 4 + \frac{7}{2}x - \frac{5}{3}x^2 + \frac{40}{x}$$

Revenue Function

In differentiation, we have seen that, if Total Revenue (TR or R) of producing an output X is given, we have to find out the Marginal Revenue (MR), that is the first order derivative of the Total Revenue (TR or R). i.e) MR = dR/dx

Thus in integration, if MR is given, we have to find out the Total Revenue (TR or R). That is the Total Revenue (TR or R) is the integral of MR (i.e dR/dx) with reference to x.

i.e, TR or R = $\int MR.dx = \int \frac{dR}{dx} dx$

Example

1) If the Marginal Revenue Function MR = 100 - 4Q find the Total Revenue Function

Solution

MR = 100 - 4Q

Therefore, TR = ∫ (100 - 4Q) dQ

$$= 100Q - \frac{4Q^2}{2} + c$$

$= 100Q - 2Q^2 + C$

2) If the Marginal Revenue Function MR = 9 - 4x^2, find out the Total Average Revenue Function

Solution

Marginal Revenue = $9 - 4x^2$

Total Revenue = $\int (9 - 4x^2)\, dx$

$$= 9x - \frac{4x^3}{3} + c$$

$$\text{Average Revenue} = \frac{Total\ Revenue}{Output}$$

$$= \frac{R}{X}$$

$$= \frac{9x - \frac{4x^3}{3}}{X}$$

$= 9 - 4x^2/3$

3) Find the Total Revenue and the Demand Functions, if MR = $3x^2 - 2x + 9$.

Solution

Marginal Revenue = $3x^2 - 2x + 9$

Total Revenue = $\int (3x^2 - 2x + 9)\, dx$

$$= \frac{3x^3}{3} - \frac{2x^2}{2} + 9x + c$$

$$= x^3 - x^2 + 9x + c$$

Demand Function P $= \dfrac{R}{X}$

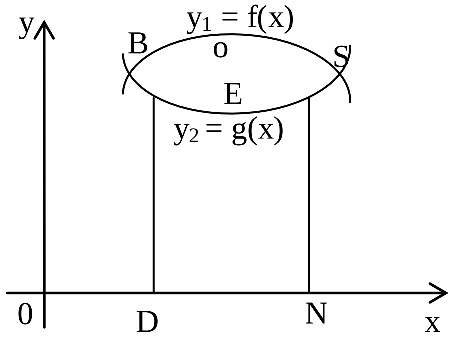
$$= \frac{x^3 - x^2 + 9x}{x}$$

$$= x^2 - x + 9$$

Area Between two curves

Let $y_1 = f(x)$ be the curve BOS and $y_2 = g(x)$ be the curve BES. Suppose they intersect at B and S where 'X' co-ordinates are OD = a and ON = b.

The above statement is illustrated in the diagram.



Area between two curves

We have to find the area BOSE, which is equal to the area DBOSN minus DBESN.

Area BOSE $= \int_a^b f(x)\,dx - \int_a^b g(x)\,dx$

Area BOSE $= \int_a^b [f(x) - g(x)]\,dx$

Example

1) Find the area beneath the curve $y = x^5$ between $x = 2$ and $x = 3$

Solution

Required area $= \int_2^3 x^5 dx$

$$= \left[\frac{x^{5+1}}{5+1}\right]_2^3$$

$$= \left[\frac{x^6}{6}\right]_2^3$$

$$= \frac{3^6}{6} - \frac{2^6}{6}$$

$$= \frac{729 - 64}{6}$$

$$= \frac{665}{6}$$

$$= 110.83$$

2) Calculate the area beneath the curve y = x between x = 3 and x = 6

Solution

The required area $= \int_3^6 x^3\, dx$

$$= \left[\frac{x^{3+1}}{3+1}\right]_3^6$$

$$= \left[\frac{x^4}{4}\right]_3^6$$

$$= \frac{6^4}{4} - \frac{3^4}{4}$$

$$= \frac{1296 - 81}{4}$$

$$= \frac{1215}{4}$$

$$= 303.75$$

Consumer Surplus

Meaning

Consumer's surplus is the difference between the price that a consumer is willing to pay for commodity rather than go without it and the actual price he pays for the commodity. Therefore,

Consumer's surplus = What a person is willing to pay - What he actually pays

(or)

Consumer's surplus = Potential price - Actual Price

Explanation

It is assumed that the marginal utility of money is constant and all the consumers have the same utility function. A Demand curve or a Demand Function for a commodity represents the amount of that commodity which will be bought by people at a given price 'p'.

Let p = f(x) be the demand function for a commodity. Suppose a consumer purchases X_0 quantity at p_0 price [$\therefore p_0 = f(x_0)$]. This implies that at price p_0, the consumer is willing to purchase and the producer is willing to sell a quantity x_0.

Thus, the total expenditure of the consumer = p_0x_0 …. (1)

However, there are buyers who would be willing to pay a price higher than p_0. Having already purchased a quantity x, he would be willing to purchase of a quantity dx at the price f(x). Thus, the expenditure would be f(x)dx.

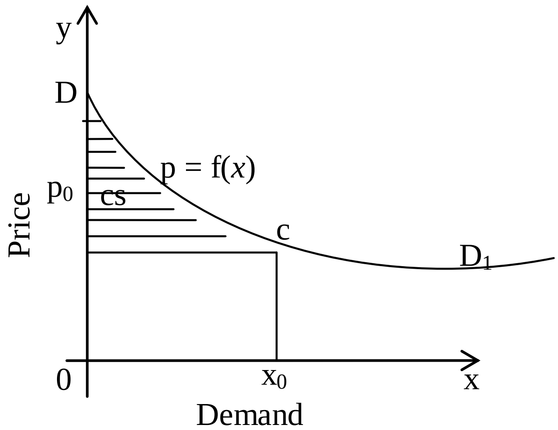
Thus, the total expenditure he would have been willing to pay for the quantity x_0, is

$$\int_0^{x_0} f(x)\, dx \ldots.(2)$$

The difference between the two is called consumer's surplus. Therefore, Consumer's surplus $= \int_0^{x_0} f(x)\,dx - p_0 x_0$

The consumer's surplus can be expressed as shown in the diagram.

Diagrammatic Repressentation



Explanation of the Diagram

In the diagram x-axis represents the Demand and y-axis represents the price. DD_1 is the Demand Curve. Demand X_0 corresponds to p_0. Then the consumer's surplus is

DCp_0 = Area $DCx_0 0$ - Area $p_0 C x_0 0$ (shaded area)

$$= \int_0^{x_0} f(x)\,dx - p_0 X_0 \;]$$

$$= \int_0^{x_0} p\,dx - p_0 x_0$$

The different is called the "Consumer's Surplus" and denoted by 'CS'.

This area (i.e., Consumer's Surplus) can easily be found out with the help of definite integral.

Example

1) If the demand function is $p = 25 - 3x - 3x^2$ and the demand x_0 is 2, what will be the Consumer's Surplus?

Solution

$p = 25 - 3x - 3x^2$

Let us substitute $x_0 = 2$ instead of x in Demand Function.

$p_0 = 25 - 3(2) - 3(2)^2$

$= 25 - 6 - 12$

$= 25 - 18$

$= 7$

Therefore $p_0 = 7$ and $x_0 = 2$

Consumer's Surplus = $\int_0^{x_0} p.dx - p_0 x_0$

$= \int_0^2 (25 - 3x - 3x^2)dx - (7 \times 2)$

$= \left[25x - \frac{3x^2}{2} - x^3\right]_0^2 - 14$

= (50 - 6 - 8) - 14

= 36 - 14

= 22

Consumer's Surplus = 22

2) The demand function for a commodity p = 30 - 2D. The supply function p = 3D. Find the Consumer's Surplus.

Solution

Demand Function p = 30 - 2D

Supply Function p = 3D

Since Demand = Supply

30 - 2D = -30

- 5D = 30

D = 30/5

= 6

Substitute the value of D = 6 in the supply function, we can get the value of p_0.

i.e, $p_0 = 3 \times 6 = 18$

Therefore, p_0 and $D_0 = 6$

Consumer's Surplus = $\int_0^6 pdD - p_0 D_0$

$= \int_0^6 (30 - 2D)\, dD - (18 \times 6)$

$= \left[30D - D^2\right]_0^6 - 108$

$= (30(6)-(6)^2) - 108$

= (180 – 36) - 108

= 36

Therefore, Consumer's Surplus = 36

Producer's Surplus

Meaning

It is assumed that Marginal Utility of money is constant and all the producers have the same production function. Let p = f(x) be the supply function or supply curve represents the amount of commodity that can be supplied at a given price p i.e., Market price p_0 ($\therefore$ $p_0 = f(x_0)$). This implies that at this price consumer is willing to purchase and producer is willing to sell a quantity x_0.

Thus the producer's revenue = $p_0 x_0$ … (1)

However, there are producers who are willing to sell or supply the commodity at a price lower than p_0. Having already sold a quantity x, he would be willing to sell a quantity d*x*, at the price f(*x*)d*x*. Hence, his true revenue in selling a quantity x would have been,

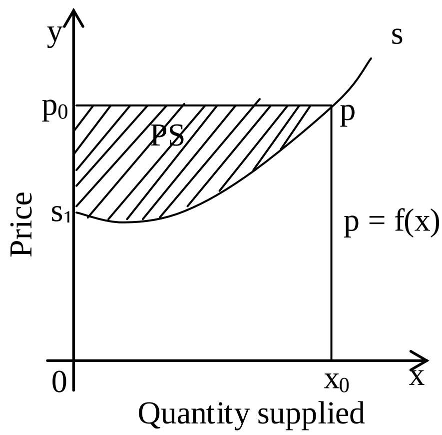
$$\int_0^{x_0} f(x)dx.....(2)$$

Therefore, the difference between the two is called "producer's Surplus" and is denoted by PS.

Producer's Surplus (PS) = $p_0x_0 - \int_0^{x_0} f(x)dx$

The producer's Surplus can be expressed as shown in the diagramme.

Diagrammatic Representation



Explanation of the Diagram

In the diagram, x - axis represents quantity supplied and y - axis represents price. SS_1 is the market supply curve. Supply X_0 corresponds to p_0. OP_0 is the market price. OS, is producer's willingness to sell or producer's willingness price.

Therefore,

Producer's Surplus = Area of the whole rectangle (p_0ox_0p) - Area under the supply curve (s_1ox_0p)

= shaded area above the supply curve (p_0s_1p)

= p_0x_0 - f(x)dx

= $p_0X_0 - \int_0^{x_0} p\,dx$

Therefore, this area (i.e., Producer's Surplus) can also be easily found out with the help of definite integral.

Example

1) The supply function for a commodity p = $x^2 - x + 5$, where x denotes supply. Find the Producer's Surplus when the price is Rs.11.

Solution

P = $x^2 - x + 5$

Let us substitute the value of p(i.e p = 1) in the supply function,

11 = $x^2 - x + 5$

$x^2 - x + 5$ = 11

$x^2 - x$ = 11 – 5 = 6

$x^2 - x - 6 = 0$

$x^2 - 3x + 2x = 0$

$x^2 - 3x + 2x - 6 = 0$

$x(x - 3) + 2(x - 3) = 0$

Therefore, $(x - 3)(x + 2) = 0$, $x = 3$ or $x = -2$

Since supply cannot be negative, $x = 3$

Therefore, $p_0 = 11$ and $x_0 = 3$ Producer's Surplus $= p_0 X_0 - \int_0^{x_0} p\,dx$

$$= 11 \times 3 - \int_0^3 (x^2 - x + 5)\,dx$$

$$= 33 - \left[\frac{x^3}{3} - \frac{x^2}{2} + 5x\right]_0^3$$

$$= 33 - \left[\frac{3^3}{3} - \frac{3^2}{2} + 5(3)\right] - 0$$

$$= 33 - \left[\frac{27}{3} - \frac{9}{2} + 15\right] - 0$$

$$= 33 - \frac{27}{3} + \frac{9}{2} - 15$$

$$= 33 - 9 + 4.5 - 15$$

$$= 13.5$$

Therefore, Producer's Surplus = 13.5

2) The supply function for a commodity $p = 2 + D^2$. Find the Producer's Surplus when price is Rs 18.

Solution

$p = 2 + D^2$

Let us substitute the value of p (i.e p = 18) in the supply function.

$18 = 2 + D^2$

$2 + D^2 = 18$

$D^2 = 18 - 2$

$= 16$

$D = \sqrt{16}$

$D = \pm 4$

Since the supply cannot be negative, D = 4

$p_0 = 18$ and $D_0 = 4$

Producer's surplus $= p_0 D_0 - \int_0^{D_0} p\,dD$

$$= 18 \times 4 - \int_0^{D_0} (2 + D^2)\,dD$$

$$= 18 \times 4 - \left[2D + \frac{D^{2+1}}{2+1}\right]_0^4$$

$$= \quad 18\times 4-\left[2(4)+\frac{4^3}{3}\right]$$

$$= \quad \left[72-8+\frac{64}{3}\right]-0$$

= 72 - [8 + 21.3]

= 72 - 29.3

= 42.7

Producer's Surplus = 42.7

3) Given the Demand function p = 8 - 2*x* and the supply function p = 2 + *x*, find the Consumer's Surplus and the Producer's Surplus

Solution

Demand Function is p = 8 - 2*x*

Supply Function is p = 2 + *x*

Since Demand = Supply

8 - 2*x* = 2 + *x*

-2*x* - *x* = 2 - 8

- 3*x* = -6

3*x* = 6

x = 6/3

= 2

Substitute the value of x(i.e., *x* = 2) in the demand function, we can get the value of p.

p = 8 - 2(2)

p = 8 - 4

p = 4

Therefore $p_0 = 4$ and $x_0 = 2$

$$\text{Consumer's Surplus} = \int_0^{x_0} p\,dx - p_0 X_0$$

$$= \int_0^{2} (8-2x)\,dx - 4\times 2$$

$$= \left[8x - x^2\right]_0^2 - 8$$

$$= \left[8(2) - 2^2\right] - 8$$

$$= \left[16-4\right] - 8$$

= 12 - 8

= 4

Therefore, Consumer Surplus = 4

$$\text{Producer's Surplus} = p_0 X_0 - \int_0^{x_0} p\,dx$$

$$= 4\times 2 - \int_0^{2} (2+x)\,dx$$

$$= 8 - \left[2x + \frac{x^2}{2}\right]_0^2$$

$$= \quad 8 - \left[2(2) + \frac{2^2}{2} \right]$$

$$= \quad 8 - [4 + 2]$$

$$= \quad 8 - 6$$

$$= \quad 2$$

Therefore Producer's Surplus = 2

UNIT-III

CORRELATION AND REGRESSION ANALYSIS

Meaning

Correlation is a statistical technique which measures and analyses the degree or extent to which two or more variables fluctuate with reference to each other. Correlation thus denotes the inter-dependence amongst variables. The degrees are expressed by a coefficient which ranges between -1 and +1. The direction of change is indicated by + or – signs; the former refers to the sympathetic movement in the same direction and the later in opposite direction. An absence of correlation is indicated by zero. Correlation thus expresses the relationship through a relative measure of change and it has nothing to do with the units in which variables are expressed.

Assumptions of simple correlation

1) Linear Relationship

The product moment-coefficient of correlation assumes essentially a linear relationship between the variables. The degree of relationship between the variables becomes progressively poorer the more the relationship departs from the straight line form.

2) Casual Relationship

In studying correlation, we expect a cause and effect relationship between the forces affecting the values in the two series. In the absence of a cause and effect relationship, the correlation is meaningless. For instance, we may find some correlation between income and height but, in the absence of cause and effect relationship between income and weight, such a correlation will have no sense.

3) Normality

A large number of independent factors are operative on each of the variables being correlated in a way so as to form normal distribution. The variables like weight, height, age, supply, demand, etc. are the examples where the forces affecting these form normal distribution.

4) Error of measurement

The coefficient of correlation is more reliable if the error of measurement is reduced to the minimum.

Limitations

1) It assumes that there is a linear relationship between the variables. This is not always true.
2) The calculation process of ‘r’ is a time consuming one.
3) A great care is needed to interpret the value of ‘r’.
4) The value of ‘r’ is affected by the value of extreme items.

Karl Pearson’s Correlation Coefficient Method

The most widely used mathematical method for measuring the intensity or the magnitude of linear relationship between two variables was suggested by great British Biometrician Karl Pearson (1967-1936). Karl Pearson’s measure, known as pearsonian correlation. The algebraic sign of pearsonian coefficient of correlation indicates the direction of relationship between two variables, i.e., whether it is positive or negative. If coefficient is in plus (+), it will be positive correlation, in case of minus (-) the correlation is negative.

Karl Pearson's coefficient of correlation is a quantitative measure of relationship between two variables, and if lies between ± 1(-1 ≤ r ≤ 1). Plus one (+1) represents perfect positive correlation and minus one (-1) perfect negative correlation. Zero represents absence of correlation. The results between (±) are interpreted as having correlation of different degrees, based on how far the coefficient is away from zero and near one.

Degrees of correlation	**Positive**	**Negative**
Perfect correlation	+1	-1
Very high degree of correlation	+ .9 or more	- .9 or more
Suffiently high degree of correlation	From + .75 to + .9	From - . 75 to - .9
Moderate degrees of correlation	From + .6 to + .75	From - .6 to - .75
Only the possibility of a correlation	From + .3 to + .6	From - .3 to - .6
Possibly no correlation	Less than + .3	Less then - .3
Absence of correlation	0	0

Pearsonian coefficient of correlation is regarded as the most satisfactory measure, because it is based on mean and standard deviation, which are amenable to algebraic manipulations due to their mathematical properties. The scale of measure (±1) is also helpful in scientific interpretation of the coefficient.

If r = + 1, there exist a perfect positive correlation between the variables.

If r = 0, there is no correlation between the variables.

If r = - 1, there exist a perfect negative correlation between two variables.

Problems

Model I

Values of the variables x and y are given to calculate the co-efficient of correlation r

1) Compute Karl-Pearson's correlation coefficient from the following data.

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

Solution

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

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{N\sum x^2 - (\sum x)^2}\sqrt{N\sum y^2 - (\sum y)^2}}$$

where

N = 8, $\sum x = 544$, $\sum y = 552$, $\sum x^2 = 37028$, $\sum y^2 = 38132$, $\sum xy = 37560$

$$r = \frac{8(37560) - (544)(552)}{\sqrt{8(37028) - (544)^2}\sqrt{8(38132) - (552)^2}}$$

$$= \frac{300480 - 300288}{\sqrt{296224 - 295936}\sqrt{305056 - 304704}}$$

$$= \frac{192}{\sqrt{288}\sqrt{352}}$$

$$= \frac{192}{16.97 \times 18.76}$$

$$= \frac{192}{318.35}$$

r = 0.603

Short-cut method

2) Compute Karl Pearson's correlation co-efficient from the following data. (same problem)

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

Solution

Let a = 67, b = 72

X	**y**	**d*x* = *x* - a** ***x* - 67**	**dy = *x* - b** ***x* – 72**	**dxdy**	**d^2x**	**d^2y**
65	67	-2	-5	10	4	25
66	68	-1	-4	4	1	16
67	65	0	-7	0	0	49
67	68	0	-4	0	0	16
68	72	1	0	0	1	0
69	72	2	0	0	4	0
70	69	3	-3	-9	9	9
72	71	5	-1	-5	25	1
$\sum x$= 544	$\sum y$ = 552	$\sum dx$ = 8	$\sum dy$ = -24	$\sum dxdy$ = 0	$\sum d^2x$ = 44	$\sum d^2y$ = 116

$$r = \frac{N\sum dx\,dy - (\sum dx)(\sum dy)}{\sqrt{N\sum d^2x - (\sum dx)^2}\sqrt{N\sum d^2y - (\sum dy)^2}}$$

$$= \frac{8(0)-(8)(-24)}{\sqrt{8(44)-8^2}\sqrt{8(116)-(-24)^2}}$$

$$= \frac{192}{\sqrt{288}\sqrt{352}}$$

$$= \frac{192}{(16.67)(18.76)}$$

$$= \frac{192}{318.35}$$

$$r = 0.603$$

3) Find out correlation coefficient for the data and interprets it's value

x	10	12	13	14	16	15
y	40	38	43	45	37	43

Solution

x	**y**	x^2	y^2	xy
10	40	100	1600	400
12	38	144	1444	456
13	43	169	1849	559
14	45	196	2025	630
16	37	256	1369	592
15	43	225	1849	645
$\sum x = 80$	$\sum y = 246$	$\sum x^2 = 1090$	$\sum y^2 = 10136$	$\sum xy = 3282$

$$r = \frac{N\sum xy-(\sum x)(\sum y)}{\sqrt{N\sum x^2-(\sum x)^2}\sqrt{N\sum y^2-(\sum y^2)}}$$

$$= \frac{(6)(3282)-(80)(246)}{\sqrt{6\,(1090)-(80)^2}\sqrt{6(10136)-(246)^2}}$$

$$= \frac{19692-19680}{\sqrt{6540-6400}\sqrt{60816-60516}}$$

$$= \frac{12}{\sqrt{140}\sqrt{300}}$$

$$= \frac{12}{(11.83)(17.32)}$$

$$= \frac{12}{204.89}$$

$$= 0.06$$

Coefficient of correlation is 0.06 which shows low degree of positive correlation.

Model II

Under this model the values of the variable are not given.

Problem

1) Calculate coefficient of correlation and interpret its value $\sum x = 136$, $\sum y = 243$, $\sum x^2 = 2278$, $\sum xy = 3476$, $\sum y^2 = 6129$, N = 10

Solution

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{N\sum x^2 - (\sum x)^2} - \sqrt{N\sum y^2 - (\sum y)^2}}$$

$$= \frac{(10)(3476) - (136)(243)}{\sqrt{(10)(2278) - (136)^2}\sqrt{(10)(6129) - (243)^2}}$$

$$= \frac{34760 - 33048}{\sqrt{22780 - 18496}\sqrt{61290 - 59049}}$$

$$= \frac{1712}{\sqrt{4289}\sqrt{2241}}$$

$$= \frac{1712}{3098.4}$$

$$= 0.55$$

Coefficient of correlation 0.55 which shows moderate degree of positive correlation.

Model III

Under this model the sum of deviation of values of the variables *x* and y from their means will be given to find coefficient of correlation.

Problem

1) Compute coefficient of correlation for $\sum xy = 58$, $\sum x^2 = 94$, $\sum y^2 = 54$

Solution

$$r = \frac{\sum xy}{\sqrt{\sum x^2}\sqrt{\sum y^2}}$$

$$= \frac{58}{\sqrt{94}\sqrt{54}}$$

$$= \frac{58}{(9.69)(7.34)}$$

$$= \frac{58}{71.2}$$

$$r = 0.81$$

The coefficient of correlation shows high degree of positive correlation in the values of *x* and y.

2) Calculate the co-efficient of correlation from the following data N = 10, $\sum x = 100$, $\sum y = 150$, $\sum(x - 10)^2 = 180$, $\sum(y - 15)^2 = 215$, $\sum(x - 10)(y - 15) = 60$

Solution

$$\bar{x} = \frac{\sum x}{n}$$

$$= \frac{100}{10}$$

$$= 10$$

$$\bar{y} = \frac{\sum y}{n}$$

$$= \frac{150}{10}$$

$$= 15$$

given

$\sum (x - 10)(y - 15) \Rightarrow \sum xy = 60$

$\sum (x - 10)^2 \Rightarrow \sum x^2 = 180$

$\sum (y - 15)^2 \Rightarrow \sum y^2 = 215$

N = 10

$$r = \frac{\sum xy}{\sqrt{\sum x^2}\sqrt{\sum y^2}}$$

$$= \frac{60}{\sqrt{180}\sqrt{215}}$$

$$= \frac{60}{(13.4)(14.66)}$$

$$= \frac{60}{196.48}$$

$$r = 0.305$$

3) From the given data compute N, if r = 0.5, $\sum xy = 120$, $\sigma y = 8$, $\sum x^2 = 90$

Solution

$$r = \frac{\sum xy}{n\sigma x\, \sigma y}$$

$$\sigma x = \sqrt{\frac{\sum (x-\bar{x})^2}{n}}$$

$$\sigma^2 x = \frac{\sum (x-\bar{x})^2}{n}$$

$$= \frac{\sum x^2}{n}$$

$$r^2 = \frac{(\sum xy)^2}{n^2 \sigma^2 x \sigma^2 y}$$

$$(0.5)^2 = \frac{(120)^2}{(n)^2 \left(\frac{\sum x2}{n}\right)(8)^2}$$

$$0.25 = \frac{14400}{n(90)(64)}$$

$$= \frac{14400}{n(5760)}$$

$$n\,(5760)\,(0.25) = 14400$$

$$n(1440) = 14400$$

$$n = \frac{14400}{1440}$$

$$n = 10$$

Assumptions, merits and Demerits of Karl Pearson's coefficient of correlation

Assumptions

The Karl Pearson's co-efficient of correlation is based on the following assumption.

1) There exists the linear relationship.
2) There exists cause and effect relationship between the variables.
3) Equal number of pairs of variables exists.
4) The data must be homogeneous.

Merits

1) It is based on mean and standard deviation of the variables. Therefore it is capable for mathematical calculations.
2) It is a quantitative measurement of relationship between variables
3) The algebraic sign and magnitude of 'r' indicates direction (movement) and closeness of relationship between variables.

Demerits

1) It assumes that there is a linear relationship between the variables. This is not always true.
2) The calculation process of 'r' is a time consuming one.
3) A great care is needed to interpret the value of 'r'.
4) The value of 'r' is affected by the value of extreme items.

Properties of co-efficient of correlation

1) The co-efficient of correlation always lies between -1 and +1

   i.e) $-1 \le r \le +1$
2) It is a pure number independent of the unit of measurements of the given data.
3) Co-efficient of correlation between two variables is always symmetric

   i.e) $r_{xy} = r_{yx}$
4) The coefficient of correlation is independence of origin and scale of the given data.
5) Two independent variables are always uncorrelated.

   i.e) $r_{xy} = 0$

Co-efficient of Rank Correlation

A method of degree of association between the ranks of observations of two variables (qualitative) is called rank correlation. If also refers to the coefficient of correlation between the ranks of two variables.

Spearman's Rank Correlation Co-efficient

A British psychologist Edwin spearman introduced a method to measure the correlation based on the ranks (orders) of the observations.

The formula for rank correlation coefficient is

$$\rho \quad = \quad 1 - \frac{6\sum D^2}{N(N^2-1)}$$

Here

ρ = rank correlation co-efficient

D = Difference between the ranks of observation.

The rank correlation coefficient row (ρ) lies between -1 and +1.

Interpretation of 'ρ'

If 'ρ' = +1, there is complete agreement in the order of ranks of the variable.

If 'ρ' = -1, there is complete disagreement in the order of ranks of the variable

If 'ρ' = 0, there is no association between the ranks of the variables.

Model - I

Under this model the actual ranks of the observations are given to calculate 'ρ'.

Problem

1) Calculate spearman's rank correlation for the ranks of the two variables *x* and y

Rank of X	1	2	3	4	5	6	7	8	9	10
Rank of Y	2	4	3	1	6	7	5	9	10	8

Solution

Rank of x	**Rank of y**	**$D = R_x - R_y$**	**D^2**
1	2	-1	1
2	4	-2	4
3	3	0	0
4	1	3	9
5	6	-1	1
6	7	-1	1
7	5	2	4
8	9	-1	1
9	10	-1	1
10	8	2	4
			$\sum D^2 = 26$

$$\rho \quad = \quad 1 - \frac{6\sum D^2}{N(N^2-1)}$$

$$= \quad 1 - \frac{6(26)}{10\,(10^2-1)}$$

$$= \quad 1 - \frac{156}{10(100-1)}$$

$$= \quad 1 - \frac{156}{(10)\,(99)}$$

$$= \quad 1 - \frac{156}{990}$$

= 1 - 0.157

ρ = 0.843

The value of row shows that it is completely agree with each other.

Model - III

Under this model the actual ranks of the observations are not given. But the values of observation will be given to compute 'p'

Procedure

Firstly ranks to be allotted to the variables according to the values of the variables.
To allot rank one can follow either ascending order or descending order.

Problem

1) Calculate Spearman's rank correlation coefficient from the given data.

x	90	102	45	75	80	92	60	110
y	85	76	51	60	90	110	58	72

Solution

X	**y**	**Rx**	**Ry**	**D = R_x - Ry**	**D^2**
90	85	4	3	1	1
102	76	2	4	-2	4
45	51	8	8	0	0
75	60	6	6	0	0
80	90	5	2	3	9
92	110	3	1	2	4
60	58	7	7	0	0
110	72	1	5	-4	16
					$\sum D^2$ = 34

$$\rho \quad = \quad 1 - \frac{6\sum D^2}{N(N^2-1)}$$

$$= \quad 1 - \frac{6 \times 34}{8(8^2-1)}$$

$$= \quad 1 - \frac{204}{8(64-1)}$$

$$= \quad 1 - \frac{204}{(8)(63)}$$

$$= \quad 1 - \frac{204}{504}$$

$= \quad 1 - 0.405$

$= \quad 0.595$

The ranks of *x* and y agree with each other.

Model III

Under this model ranks of the observations are repeated. Following formula is used

$$\rho = 1 - \frac{6\left[\sum D^2 + \frac{1}{12}(m^3 - m)\right]}{N(N^2 - 1)}$$

D - Difference between the rank of variables *x* and y.

M - No.of times the value of variables repeated.

Problem

1) Find the spearman's rank correlation coefficient from the given data.

x	50	50	55	60	65	65	65	60	60	50
y	11	13	14	16	16	15	15	14	13	13

Solution

X	**y**	**R_x**	**R_y**	**D = R_x - R_y**	**D^2**
50	11	9	10	-1	1
50	13	9	8	1	1
55	14	7	5.5	1.5	2.25
60	16	5	1.5	3.5	12.25
65	16	2	1.5	0.5	0.25
65	15	2	3.5	-1.5	2.25
65	15	2	3.5	-1.5	2.25
60	14	5	5.5	-0.5	0.25
60	13	5	8	-3	9
50	13	9	8	1	1
					$\sum D^2 = 31.5$

$$\rho \quad = \quad \frac{6\left[3D^2 + \frac{1}{12}(m^3 - m)\right]}{N(N^2 - 1)}$$

$$\frac{1}{12}(m^3 - m) \quad = \quad \frac{1}{12}\left[(3^3 - 3) + (3^3 - 3) + (3^3 - 3) + (2^3 - 2) + (2^3 - 2) + (2^3 - 2)\right]$$

$$= \quad \frac{1}{12}\left[24 + 24 + 24 + 6 + 6 + 6\right]$$

$$= \quad \frac{1}{12}\left[114\right]$$

$$= \quad 9.5$$

$$\rho = \frac{6\left[3D^2 + \frac{1}{12}(m^3 - m)\right]}{N(N^2-1)}$$

$$= \frac{1 - 6\times 31.5 + (9.5)}{10(10^2-1)}$$

$$= 1 - \frac{198.5}{10\times 99}$$

$$= 1 - \frac{198.5}{990}$$

$$= 1 - 0.20$$

$$\rho = 0.8$$

Probable Error

Like other measures, coefficients of correlation are generally calculated from samples. Therefore for determining their reliability, a measure called probable Error is used . Probable error is an amount which if added to and deducted from the coefficient of correlation produces range within which coefficients of correlation of other groups selected from same series at random will fall, with a probability of 50%. The probable error (P.E) of the coefficient of correlation is calculated by using the formula P.E (r) = $\frac{0.6745\,(1-r^2)}{\sqrt{n}}$, where n is size of sample drawn randomly from a bivariate normal population.
There are two functions of the probable error

1) Determination of Limits

Using the value of simple correlation coefficient and its probable error, one can determine the two limits (r ± P.E) within which, there is 50% probability that coefficients of correlation of randomly selected from the same population will be.

2) Interpretation of r

Probable error is regarded as a measure of significance of Karl Pearson's coefficient of correlation. In this regards, a few interpretations are as under.
If r < P.E (r), i.e., if the observed value of r is less than its P.E., then correlation is not at all significant.
If r > 6 P.E. (r), i.e., if observed value of r is greater than 6 times its P.E, then r is definitely significant.
In other situations, nothing can be concluded with certainly

Problem

Find the probable error if r = 0.6 and N = 64

Solution

$$\text{P.E (r)} = (0.6745)\,\frac{1-r^2}{\sqrt{N}}$$

$$= (0.6745)\,\frac{1-(0.6)^2}{\sqrt{64}}$$

$$= \quad (0.6745)\frac{1-0.36}{8}$$

$$= \quad (0.6745)\frac{(0.64)}{8}$$

$$= \quad (0.6745)\ (0.08)$$

$$P.E\ (r) = 0.053$$

Regression

Relationship between two or more variable in the form of an equation to estimate the value of one variable, given the value of other variables, is called 'Regression Analysis'. The linear algebraic equation that is used for expressing a dependent variable in terms of independent variable is called 'linear regression equation'. The literal or dictionary meaning of regression is 'moving backward', 'going back' or 'the return to the mean value'.

Definition

Various experts have defined regression in their own words in different ways and some important definitions of regression are given below.

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

According to Ya Lun Chou, "Regression analysis attempts to establish the nature of the relationship, between variables, i.e., to study the functional relationship between the variables and thereby provide a mechanism for prediction, or forecasting."

Regression analysis is a statistical tool with the help of which one is in a position to estimate (or predict) the unknown values of one variable from known values of another variable. The variable whose value is influenced or is to be predicted is called dependent variable (also known as regressed or explained variable) and the variable which influences the values or is used for prediction, is called independent variable (also known as regresser or predictor or explanatory variable). The independent is usually denoted by x and the dependent variable by y.

Difference between correlation and regression

Correlation	Regression
1) It studies the degree of relationship between variables	It studies the nature of relationship between variables.
2) It need not imply cause and effect relationship between variable	It implies cause and effect relationship between variables
3) There may be nonsense correlation between the variables.	There is nothing like nonsense regression.
4) The correlation coefficient is independent of change of origin and scale.	The regression coefficients are independent of only change of origin but not of scale.
5) The correlation coefficient can not be used for prediction.	The regression lines can be used for prediction.

Types of regression

1) Simple Regression

The measure of average relationship between one dependent variable and one independent variable is called simple regression Example d = f(p)

2) Multiple regression

The measure of average relationship between one dependent variable and two or more independent variable is called multiple regression Example Q = f (L, K, D, O)

3) Partial regression

The measure of average relationship between a dependent variable and anyone independent variable keeping other independent variable as constant in a group of independent variable is called as partial regression.

Simple linear regression

In regression, if the ratio of change in the value of dependent variable (x) to the change in the value of independent variable (y) remains constant is called simple linear regression.

Simple non-linear regression

In regression if the ratio of change in the value of dependent variable (x) to the change in the value of independent variable (y) do not remain constant is called non linear regression.

Regression Line

A line fitted to a given set of dependent variable and independent variable to estimate the value of one variable for a given specific value of other variable is called regression line .

There are two regression lines

1) Regression line y on *x*

In this case the variable y is taken as dependent variable and x is taken as independent variable. This gives the most probable value of y for a given value of x.

2) Regression line *x* on y

In this case, the variable x is taken as dependent variable and y is taken as independent variable. This gives the most probable value of x for a given value of y.

Regression Equation

An algebraic equation explains the relationship between a set of dependent variable and an independent variable is called regression equation.

1) Regression equation y on *x*

Regression equation y on x is used to describe the variation in the value of y for a given value of x.

2) Regression equation *x* on y

Regression equation x on y is used to describe the variation in the value of x for a given value of y.

Least Square Method

The average relationship between the dependent variable and one independent variable is generally explained in the form of an equation of a line. A line fitted on the basis of principle of least square method is called line of best fit.

The Least Square Method is based on the following principle

1) The sum of deviations of individual observation from the fitted line of regression on both sides are zero

   i.e) $\sum (y - y_e) = 0$ and $\sum (x - x_e) = 0$

2) The sum of squares of deviations of individual observation from the refitted line of regression on both sides are minimum

   i.e) $\sum(y - y_e)^2 = \text{minimum}$

   $\sum(x - x_e)^2 = \text{minimum}$

   x_e and y_e values are obtained from regression.

Merits

1) More scientific and accurate because it is based on mathematical method.
2) Free from personal bias.
3) It can be used for both linear and non-linear regression.
4) It is used to estimate all the variables of one variable for the given values of other variables.

Demerits

1) It is not a flexible method.
2) The calculation is difficult
3) It is time consuming method.

Procedure to fit a regression line for the given data (observed data)

Let the regression equation y on x be $y = a + bx$ for a given set of N observations $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$. In this equation a and b are the two unknown concepts. Least square method is applied to find the values of a and b.

The least square method gives two normal equations

$\sum y = na + b\sum x$ …. (1)

$\sum xy = a\sum x + b\sum x^2$ … (2)

Solving the above equations we get

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

$$a = \bar{y} - b\bar{x}$$

Let the regression equation x on y be

$x = a + by$

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

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

Solving the above equations we get

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum y^2 - (\sum y)^2}$$

$$b = \bar{x} - b\bar{y}$$

Model I

In this model the values of the variable x and y are given to fit the regression lines (required).

Problem

Obtain the two regression lines in the following data.

x	6	2	10	4	8
y	9	11	5	8	7

Solution

x	y	xy	x^2	y^2
6	9	54	36	81
2	11	22	4	121
10	5	50	100	25
4	8	32	16	64
8	7	56	64	49
$\sum x = 30$	$\sum y = 40$	$\sum x^2 = 214$	$\sum x^2 = 220$	$\sum y^2 = 340$

Let the regression equation y on x be

$$y = a + bx$$

Applying Least Square method we get normal equations as

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

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

Solving this we get

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

$$= \frac{(5)(214) - (30)(40)}{(5)(220) - (30)^2}$$

$$= \frac{1070 - 1200}{1100 - 900}$$

$$= -\frac{130}{200}$$

$$b = -0.65$$

$$a = \bar{x} - b\bar{y}$$

$$a = \frac{\sum x}{n} - (b)\frac{\sum y}{n}$$

$$= \frac{30}{5} - (-0.65)\frac{40}{5}$$

$$= 6 + (0.65)(8)$$

$$= 6 + 5.2$$

$$= 11.2$$

The regression line y on x is $y = 11.2 - 0.65x$

Let the regression equation x on y be

$$x = a + by$$

Applying the least square method we get normal equations as

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

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

Solving this we get

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum y^2 - (\sum y)^2}$$

$$= \frac{(5)(214)-(30)(40)}{(5)(340)-(40)^2}$$

$$= \frac{1070-1200}{1700-1600}$$

$$= -\frac{130}{100}$$

b = - 1.3

a = $\bar{x} - b\bar{y}$

= 6 - (-1.3) (8)

= 6 + 10.4

= 16.4

The regression equation x on y is x = 16.4 - 1.3y

Deviation Method (Short Cut Method)

Deviation from actual mean

When the deviations of x and y observations taken from their actual mean $\bar{x}$ *and* $\bar{y}$, then the regression equations are

i) Regression equation y on x is

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

where

$$b = r\frac{\sigma y}{\sigma x} \ (or) \ b = \frac{\sum xy}{\sum y^2}$$

ii) The regression equation x on y is

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

where

$$b = r\frac{\sigma x}{\sigma y} \quad (or) \ b = \frac{\sum xy}{\sum x^2}$$

(or)

$$x-\bar{x} = r.\frac{\sigma x}{\sigma y}(y-\bar{y})$$

Deviation from Assumed Mean

When the deviations of x and y observations taken from their assumed mean A and B then the regression equations are

i) Regression equation y on x is

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

byx = Regression coefficient y on x

byx $\dfrac{n\sum dxdy-(\sum dx)(\sum dy)}{n\sum d^2x-(\sum dx)^2}$

ii) Regression equation x on y is

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

bxy = Regression co-efficient x on y

bxy = $\dfrac{n\sum dx\,dy-(\sum dx)(\sum dy)}{n\sum d^2y-(\sum dy)^2}$

Model - II

1) Compute the regression lines for $\bar{x}$ = 40, $\bar{y}=50, \sigma x=10, \sigma y=16, r=0.6$

i) Let the regression equation y on x be

$(y-\bar{y}) = b\text{yx}(x-\bar{x})$

$$\text{b}yx = r\frac{\sigma y}{\sigma x}$$

$$= 0.6\ \frac{16}{10}$$

$= (0.6)\ (1.6)$

$= 0.96$

Now

y - 50 = 0.96 (x - 40)

y - 50 = 0.96x - 38.4

y = 0.96x - 38.4 + 50

y = 0.96x + 11.6

y = 11.6 + 0.96 x

ii) Let the regression equation x on y be

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

$$\text{b}xy = r.\frac{\sigma x}{\sigma y}$$

$$= (0.6)\ \frac{10}{16}$$

$= (0.6)\ (0.625)$

Now

x - 40 = 0.375 (y - 50)

x - 40 = 0.375 y - 18.75

x = 0.375y - 18.75 + 40

x = 0.375 + 21.25

x = 21.25 + 0.375y

Model III

Deviation from Assumed mean

Problem

1) Fit a linear regression y = a + bx for the given data.

x	15	10	15	20	25	30	35
y	12	19	31	40	51	58	67

Solution

X	Y	dx = x - 20	dy = y - 40	dxdy	d^2x	d^2y
15	12	-5	-28	140	25	784
10	19	-10	-21	210	100	441
15	31	-5	-9	45	25	81
20	40	0	0	0	0	0
25	51	5	11	55	25	121
30	58	10	18	180	100	729
35	67	15	27	405	225	729
$\sum x = 150$	$\sum y = 278$	$\sum dx = 10$	$\sum dy = -2$	$\sum dxdy = 1035$	$\sum dx^2 = 500$	$\sum dy^2 = 2480$

i) Let the regression equation y on x be

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

$$byx = \frac{n\sum dxdy - \sum dx \sum dy}{n\sum d^2x - (\sum dx)^2}$$

$$= \frac{(7)(1035) - (10)(-2)}{(7)(500) - (10)^2}$$

$$= \frac{7245 + 20}{3500 - 100}$$

$$= \frac{7265}{3400}$$

$$byx = 2.13$$

The regression equation y on x is

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

$$\bar{x} = \frac{\sum x}{n}$$

$$= \frac{150}{7}$$

$$= 21.\ 42$$

$$\bar{y} = \frac{\sum y}{n}$$

$$= \frac{278}{7}$$

$$\bar{y} = 39.71$$

(y - 39.71) = 2.13 (x - 21.42)

y - 39.71 = 2.13 x - (2.13) (21.42)

y - 39.71 = 2.13x - 45.62

y = 2.13x - 45.62 + 39.71

y = -5.91 + 2.13x

ii) Let the regression equation x on y be

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

$$bxy = \frac{n\sum dx\,dy - \sum dx\,dy}{n\sum d^2y - (\sum dy)^2}$$

$$= \frac{(7)(1035)-(10)(-2)}{(7)(2480)-(278)^2}$$

$$= \frac{7245+20}{17360-77284}$$

$$= \frac{7265}{-59924}$$

$bxy = -0.12$

The regression equation of x on y be

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

$\bar{x} = 21.42$, $\bar{y} = 39.71$

$(x - 21.42) = -0.12(y - 39.71)$

$x - 21.42 = -0.12\,y + (0.12)\,(39.71)$

$x - 21.42 = -0.12\,y + 4.76$

$x = -0.12y + 4.76 + 21.42$

$x = 26.18 - 0.12y$

Properties of regression co-efficient

i) The geometric mean of regression co-efficient is correlation coefficient

   a. i.e) $\sqrt{(byx)(bxy)} = r$

ii) The arithmetic mean of regression co-efficient is greater than or equal to correlation coefficient i.e) $\frac{byx+bxy}{2} \geq r$

iii) Both the regression coefficients have same sign.

iv) The regression co-efficient are independent origin and not independent of scale.

v) If one regression co-efficient is greater than one, then other regression co-efficient will be less than one. If $byx > 1$ then $bxy < 1$.

vi) If the correlation co-efficient r is ±1 then the regression line y on x and x on y coincides each other.

vii) If r = o regression lines are perpendicular (⊥r) to each other.

viii) The co-ordinates of points of intersection of the two regression lines give the mean of x- series and y-series.

Model IV

Under this the value of the observation x and y are not given. The equation of the regression lines y on x and x on y will be given to find out $\bar{x}$, $\bar{y}$ and r.

Problem

Calculate $\bar{x}$, $\bar{y}$, r, byx, bxy when the regression lines are $2y - x = 50$ and $-3y + 2x=10$.

Solution

Consider $2y - x = 50$ as regression equation y on x.

$2y = 50 + x$

$y = \frac{50}{2} + \frac{x}{2}$

$y = 25 + 0.5x$

$byx = 0.5$

Consider - 3y + 2*x* = 10 as regression equation *x* on y.

$2x = 10 + 3y$

$x = \frac{10}{2} + \frac{3}{2}y$

$x = 5 + 1.5y$

$bxy = 1.5$

Now

$r = \sqrt{(byx)(bxy)}$

$= \sqrt{(0.5)(1.5)}$

$= \sqrt{(0.75)}$

$r = 0.86$

$2y - x = 50 \rightarrow (1)$

$-3y + 2x = 10 \rightarrow (2)$

$$equation\ (1) \times 2 \Rightarrow 4y - 2x = 100$$

$$equation (2) \Rightarrow \frac{-3y + 2x = 10}{y + 0 = 110}$$

$y = 110$

put y = 110 in equation (2)

$-3(110) + 2x = 10$

$-330 + 2x = 10$

$2x = 10 + 330$

$2x = 340$

$x = 340/2$

$byx = 0.5$, $bxy = 1.5$, $r = 0.86$, $\bar{x} = 170$, $\bar{y} = 110$

Standard Error of Estimate

It is a measure of reliability of regression equation (estimating equation). It measures the variability of observed values of dependent variable and the estimated values of the independent variable using regression equation

$$S.E_{yx} = \sqrt{\frac{(y - y_e)^2}{n}},$$

$$S.E_{xy} = \sqrt{\frac{(x - x_e)^2}{n}}$$

y - observed values

e - estimated values

Problem

Given are the following data

X	6	2	10	4	8
Y	9	11	5	8	7

Find the regression lines and standard errors of the estimates (S_{yx} and S_{xy})

Solution

x	y	x_e	y_e	$(y - y_e)^2$	$(x - x_e)^2$
6	9	4.7	8.0	1.00	1.69
2	11	2.1	10.6	0.16	0.01
10	5	9.9	5.4	0.16	0.01
4	8	6.0	9.3	1.69	4.00
8	7	7.3	6.7	0.09	0.49
$\sum x = 30$	$\sum y = 40$	$\sum x_e = 30$	$\sum y_e = 40$	$\sum(y - y_e)^2 = 3.1$	$\sum(x - x_e)^2 = 6.2$

Regression Equation of y on x is y = a + b*x* we have two normal equations

$\sum y = na + b\sum x$ (or) $40 = 5a + 30b$

and $\sum xy = a\sum x + b\sum x^2$ (or) $214 = 30a + 220b$

solving above two equations, we get

a = 11.9 and b = - 0.65

Hence regression equation of y on *x* is

$y = 11.9 - 0.65x$

Regression Equation of *x* on y is

$x = c + dy$

The two normal equations are

$\sum x = nc + d\sum y$ (or) $30 = 5c + 40d$

and $\sum xy = n\sum y + d\sum y^2$ (or) $214 = 40c + 340d$

Solving above two equations, we get

c = 16.4 and d = - 1.3

Hence the regression equation of x on y is

$x = 16.4 - 1.3y$

x_e can be obtained through Regression

$x = 16.4 - 1.3y$

y	9	11	5	8	7
x	4.7	2.1	9.9	6	7.3

y_e can be obtained through Regression line

$y = 11.9 - 0.65x$

x	6	2	10	4	8
y	8	1.6	5.4	9.3	6.7

$$S_{yx} = \sqrt{\frac{\sum(y - y_e)^2}{n}}$$

$= \sqrt{\frac{3.10}{5}}$

$= \sqrt{0.62}$

$=$ 0.787

$S_{xy} = \sqrt{\frac{\sum(x - x_e)^2}{n}}$

$= \sqrt{\frac{6.20}{5}}$

$= \sqrt{1.24}$

$=$ 1.114

Multiple Regression

Multiple regression allows the use of two or more independent variable. Thus it uses more information to estimate the dependent variable y. The technique can be used to fit non-linear as well as linear relations. It can be manipulated to use qualitative variables.

If x_1, x_2 are independent variables which explain y we have the regression of y on x_1, x_2 given by

$y = a + b_1x_1 + b_2x_2$ where a, b_1, b_2 can be estimated by the principle of least squares

For this, solve the normal equation

$\sum y = na + b_1\sum x_1 + b_2\sum x_2$

$\sum x_1y = a\sum x_1 + b_1\sum x_1^2 + b_2 \sum x_1x_2$

$\sum x_2y = a\sum x_2 + b_1\sum x_1x_2 + b_2\sum x_2^2$

Problem

1) Fit a regression equation y on x_1 and x_2 from the following data.

x_1	4	7	9	12
x_2	1	2	5	8
y	2	12	17	20

Solution

x_1	x_2	**Y**	$x_1 x_2$	x_1^2	x_2^2	x_1y	x_2y
4	1	2	4	16	1	8	2
7	2	12	14	49	4	84	24
9	5	17	45	81	25	153	85
12	8	20	96	144	64	240	160
$\sum x_1 = 32$	$\sum x_2 = 16$	$\sum y = 51$	$\sum x_1x_2 = 159$	$\sum x_2^2 = 290$	$\sum x_2^2 = 94$	$\sum x_1y = 485$	$\sum x_2y = 271$

Let the regression equation y on x_1, x_2 be

$y = a + b_1x + b_2x_2$

The normal equations are

$\sum y = na + b_1\sum x_1 + b_2\sum x_2$

$\sum x_1y = a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2$

$\sum x_2y = a\sum x_2 + b_1\sum x_1x_2 + b_2\sum x_2^2$

Now

$4a + 32b_1 + 16b_2 = 51$ …. (1)

$32a + 290b_1 + 159b_2 = 485$ …. (2)

$16a + 159b_1 + 94b_2 = 271$ …. (3)

Consider equation (1) and (3)

$4a + 32b_1 + 16b_2 = 51$ …. (1)

$16a + 159\ b_1 + 94b_2 = 271$ …. (3)

Multiply equation (1) by 4

$$16a + 128b_1 + 64b_2 = 204$$

$$16a + 159b_1 + 94b_2 = 271 \quad \ldots. (4)$$

$$0 - 31b_1 - 30b_2 = -67$$

Consider equation (2) and (3)

$32a + 290b_1 + 159\ b_2 = 485$ …. (2)

$16a + 159b_1 + 94\ b_2 = 271$ …. (3)

Multiply equation (3) by 2

$$32a + 290b_1 + 159b_2 = 485$$

$$32a + 318b_1 + 188b_2 = 542 \quad \ldots. (5)$$

$$0 - 28b_1 - 29b_2 = -57$$

$- 31b_1 - 30b_2 = - 67$ …. (4)

$- 28b_1 - 29b_2 = -57$ …. (5)

Multiply equation (4) by 28 and equation (5) by 31

$$-868b_1 - 840b_2 = -1876$$

$$-868b_1 - 899b_2 = -1767$$

$$0 + 59b_2 = -109$$

$$b_2 = -\frac{109}{59}$$

$$b_2 = - 1.84$$

put $b_2 = 1.84$ in equation (4)

$- 31b_1 - 30 (-1.84) = - 67$

$- 31b_1 + 55.2 = - 67$

$- 31b_1 = - 67 - 55.2$

$$b_1 = \frac{-122.2}{-31}$$

$$b_1 = 3.94$$

Put $b_1 = 3.94$ and $b_2 = -1.84$ in equation (1)

$4a + 32 (3.94) + 16(-1.84) = 51$

$4a + 126.14 + (-29.44) = 51$

$4a + 96.7 = 51$

$4a = 51 - 96.7$

$4a = - 45.7$

$$a = \frac{-45.7}{4}$$

$$a = - 11.42$$

Verification

$4a + 32b_1 + 16b_2 = 51$

$4(- 11.42) + 32(3.94) + 16(-1.84) = 51$

$- 45.68 + 126.\ 08 + (- 29.44) = 51$

$- 75.12 + 126.08 = 51$

verified.

Multiple correlation, partial correlation

The correlation between a variable x_1 with the group of variables x_2, x_3 is measured by the coefficient of multiple correlation, written $R_{1.23}$

Variable 1 corresponds to the dependent variable; 2, 3 corresponds to the independent variables.

If r_{12} stands for the correlation between x_1, x_2 (same as r_{21})

r_{23} between x_2, x_3 (same as r_{32})

r_{31} between x_3, x_1 (same as r_{13})

We can show that the multiple correlation coefficient

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - r_{12} r_{23} r_{31}}{1-r_{23}^2}} \quad (0 \le R_{1.23} \le 1)$$

We can similarly define $R_{2.31}$, $R_{3.12}$. In the case of n variables, the multiple correlation between x_1 and x_2, x_3, … x_{n-1} can be written

$R_{1.23 \ldots n-1}$

The correlation between x_1, x_2 keeping x_3 constant is expressed by the partial correlation coefficient.

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}} \quad (-1 \le r_{12.3} \le 1)$$

$$r_{23.1} = \frac{r_{23} - r_{12} r_{13}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{13}^2}} \quad \textit{and so on } (r_{21} = r_{12}, r_{31} = r_{13})$$

The partial correlation between x_1, x_2 in the presence of variables x_3, x_4, …., x_n is written $r_{12.34 \ldots n}$

Problem

1) Find $R_{1.23}$ and $r_{12.3}$ given $r_{12} = .2$, $r_{13} = .3$, $r_{23} = .4$

$$R_{1.23} = \sqrt{\frac{.04 + .09 - 2 \times .2 \times .3 \times .4}{1-.16}} = 0.31$$

$$r_{12.3} = \frac{0.2 - 0.3 \times 0.4}{\sqrt{1-.09}\sqrt{1-.16}} = 0.092$$

Methods of estimation of non - linear equation

Exponential

We use mathematical tools for decision problems in which the first requirement is to define all significant interactions (or) relationships among primary factor relevant to the problems. These relationships usually are stated in the forms of an equation or a set of inequations.

A function having a variable base and a constant exponent namely $y = x, x^3, x^9$ etc… is called a power function. While a function having a constant base and variable exponents is called Exponential function

For example,

$y = a^x$ $(a > 0)$ is an exponential function where 'a' is the base and x is a real number exponent.

Characteristics of Exponential function

1) An exponential function $y = a^x$ is defined only for positive values of a and a^x is always positive.
2) The domain of the exponential function $y = a^x$ is R and range is $(0, \infty)$
3) If $a > 0$ and $a \neq 1$, then the function $y = a^x$ is into R^+
4) If $x > z$ and $a > 1$, then $a^x > a^z$.
5) If $0 < a < 1$ then $x > z$ or $a^z < a^x$. In other words for $a > 1$, $y = a^x$ is a decreasing function of x.
6) Of a = a then a^x, $x \in R$ and thus the function $y = a^x$ becomes a constant function.
7) If $a > 0$, $a \neq 1$, then $y = a^x$ is a continuous function with domain R and range $(0, \infty)$
8) The graphs of $y = a^x$ and $y = \left(1/a\right)^x = a - x$ are the reflection of each other about the y axis.

Parabolic

The form of function $y = f(x)$ (or) its estimate for any given value of x can be obtained by fitting a polynomial curve to the given set of observations provided the values of x are at equal intervals. the method is based on the fundamental theorem of algebra namely, one and only one polynomial curve of degree less than or equal to n passes through a given set of distinct points. Thus if we are given equidistant arguments and entries then we can represent the function $y = f(x)$ by a polynomial of n^{th} degree,

$y = f(x) = a_0x^n + a_1x^{n-1} + \ldots\ a_{n-1}x + a_n$

where $a_0, a_1, \ldots a_n$ we get the required form of function $y = f(x)$

UNIT - IV

PROBABILITY

Probability refers to the chance of happening (or) occurring (or) not happening of an event with reference to a series of events in an experiment.

The probability of a given event is an expression of the likelihood of occurrence of the event. Furthermore, a probability is a number which ranges from zero, for an event which cannot occur, to one for an event which is certain to occur.

As with basic concepts in any scientific discipline, leading authorities, disagree on the definition of probability. Three major schools of thought namely, the 'classical', 'relative', 'frequency' and subjective, theories are available.

The Two approaches to probability are (1) Classical approach (2) Empirical approach.

1) Classical approach

It is also called as a-priority definition of probability. The reason is in this method the result can be stated in advance without conducting any experiment. To describe the probability of an event deductive logic is used.

Definition

The probability of an event is the ratio of no.of favorable outcome. To happen the event A to total possible outcomes of the experiment.

$$P(A) = \frac{no.of.\ favourable\ outcomes\ of\ eventA}{Total\ possible\ outcome\ of\ the\ \exp eriment}$$

Assumption

The classical definition of probability based on the following assumptions. The events of the experiment are

1) (MEE) Mutually exclusive events.
2) (ELE) Equally likely events.
3) (CEE) Collectively exhaustive events.

2) Empirical (or) Relative frequency/ Statistical definition of probability

It is based on past experience of data or evidence of experiment and observation. In this an experiment is repeated large number of times to find the probability of event in a single trial. Inductive logic is used to derive the probability measurement of an event.

Definition

The probability of event is defined as a limiting value of the ratio of number of times the event happens to the total number of trials of the experiments as the no.of trial increases.

Theorem's of probability

The important theorem's of probability are (1) addition Theorem (2) Multiplication Theorem

(i) Addition Theorem of probability

It states that if A and B are any two mutually exclusive events with probability P(A) and P(B) respectively, then the probability of either A or B is equal to sum of their individual problems.

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

proof

Let 'N' be the total possible outcomes of an experiment. Let M_1 and M_2 be the no.of favorable outcomes to happen the event A and B respectively.

$$P(A) = \frac{M_1}{n}; \quad P(B) = \frac{M_2}{n}$$

Since A and B are mutually exclusive events, there are exactly M_1+M_2 outcomes favourable to happen the event A or B.

$$\text{Now,} \quad P(A \text{ or } B) = \frac{M_1 + M_2}{n}$$

$$= \frac{M_1}{n} + \frac{M_2}{n}$$

$$P(A \text{ or } B) = P(A) \text{ or } P(B)$$

Hence proved the addition theorem.

Generally if A_1, A_2, A_3 …. A_n are mutually exclusive events, then $P(A_1$ or A_2 or A_3 or …. $A_n)$ = $P(A_1) + P(A_2) + P(A_3) + …. + P(A_n)$.

Example

What is the probability of geting a king or a queen when a card is drawn at random,

Solution

Total possible outcomes to draw a card	=	52.
No.of favourable outcome to get a king	=	4
No.of favourable outcome to get a queen	=	4
No.of favourable to get a king (or) queen	=	4 + 4
	=	8

by Addition theorem

P(A King or Queen)	=	8/52
	=	2/13

(ii) Multiplication Theorem of Probability

Statement

It states that if A and B are two independent events with probability P(A) and P(B) respectively, then the probability of both events A and B is equal to the product of their individual probabilities.

P(A and B) = P(A).P(B) (or)

if A and B are two independent events, then $P(A) \times P(B)$

Proof

Let N_1 be the total No.of possible outcomes and M_1 be the no.of favourable outcomes for the event A. Let N_2 be the total possible outcomes and M_2 be the no.of favourable outcomes for the event B.

$$P(A) = \frac{M_1}{n_1}$$

$$P(B) = \frac{M_2}{n_2}$$

Since the events A and B are independent events, each favourable outcomes of the event A can be combined with each favourable outcomes of B. Therefore total no.of

favourable outcomes for both the events A and B is M_1 and M_2. Similarly the total possible outcome of the event A and B is n_1 and n_2.
Now the probability of A and B is

$$P(A \text{ and } B) = \frac{M_1 M_2}{n_1 n_2}$$

$$= \frac{M_1}{n_1} \times \frac{M_2}{n_2}$$

$$= P(A).P(B)$$

Hence proved.

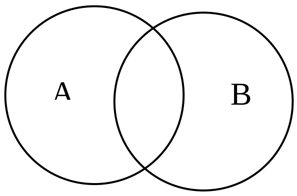
Problem

Prove that if A and B are any two not mutually exclusive events.

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

Proof

Given that A and B are not mutually exclusive.
Let set A represents the favourable outcomes of the event A.
Let set B represent the favourable outcomes of the event B.
Since A and B are not mutually exclusive events $A \cap B$ exists.



From the diagram,

n(A or B) = n(A) + n(B) - n(A and B) divided by n(s)

divided by n(s)

$$\frac{n(A or B)}{n(s)} = \frac{n(A)}{n(s)} + \frac{n(B)}{n(s)} - \frac{n(A and B)}{n(s)}$$

i.e) P(A or B) = P(A) + P(B) - P(A and B)

Hence proved.

Conditional Probability

In many situation, it is found that the occurrence of an event supplies information which affects the changes of occurrence (or non-occurrence) of some other event. This is true of almost all the business situations where knowledge about some outcome reduces (or increases) the uncertainty about some other. For example, if A denotes the event ‘Mr. X is promoted to a senior executive position and B denotes the event ‘Mr. X gets his MBA degree’ then the knowledge that ‘B has already occurred’ reduces the uncertainty about A’s occurrence. On the other hand, the occurrence of B increases the uncertainty about the occurrence of another event c_ ‘the rival of x (a non- MBA) gets the promotion.

The probability that is assigned to an event A when it is known that another event B has already occurred is called the ‘conditional probability of A given B’, which is denoted by P(A\B) and defined by

$$P(A \backslash B) = \frac{P(A \cap B)}{P(B)}, \text{ provided } P(B) > 0$$

Similarly, the conditional probability of event B when it is known that another event A has already occurred is defined as

$$P(B\backslash A) = \frac{P(A \cap B)}{P(A)}, \text{ provided } P(A) > 0$$

Problem

If two fair dice are tossed, what is the probability that the sum is seven if it is given that at least one face shows a 4?

Solution

Let A and B denote the events that 'sum is seven' & 'At least one face shows a 4'.

Then A = {(6, 1), (5, 2), (4, 3), (3, 4), (2, 5), (1,6)}

B = {(4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (2, 4), (3, 4), (5, 4), (6, 4)}

So that $A \cap B$ = {(4, 3), (3, 4)}

Thus $P(A \cap B)$ = 2/36 and P(B) = 11/36

The required probability is given by

$$P(A\backslash B) = \frac{P(A \cap B)}{P(B)} = \frac{2/36}{11/36} = \frac{2}{11}$$

Elementary concept of Random variable

Random variable

A variable whose value is determined by the outcome of a random experiment is called a random variable. It is a rule that assigns a numerical value to each outcome of an experiment.

Discrete Random variable

A random variable that may only take a finite or countable number of different value S is referred to as a 'discrete random variables'.

Continuous random variable

It is one that can assume the infinitely large number of values corresponding to the points on a line interval. Examples are weight, time, temperature, humidity, pressure, life of an electric bulb and so on which may take any value in a certain interval or collection of intervals.

Probability Mass Function (PMF)

Probability distributions of discrete random variables are called probability mass function. A P.M.F has the following properties.

If X is a discrete random variable with values x_1, x_2, …. x_n, then

$p(x_i) \geq 0$, for i = 1, 2, ….n

$p(x_1) + p(x_2) + \ldots = 1$

Problem

1) Obtain the probability distribution of *x*, the number of heads in three tosses of a coin C or simultaneous toss of free coins.

Solution

When three coins are tossed, the sample space S consists of $2^3 = 8$ sample points, as given below

S = {(H, T) × (H, T) × (H, T)} = { (HH, HT, THT) × (H, T)}

= {HHH, HTH, THH, TTH, HHT, HTT, THT}

Probability distribution of number of heads

No.of Heads X	Favourite sample points	Number of Favourable cases	Probability P(X = X) = P(X)
0	(TTT)	1	1/8
1	(TTH, HTT, HHT)	3	3/8
2	(HTH, THH, HHT)	3	3/8
3	(HHH)	1	1/8

2) A box contains number of defective bulbs drawn. Then X is a random variable which can take the values 0, 1, 2 and 3 with the associated probabilities as shown below

Solution

P(x = 0) = P(0) = p (no defective bulb)

$$= \frac{9c_3}{12c_3}$$

$$= \frac{9\times8\times7}{12\times11\times10}$$

$$= \frac{21}{55}$$

P(x = 1) = P(1) = P(one defective bulb)

$$= \frac{3c_1\times9c_2}{12c_3}$$

$$= \frac{3\times9\times8}{12\times11\times10}$$

$$= \frac{9}{55}$$

P(x = 2) = P(2) = P(Two defective bulbs)

$$= \frac{3c_2\times9c_1}{12c_3}$$

$$= \frac{3\times2\times9}{12\times11\times10}$$

$$= \frac{9}{220}$$

P(x = 3) = P(3) = P(three defective bulbs)

$$= \frac{3c_3}{12c_3}$$

$$= \frac{3\times2\times1}{12\times11\times10}$$

$$= \frac{1}{220}$$

Thus, the probability distribution of x is given by

Values of x	0	1	2	3
P(x)	21/55	9/55	27/220	1/220

Probability Density Function (p.d.f)

To assign the probability measure to a continuous random variable, a continuous probability distribution called 'probability density function' is used. This is defined as follows

A probability density function is any function f(x) of a continuous random variable x, where

i) f(x) ≥ 0 for all x, and
ii) The total area under the curve f(x), equal one.

The definition, characteristics, and application of probability density function may be formally expressed using integral calculus. The following definition of a probability density function is equivalent to that given above.

A probability density function is any function f(x) of a continuous random variable x defined over the range r, where

i) f(x) ≥ 0 for all x
ii) $\int_R f(x)\,dx = 1$

Probability for a continuous random variable is defined in terms of the area under a curve. The probability, P(a ≤ x ≤ b), that a continuous random variable x takes on values between two limits a and b is the proportion of total area between a and b and is given by

$$P(a \le x \le b) = \int_a^b f(x)\,dx$$

The expected value, E(x), and variance of σx^2 of a continuous random variable are defined as

$$E(x) = \int_R x\, f(x)\,dx \; and \; \text{var}(x) \text{ (or)}$$

$$\sigma x^2 = \int_R x^2 f(x)\,dx - [E(x)]$$

Theoretical Distribution

The frequency distributions in the form of data considered earlier refer to samples drawn from a population. Measures of central tendency and dispersion etc… are the main features of a frequency distribution but their knowledge is not sufficient if complete information about the distribution is needed. If we take a large number of observations of an actual distribution, the law of that distribution can be established. The variate valued follow some mathematical law which can be found.

It is possible to fit such a theoretical distribution to the given observation if its characteristics and laws are known. We may be also discover additional features of the observed values with its help.

In statistics we come across many distributions either experimentally or theoretically. The most important theoretical distributions are

i) The Binomial Distribution
ii) The Poisson Distribution
iii) The Normal Distribution

(i) The Binomial Distribution

A Binomial distribution is one which the frequencies are proportional to the successive terms of the binomial expansion. This Theoretical distribution is very useful in practice. It can be applied to a number of commonly occurring situations.

A discrete random variable x is set to follow a binomial distribution. It is probability mass function is given by the equation $P(x) = ncxp^{x}\, q^{n-x}$ where $x = 0, 1, 2, \ldots n$.

n - number of trials
x - number of success required,
p - probability success of an event in a trial
q = 1-p
n, p are parameters.

Assumptions

1) Each trials has 2 mutually exclusive events, success and failure.
2) The trials are independent
3) The no. of trials is finite and fixed.
4) The probability of success in each trial remains constant.

Properties

It is a discrete probability distribution. It is a two parameter. (i.e) $P(x) = n(x.\ p^{r}.\ q^{n-x})$

n = no. of trials
p = probability of success in a trial

The no. of trials 'n' is finite

The Mean $(\bar{x})$ of B.D. is "np"

The standard derivation (σ) of B.D is $\sqrt{npq}$ variance σ^2 = npq

Co-efficient of skewness (β_1) of B.D

$$(\beta_1) = \frac{(q-p)^2}{npq}$$

Co-efficient of kurtosis (β_2) of B.D

$$\beta_2 = \frac{3+1-6pq}{npq}$$

B.D is positively skew if P < 1/2; B.D is negatively skew if P > 1/2 and symmetrical if p = 1/2

The mean and more value of B.D increases as the value of N increases for a fixed "p".

If x_1 and x_2 are the two independent B.D variable. Then $x_1 + x_2$ is also Binomial variables.

Problem

A set of 7 coins are tossed 28 times with a binomial distribution to this experiment and calculate the expected frequencies and calculate the mean and standard deviation in Binomial distribution.

Solution

No of trials (n) = 7

No.of tossed (N) = 128

Let x denoted no. of heads required

p = probability of getting a head in a trial.

p = 1/2

q = 1 - p (p + q = 1)

= 1 - 1/2

$$nC_r = \frac{n!}{(n-r)!r!}$$

Value of 0! = 1

x	p(x)
0	$P(0) = 7C_0 (1/2)^0 (1/2)^7 = 1 \times 1 \times (1/2)^7 = (1)\ 1/128 = 1/128$
1	$P(1) = 7C_1 (1/2)^1 (1/2)^6 = (7) (1/2)^1 (1/2)^7 = (7)\ 1/128 = 7/128$
2	$P(2) = 7C_2 (1/2)^2 (1/2)^5 = (21) (1/2)^2 (1/2)^5 = (21)\ 1/128 = 21/128$
3	$P(3) = 7C_3 (1/2)^3 (1/2)^4 = (35) (1/2)^4 (1/2)^4 = (35)\ 1/128 = 35/128$
4	$P(4) = 7C_4 (1/2)^4 (1/2)^3 = (35) (1/2)^4 (1/2)^4 = (35)\ 1/128 = 35/128$
5	$P(5) = 7C_5 (1/2)^5 (1/2)^2 = (21) (1/2)^5 (1/2)^2 = (21)\ 1/128 = 21/128$
6	$P(6) = 7C_6 (1/2)^6 (1/2)^1 = (1) (1/2)^7 (1/2)^0 = (7)\ 1/128 = 7/128$
7	$P(7) = 7C_7 (1/2)^7 (1/2)^0 = (1) (1/2)^7 (1/2)^0 = (1)\ 1/128 = 1/128$

E(x) - (N) P(x)

E(0)	=	(128) 1/128	=	1
E(1)	=	(128) 7/128	=	7
E(2)	=	(128) 21/128	=	21
E(3)	=	(128) 35/128	=	35
E(4)	=	(128) 35/128	=	35
E(5)	=	(128) 21/128	=	21
E(6)	=	(128) 7/128	=	7
E(7)	=	(128) 1/128	=	1
		$\sum E(x)$	=	128

Mean of Binomial Distribution = $n \times p$

= (7) (1/2)

= 7/2

= 3.5

Standard Deviation = $\sqrt{npq}$

= $\sqrt{(7)(1/2)(1/2)}$

= $\sqrt{\frac{7}{4}}$

3) One half of the population of a town consists of smokers. 100 investigators each check 10 persons for smokers. How many investigators may be expected to find that 3 or less people are smokers?

Solution

Probability of anyone being a smoker = 1/2

Probability of 3 or less being found smokers is P($x \geq 3$) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)

= $10c_0(1/2)^{10} + 10c_1(1/2)^1(1/2)^9 + 10c_2\,(1/2)^2\,(1/2)^8 + 10c_3\,(1/2)^3\,(1/2)^7$

= $\left(1/2^{10}\right)(176)$

= $\frac{176}{1024}$

The number reporting 3 or less = (100) $\left(\frac{176}{1024}\right)$

= 17

The Poisson Distribution

When the probability of success of an event is very small i.e, the even is rare in a large number of trials over a time, the resulting distribution is known as Poisson Distribution. The number of deaths from horse licks in battles, the number of printing errors per page in a book, the number of blind children born in a city etc., provide examples of the Poisson distribution which has many uses in industry, demography and elsewhere.

Poisson Distribution was discard in 1837 by a French mathematician Simon Denies Poisson. It is used describe the behaviour of rare event. Therefore it is also called as law of improbable events.

The various probabilities of Poisson Distribution can be obtained, using the formula

$$P(x) = \frac{e^{-m}\,m^x}{\angle x}$$

where x = 0, 1, 2, 3, …

m = mean of Poisson Distribution

e = 2.7182

m = n × p and m > 0

Assumptions of Poisson Distribution

1) No. of trials is very large
   i.e) n ⇒ α
2) The probability of success for each trial is very small
   i.e) p ⇒ 0
3) m = np is a finite
   i.e) m = np

Problem

3% of electric bulbs given by a company are defective. Find the probability that in a sample of 100 bulbs exactly 5 are defective.

Solution

Sample size (n) = 100

defective item = 3%

Probability of a defective item (p) = 3/100 = 0.03

Mean of Poisson Distribution (m) = n × p

= (100) (0.03)

= 3

Let '*x*' denotes no. of defective items probability of exactly 5 defective items.

$= p(x = 5) \quad [\therefore p(x) = \frac{e^{-m} m^x}{\angle x}]$

$= \frac{e^{-3}.3^5}{\angle 5}$

$= \frac{(e^{-3})(243)}{5 \times 4 \times 3 \times 2 \times 1}$

$= \frac{(e^{-3})(81)}{40}$

$e^{-3} = 1/e^3$

$= \frac{1}{2.718^3}$

$= \frac{1}{20.08}$

$= 0.0498$

$= \frac{0.0498 \times 81}{40}$

$= \frac{4.032}{40}$

$= 0.1008$

Model II

Explain the procedure to fit a Poisson Distribution in a given set of data.

To fit a Poisson Distribution and to calculate an expected frequencies to a given set of data.

Step: 1

Calculate the mean of the given set of data $(\bar{x})$ observed distribution.

Step: 2

Assume that the mean of the given set of data is equal to mean of the required Poisson distribution i.e) $(\bar{x}) = m$

Step: 3

Compute the probability of each value of the random variable *x*.

Step: 4

Calculate the expected frequency of each value of *x* by using the formula

$E(x) = N\, p(x)$

where N - Total frequency of given distribution.

Problem

Fit a Poisson Distribution and calculate Theoretical (expected) frequencies.

X	0	1	2	3	4
F	21	18	7	3	1

Solution

x	f	fx	p(x)						
0	21	0	p(0)	=	$\frac{e^{-0.9}.(0.9)^0}{\angle 0}$	=	$\frac{0.4065\times 1}{1}$	=	0.4065
1	18	18	p(1)	=	$\frac{e^{-0.9}.(0.9)^1}{\angle 1}$	=	$\frac{0.4065\times 0.9}{1}$	=	0.3658
2	7	14	p(2)	=	$\frac{e^{-0.9}.(0.9)^2}{\angle 2}$	=	$\frac{0.4065\times 0.81}{2}$	=	0.1646
3	3	9	p(3)	=	$\frac{e^{-0.9}.(0.9)^3}{\angle 3}$	=	$\frac{0.4065\times 0.729}{1\times 2\times 3}$	=	0.0493
4	1	4	p(4)	=	$\frac{e^{-0.9}.(0.9)^4}{\angle 4}$	=	$\frac{0.4065\times 0.6561}{1\times 2\times 3\times 4}$	=	0.0111

$\sum f = 50$ $\sum fx = 45$

$$\text{Mean } (\bar{x}) = \frac{\sum fx}{\sum f}$$

$$= \frac{45}{50}$$

$$= 0.9$$

Assumed mean of Poisson Distribution = $\bar{x}$

i.e m = 0.9

$E(x) = N.P(x)$

E(0) = 50 × 0.4065 = 20

E(1) = 50 × 0. 3658 = 18

E(2) = 50 × 0.1646 = 8

E(3) = 50 × 0.0493 = 2

E(4) = 50 × 0.0111 = 0.5

Properties of Poisson Distribution

1) It is a discrete Probability Distribution in single parameter m (mean of Poisson Distribution) where m = np
2) The no. of trials (n) of P.D is very large and (n $\Rightarrow \alpha$)
3) The probability of success in a trial is very small (p $\Rightarrow$ 0)
4) The mean of the Poisson Distribution is m and m > 0
5) The standard deviation of Poisson Distribution (σ)

   (i.e) P.D (σ) = $\sqrt{m}$ and

Variance(σ^2) = m

6) The coefficient of skewness of P.D(β_1) = 1/m
7) The co-efficient of kurtosis in P.D (β_2) = 3 + 1/m
8) In a P.D if m is an integer then P.D has two modes.
9) Poisson distribution is positively skew as m value increase as skewness decreases.

The Normal Distribution

The normal distribution provides a standard pattern with which other distributions may be compared and which provides a common feature of interest among economists, sociologists, psychologists etc.

It has been noticed that empirical distributions of various types of observations in natural and social sciences are often very close to the normal distribution. In statistical analysis the distributions of observations is frequently assumed to be approximately normal. In statistical estimation and testing of hypothesis the normal distribution plays an important role.

The normal distribution is defined by the equation π

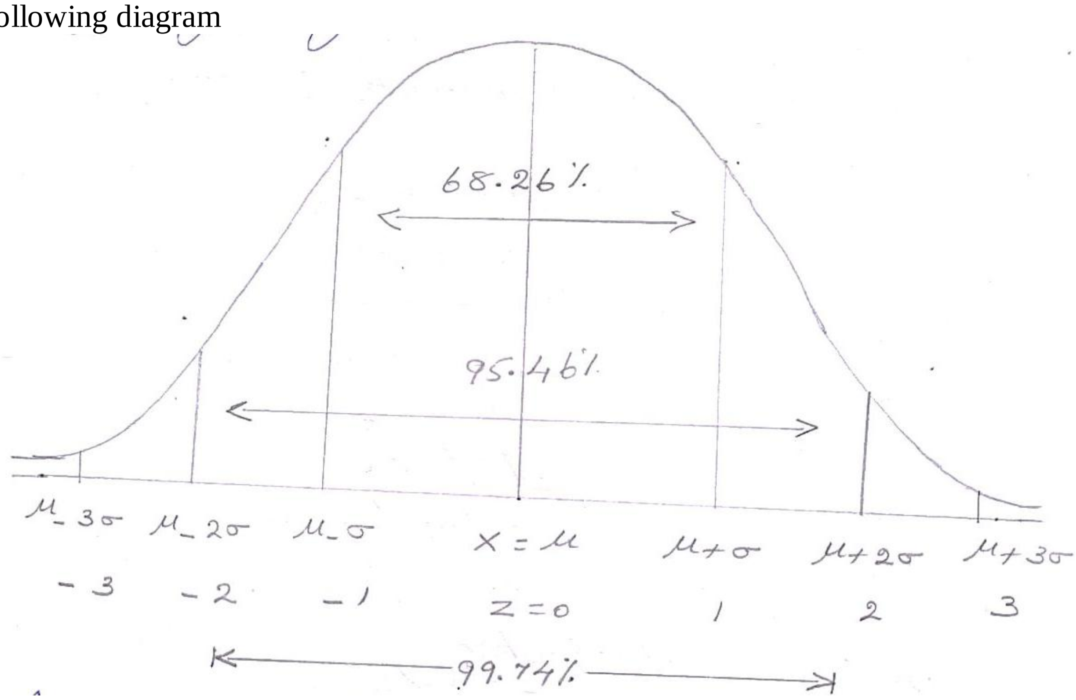
$$y = \frac{1}{\sigma\sqrt{2\Pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}$$

where μ is the mean value of x in the population and σ is the standard deviation of x in the population μ and σ are the parameters of the equation and completely define the distribution.

y is the height of the curve at any particular value of *x*

e(= 2.7183) and π (= 3.1416) are constants.

The percentage distribution of area under standard normal curve is broadly shown in the following diagram



Areas under standard Normal curve

Distance from the Mean ordinates in terms of $\pm\sigma$	**Area under the curve**
z = ± 0.6745	50% = 0.50
z = ± 1.00	68.26% = 0.6826
z = ± 1.96	95% = 0.95
z = ± 2.0	95.44% = 0.9544
z = ± 2.58	99% = 0.99
z = ± 3.0	99.73% = 0.9973

Properties of Normal Distribution

1) It is a continuous probability distribution with two parameters mean μ and standard deviation σ.
2) It is symmetrical about mean (μ) x = mean
3) A normal distribution mean = median = mode
4) It has only one mode. There is unimodal distribution.
5) In normal distribution, mean deviation = 4/5 σ, Quartile deviation = 2/3 σ.
6) The co-efficient of skewness of (β_1) of normal deviation = 2/3σ
7) Co-efficient of kurtosis (β_2) of the normal distribution is three.
8) The first and third quartiles (q_1, q_3) are equal distance from median.
9) The various descriptive measures of the normal distribution are: mean = $\bar{x}$ or μ, standard deviation = σ, variance or $\mu_2 = \sigma^2$, third central moment, $\mu_3 = 0$.
10) The height of the normal curve is the maximum at x = e and it's value is $\frac{1}{\sigma\sqrt{2\Pi}}$
11) 68.26% of the area lies between $x = \mu \pm \sigma$, 95.46% of the area lies between $x = \mu \pm 2\sigma$, 99.74% of the area lies between $x = \mu \pm 3\sigma$
12) The normal curve is concare at x = mean and curve at $x = \mu \pm 3\sigma$
13) The points of inflexion (the points at which the curve changes its direction) are each at a distance of one standard deviation from the mean.
14) The two tails of the normal curve extend to infinity in both directions and theoretically never touch the horizontal axis.

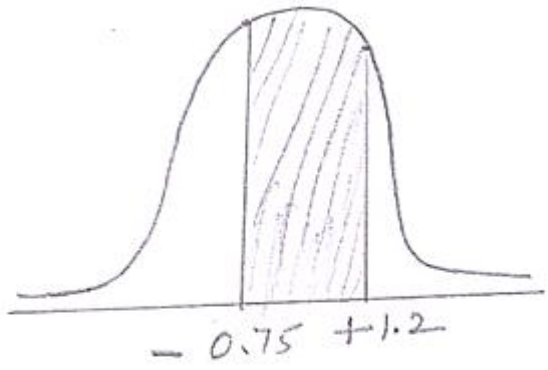
Problem

For the problems given, the normal curve table should be used. In each case if is preferable to draw the diagram

1) Find the area lies between z = -0.75 and z = 1.2

Solution

The area between z = - 0.75 and z = 1.2 is given by the sum of (the area between z = -0.75 and z = 0) and (the area between z = 0 and z = 1.2)



i.e) p(-0.75 ≤ z ≤ 0) + p(0 ≤ z ≤ 1.2)

= p(0 ≤ z ≤ 0.75) + p(0 ≤ z ≤ 1.2)

= 0.2734 + 0.3849

= 0.6583

2) Find the area between z = 0.46 and z = 1.28

Solution

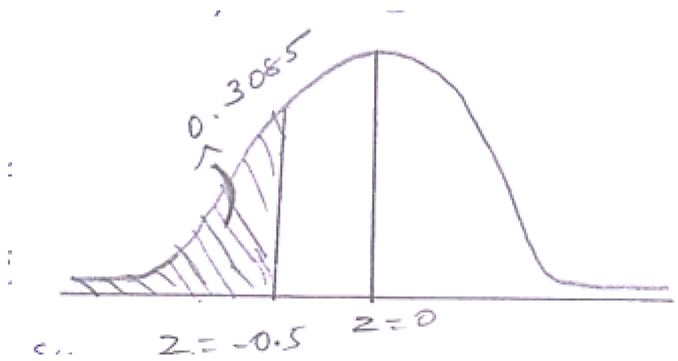
The area between z = 0 and z = 1.28 = (Area between z = 0 and z = 1.28) - Area between z = 0 and z = 0.46)

= 0.3997 - 0.1772

= 0.2225

3) The marks obtained in a certain examination follow the normal distribution with mean 45 and standard deviation 10. If 1,000 students appeared at the examination, calculate the number of students scoring i) Less then 40 marks, ii) more than 60 marks and iii) between 40 and 50 marks.

Solution



$Z = \frac{x-45}{10}$

i) when x = 40, z = $\frac{40-45}{10} = -0.5$

Thus, p(x ≤ 40) = p (z ≤ - 0.5)

= p (z ≥ 0.5) (Due to symmetry)

= 0.5 - p (0 ≤ z ≤ 0.5)

= 0.5 - 0.1915

= 0.3085

Hence the number of students scoring less than 40 marks are 1000 × 0.3085 = 309.

ii) p(x ≥ 60)= p (z ≥ 1.5) [∴ when x = 60, z =(60 - 45/10) = 1.5]

= 0.5- p(0 ≤ z ≤ 1.5)

= 0.5 - 0.4332

= 0.0668

when x = 40, z = $\frac{40-45}{10}$ = 0.5

when x = 50, z = $\frac{50-45}{10}$ = 0.5

Thus, p(40 ≤ x ≤ 50) = p(-0.5 ≤ z ≤ 0.5)

= p(0 ≤ z ≤ 0.5)

= p(0 ≤ z ≤ 0.5)

= 2 × 0.1915

= 0.3820

Hence the number of students scoring marks between 40 and 50 are 1000 × 0.3830 = 383

4) In a sample of 1,000 items, the mean weight and standard deviation are 50 and 10 kilograms respectively. Assuming the distribution to be normal, find the number of items weighing between 40 and 70 kilograms.

Let the random variable x denote the weight (in kg) of items. Then x is normally distributed with mean, μ = 50 and s.d, σ = 10

Standard normal variate corresponding to 40,

$z = \frac{40-50}{10} = -1$

Standard normal variate corresponding to 70,

$$z = \frac{70 - 50}{10} = 2$$

Number of items weighing between 40 and 70kg are 1,000 × 0.8185 = 819.

Central Limit Theorem

A central limit theorem is the most important theorem explains the relationship between the shape of the population distribution and shape of the sampling distribution of mean.

Statement of Central limit theorem

It states that the distribution of mean of random samples taken from a population having mean μ and finite variance σ^2 approaches the normal distribution with mean μ and variance σ^2/n as the sample size increases to infinity

Significance of Central limit theorem

1) It permits the use of sample statistic to make generalization about the population parameter on the basis of the sample we get from the population.
2) It points out that many statistical constant tend to be normally distributed in the sample size becomes sufficiently large.

Conditions or Assumptions of Central limit theorem

1) The variables must be independent
2) The variables should have common mean and standard deviation.
3) The size of the sample is enlarged

UNIT V

SAMPLING THEORY

Sampling theory is a study of relation existing between population and the samples drawn from that population.

Sampling

The process of obtaining information about the population by examining a part of the population is called sampling.

It is a process or technique of selecting a suitable sampling with the objective of getting information above the entire population.

Uses / advantages / need for sampling

1. A sample study is more economic and time saving compare to census study.
2. Since the sample size is limited one, both the execution of field work and analysis of research can be carried out quickly.
3. More detailed information is connected in sample survey because the magnitude of operations involved in sample survey is solved.
4. Sampling is the only way then the population size is infinite (very large). Because census method will not be used in such causes.
5. More accurate measurement can be obtained better than census survey.

Principles of sampling

Sampling is based on two principles.

1. Law of statistical regularity
2. Law of inertia of large numbers.

1) Law of statistical regularity

This law is derived from the mathematical theory of probability. It points out that if a large sample is taken at random from a population, the sample is almost likely to possess all the same characteristics of that population this principle directs the important point of the disability of choosing a sample at random.

2) Law of inertia of large numbers

This law states that large aggregates are more stable than small ones. This means that total change is likely to be very small when the large no of items are taken in a sample. This law simply points out that other things being equal as the sample size increases. The sample gives more accurate and reliable results about the population.

Methods of sampling (or) sample design

The various methods of sampling is called design – It can be classified in to

1. Random sampling
2. Non – Random sampling

Random sampling

A sample is selected in a way that each unit of the population has an equal chance of being selected is called random sampling.

The various types of random sampling are

Simple Random Sampling

1. This is a method of sampling in which each unit of the population has exactly the same chance of being included in the sample.
2. Random sampling is of two types: - Random sampling with replacement – the same unit of population may occur more than once and Random sampling without replacement-the same population unit cannot come more than once in the same sample.
3. Random sampling can be done with the help of random numbers or using lottery method.
4. It is an easiest technique of sampling and involves less work and time.

Stratified sampling

1. In this method of sampling, the entire heterogeneous population is divided into stares (or groups) consisting of homogeneous units and than samples are drawn from every group by simple random sampling technique.
2. There is only one type of stratified random sampling.
3. Stratified sampling is done by dividing the heterogenous population into a number of homogeneous group which differ from one another but each group is homogeneous within itself. Then units are sampled at random from each of these stratum. The sample which is the aggregate of the sampled units of each stratum is called a stratified sample.
4. It is tedious than simple random sampling.

Systematic sampling

This method of sampling involves the selection of sample units at equal intervals after all the units in the population have been arranged in some order. The units in the population are serially numbered from the first k of these, a single unit chosen at random. This unit and every k^{th} unit there after constitute a systematic sample. To obtain a systematic sample of 100 villages out of 10,000, all the villages are numbered serially from the first 100 of these a village is selected at random suppose with the serial number 7, Then the villages with serial numbers 7, 107, 201, 307, ….. constitute a systematic sample.

Multi-stage sampling

This method of sampling refers to a sampling procedure which is carried out in several stages. Under this method, the population is first divided into large groups, called first-stage units. These first stage units are again divided into smaller units called second – stage units the second stage units into third stage units and so on until we reach the ultimate units. In a survey of rural debt in India, one may like to apply a sampling technique for selection of districts, then villages, then households, and so on.

Non Random Sampling

Deliberate, purposive or judgment sampling is a method under which items for the sampling are selected generally on certain pre-determined criteria. The fixation of criteria and deliberate choice of sampling units may bring in personal element and introduce bias. The selection of items would differ from person to person, at times by personal fancy and judgement of the individual determining the sample. This method is generally used and considered appropriate in small inquiries and researches by individuals, specially when they are familiar with almost all items of universe. Inferences drawn under this method are not amenable to statistical treatment.

Quota sampling is that method in which each person engaged in the primary collection of data is assigned a certain quota of investigations. The actual selection of items for the sample is left to the investigator's discretion. This method is convenient and is relatively inexpensive but this allows some bias to enter into the inquiry. Inferences drawn using this method are not amenable to statistical treatment in a formal way.

Convenience sampling is a method under which units in the sample are selected at the convenience of the investigator from an available source like that of telephone directory, automobile registration records, industrial or stock exchange directories, etc.

Concept of sampling

1) Population

The aggregate of all possible values that a random variable can assume is called population.

The aggregate of items above which information is desired. The population may be finite or infinite.

2) Sample

A part (or) portion of the population selected for direct examination (or) measurement is called sample.

The no. of items (observations) included in a population / sample is called population size or sample size.

3) Parameter

The measures obtain from population to describe the characteristic of the population is called parameter.

4) Statistic

The measure obtain from sample to describe the characteristic of the sample is called statistic.

Uses of Sampling Theory

1. It helps in estimating the unknown population parameters from the knowledge of sample statistics.
2. It is useful in determining the observed difference between two samples is significant or not.
3. It helps in making generalization about population on the basis of sample drawn from it also helps to find out the accuracy of the generalization of the population.

Statistical Inference

It retex to the process of selecting and using a sample statistic to make conclusion about the population parameter.

Sample study have no use unless it helps to get an idea about the population. The following problem arises when sample study are conducted.

1. How to generalize the result of sample to the population.
2. To what extend the generalization are valid.

Statistical inference is divided into

1. Theory of estimation
2. Test of hypothesis (Hypothesis Testing)

1. Theory of estimation

Theory of estimation was developed by R.A. Fisher in 1930. In estimation the value of the parameter is determined from a possible alternative sample statistics selected from the population.

Definition

A statistical methods used to evaluate the unknown population parameter from the known sample statistic is called estimation.

Estimation of parameter is essential whenever sample study is conducted.

Estimator

A sample statistic used to estimate the population parameter is called an estimator.

Type of Estimator

The estimator of the population parameter may be one single value (or) a range of value. There are two types of estimator.

1. Point Estimator

A single value of the estimator used find out the unknown values of the population parameter is called point estimate.

2. Internal Estimator

A Range of values of the estimator used to find out the unknown values of the population parameter is called Interval estimator. The interval estimator consist of two values with in that the values of the population parameter lie.

(i.e) Population parameter = Value of estimator ± Table value of test statistic × Standard error

Properties of good estimator

1) It is very difficult to give a correct detinition of a good estimator. There are 4 criteria by which one can estimate a quality of a statistics as a good estimator. The 4 criteria are;

a) Unbiasedness

It is a very important property of a good estimator, estimator is said to be an unbiased estimator. If the mean of all the statistic computed from all possible sample of same size taken from a population is equal to population parameter.

b) Efficiency

Another important desirable property of a good estimator is efficiency.

The efficiency of the estimator is measured in terms of size of the standard error. An estimator as the smallest standard error compared with any other estimator of the sample parameter, is called efficient estimator of that parameter.

c) Consistency

An estimator is said to be consistent estimator of a population parameter if the value of the estimator comes closer and closer to the population parameter and the sample size increase.

d) Sufficiency

An estimator is said to be sufficient estimator. It the estimator used as much as possible information, available from the sample about the parameter. No other estimator can provide any additional information about the parameter.

2. Test of Hypothesis

A Hypothesis Testing the researcher begins with some assumptions about the true value of the population parameter. Then he will decide to accept or reject the hypothesis on the basis of the test statistics calculation from the sample statistics.

The process of accepting or rejecting a hypothesis on the basis of sample result is called test of hypothesis.

The purpose of test of hypothesis is to make the judgement about the difference between sample statistic and population parameter.

Definition of Hypothesis

The Hypothesis is a quantitative statement about the population parameter (or) probability distribution.

Hypothesis is nothing but a statement about the parameter which being tested and there by accepted (or) rejected.

The Hypothesis is the tentative statement about any parameter of the population which may or may not be true on the basis of sample information.

Characteristics of Hypothesis

1. Hypothesis should be clear and precious for reliable inference capable of being tested.
2. The relationship between the variable if it is a relational hypothesis.
3. It should be consistent which most known tests.
4. Stated in most simple terms are easy understanding.
5. Limited in scope

Procedure of Hypothesis Testing

The tests conducted to accept (or) reject the hypothesis are known as Test of Hypothesis. The procedure for testing a hypothesis include a following steps.

Step I: Formulate the Null and alternative hypothesis

The First step of test of hypothesis is to formulate the Null and Alternative hypothesis.

The Reason for setting two hypothesis is if one hypothesis is false (or) rejected.

Null Hypothesis

The Assumption a researcher is trying to rejects is called Null Hypothesis. It is denoted by the symbol 'Ho'. Null Hypothesis is the original Hypothesis. A Hypothesis which we want to test on the basis of the evidence of sample observation is called Null Hypothesis. Generally the statistician formulate the Null hypothesis exactly opposite of what he wants to verify. A common of stating the 'Ho' is there is no significant difference between the sample value and population value. The word no significant difference means that lase of there is any different it is due to sampling fluctuation.

Alternative Hypothesis

Any Hypothesis which is different from Null Hypothesis is called alternative Hypothesis. Simply 'AH' is just opposite to Null Hypothesis. It is denoted by the symbol 'H_1'.

For example, the H_0 and H_1 relating to the average Height of the soldiers may be stated as follows

$H_0 = \mu = 165$ cm

$H_1 = \mu \neq 165$ cm

2) Choose (or) select suitable level of Significance

This is an important concept in the context of hypothesis Testing. The Maximum probability of rejecting the Null Hypothesis when it is true is called level of Significance. The level of significance should be stated usually before a test is maid. If we adopt 5 % level of significance, it means 95 % confidence, that we will make correct decision.

3) Select Appropriate Test Statistic

The decision to accept the Null Hypothesis or H_1 is made on the basis of the statistic calculated from the sample. Such a statistic is called test statistic. Test statistics are generally based on some probability distribution. Some of the common probability distribution used in test of hypothesis are;

'z' Test, 't' Test, χ^2(Chi - Square) Test, 'F' Test and So on.

4) Decision Making

It the calculated value of the test statistic is less than are equal to table value of test statistic at specified level of significance then the stated Null Hypothesis is accepted. It is calculated value of the test statistic is greater than table value of test statistic at specified level of significance, then the stated Null hypothesis is rejected.

Level of Type I and Type II error

The decision to accept (or) reject H_0 is made on the basis the informations supply by the sample observations. Therefore the conclusions made on the basis of a particular sample above the population may or may not be true. There is chance of making error in the decision. Generally two types of errors are committed in the test of Hypothesis. They are Type I and Type II error.

Type I error

An error caused by rejecting the null hypothesis, when it is true is called type I error.

(i.e) Type I error: Reject H_0 when H_0 is true. The probability of committing Type I error is denoted by the symbol α

(i.e) P(Type I error) = α

The Type I error is also known as rejection errors.

Type II error

An error caused by accepting the Null Hypothesis, when it is falls is called Type II error.

(i.e) Type II error = Accept H_0 when H_0 is false.

Probability of commiting Type II error is denoted by the symbol β.

(i.e) P(Type II error) = β

The Type II error is also known as acceptance error.

When a hypothesis tested there are 4 possibilities. The 4 possibilities can be shown the following Table.

Nature of Hypothesis

Decision	**H_0 is True**	**H_0 is false**
Accept H_0	Correct decision	Wrong decision (Type II error)
Reject H_0	Wrong decision (Type I error)	Correct decision

In any Test of hypothesis there is a nisk of committing type I and Type II error. It is not possible to minimize both the error at a time for a given sample size.

In general committing Type II error, is more dangerous than Type I error. Therefore a test of hypothesis, a probability type I error is fixed at 5 % level (or)1% level and then probability of committing type II error is minimised.

CHI-SQURARE DISTRIBUTION (χ^2)

Chi-square test is one of the most important non-parometric (or) distribution free test introduced in 1990 by Karl pearson. Chi-square distribution with a mathematically expressed frequency function. The chi-square distribution may be defined as the sum of the square of independent, normally distributed variables with zero means and unit variances.

The square of a standard normal variate is called a chi-square variate with 1 degree of freedom.

Features of χ^2- distribution

1. The distribution has only one parameter.
2. Chi-square is non-negative in value. It is either zero (or) positively valued.
3. There are many chi-square distributions.
4. The shape of a chi-square distribution curve is skewed.
5. The total area under a chi-square distribution curve is unity.

Uses of chi-square Test

1. Test of goodness of fit
2. Test of independence
3. Test of population variance through confidence intervals suggested by χ^2 test.

4. Test of homogeneity to determine whether 2 or more independent random samples are drawn from the same population.

Properties of Chi-square Test

Mean of χ^2distribution = Degree of freedom = df and SD of χ^2distribution = $\sqrt{2df}$
Median of χ^2 distribution divides the area of the curve into two equal parts each being 0.5.
Mode of χ^2 distribution = df-2.
Since Chi-square values are always positive, the chi-square curve is always positively skewed.
The lowest value of $\chi^2$is zero and the highest value is infinite.

'Z' Test

Under this heading we will discuss tests of significance when samples are large (i.e) when the samples are of size n>30. Further since, for large samples almost all distributions are closely approximated by normal distribution. Therefore normal distribution forms the statistical basis of all large sample test. Thus, in all large sample tests, we compute the test statistic z under H_0, where z is a standard normal distribution with mean o and variance 1.

Steps

1. State null and alternative hypothesis
2. Set up a suitable significance level.
3. Select a Suitable sample statistic whose sampling distribution is known.
4. Define and compute the test statistic under null hypothesis H_0.
5. Find the critical region and establish the decision rule.
6. The final step in hypothesis testing is to draw a statistical decision, involving the acceptance (or) rejection of the null hypothesis.

Hypothesis Test concerning the differences between two standard deviations

As an extension to what we have discussed earlier, let us now consider whether or not two samples having different standard deviations could have come from the sample population. We know that the standard diviation of any distribution is determined by the extent to which the individual items are dispersed around the mean. It was explained earlier that we usually do not know the standard deviation of the population and have to resort to the standard deviation of the sample, provided that the sample is large.

Under the null hypothesis H_0: that sample standard deviation do not differ significantly,

$$z = \frac{S_1 - S_2}{\sqrt{\left[\left(\frac{\sigma_1^2}{2n_1}\right)+\left(\frac{\sigma_2^2}{2n_2}\right)\right]}}$$

$$= \frac{S_1 - S_2}{\sqrt{\left[\left(\frac{s_1^2}{2n_1}\right)+\left(\frac{s_2^2}{2n_2}\right)\right]}}$$

($\therefore\ \sigma_1^2 = s_1^2$ and $\sigma_2^2 = s_2^2$ for large samples)

'F' - Distribution

The 'F' distribution which was developed in 1924 by the great English statisticion R.A. Fisher.

The random variable defined by the ratio of two independent chi-square distributed random variables each divided by its respective degrees of freedom. $\dfrac{\left(\dfrac{x_1^2}{v_1}\right)}{\left(\dfrac{x_2^2}{v_2}\right)}$ follows the F-distribution. Its parameters are the number of degrees of freedom for the numerator, v_1 and for the denominator v_2. Given v_1 and v_2 and f-distribution is completely specified.

The values of F-range from zero to positive infinity. Generally the F-distribution is positively skewed.

Given two independent random samples of sizes n_1 and n_2 from two normal populations with unknown means, we may be required to test the null hypothesis. H_0: σ_1^2, (i.e) population variances are same.

The unbiased estimates of population variances σ_1^2 and σ_2^2 are computed from the samples.

$$\hat{\sigma}_1^2 = s_1^2 = \frac{1}{n_1 - 1}\sum(x-\bar{x})^2 \quad \text{(or)} \quad \frac{n_1 s_1^2}{n_1 - 1}$$

$$\hat{\sigma}_{21}^2 = \frac{1}{n_2 - 1}\sum(y-\bar{y})^2 \quad \text{(or)} \quad \frac{n_2 s_2^2}{n_2 - 1}$$

The technique of Analysis of Variance

The first step in the analysis of variance is to partition the total variation in the sample data into the following two components variations in such a way that it is possible to estimate the contribution of factors that may cause variation.

1. The amount of variation among the sample means
2. The amount of variation within the sample observations.

The observations in the sample data may be classified according to one factor (criterion) or two factors (criteria).The classifications according to one factor and two factors are called one-way classification and two-way classification, respectively. The calculations for total variation and its components may be carried out in each of the two-types of classifications by (i) direct method, (ii) short-cut and (iii) coding method.

One-way classification to Test Equality of population mean

In one-way analysis of variance, observations are classified into groups or samples on the basis of single criterion. Suppose k independent random samples (or groups of observations) are drawn on from each of k normal populations with means μ_1, μ_2, ...μ_k and common variance σ^2 (unknown). On the basis of the data, it is required to test the null hypothesis that the population means are equal, i.e.

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

Steps for Testing Null Hypothesis

1. State hypothesis to test the equality of population mean as

Null hypothesis, $H_0 : \mu_1 = \mu_2 = \mu_k$

Alternative hypothesis, H_1 : Not μ_is are equal (j = 1, 2, …. k)

α = level of significance

2. Calculate total variation. Total variation is represented by the 'Sum of Squares Total' (SST) and is equal to the sum of the squared differences between each sample value from the grand mean $\overline{\overline{X}}$, i.e.

$$SST = \sum_{i=1}^{r}\sum_{j=1}^{k}\left(X_{ij} - \overline{\overline{X}}\right)^2$$

3. Calculate variation between sample means. This is usually called the 'Sum of squares between'(SSB). In statistical terms, variation between sample means is also called the between – column variance, difference between the means of the various samples and the grand mean square the deviations and multiply them by the numbers of respective samples to get sum of squares between samples (or groups). i.e,

$$SSB = n_1\left(\overline{X} - \overline{\overline{X}}\right)^2 + n_2\left(\overline{X}_2 - \overline{\overline{X}}\right)^2 + ...+ n_k\left(\overline{X}_k - \overline{\overline{X}}\right)^2$$

4. Calculate variation within samples. This is usually called the 'Sum of squares within'(SSW) and measures the difference within samples due to chance error.

5. Calculate average variation between and within samples mean squares. Since k independent samples are being compared, k-1 degrees of freedom are associated with the sum of the squares among samples. n – k degree of freedom associated with the sum of the squares within samples.

The mean squares are now calculated on dividing the sum of squares by the corresponding degree of freedom.

6. When the null hypothesis H_0 is true, both the mean squares MSB and MSW provide independent unbiased estimates of the same population variance (σ^2).

7. Make decision regarding null hypothesis. The calculated value of F-ratio is compared with the table value of F at 5% ($F_{.05}$) or 1% (F_{01}) level of significance for given number of degrees of freedom (v_1,v_2). If F (calculated) > $F_{.05}$, the null hypothesis is rejected, that means that the difference between sample means is deemed significant.

Short – cut Method

The values of SSB and SSW can be calculated by applying the following short – cut methods.

- Calculate the total G of the observations in samples from each of k samples, i.e.

  $G = \Sigma X_1 + \Sigma X_2 + …. + \Sigma X_k$ or $T_1 + T_2 + ….. + T_k$

- Compute the Correction Factor, $CF = \frac{G^2}{n}$; $n = n_1 + n_2 + …. + n_k$

- Find the sum of squares of all the observations in samples from each of k samples and subtract CF from this sum to obtain the total sum of squares of deviations (SST), i.e.,

  $$SST = \left(\sum X_1^2 + \sum X_2^2 + ... + \sum X_k^2\right) - CF = \sum_{i=1}^{k}\sum_{j=1}^{r} X_{ij}^2 - CF$$

  SSW = TSS – SSB

Coding method

Coding refers to the addition, multiplication, substraction or division of data by constant or common factor. In the computation of analysis of variance, the final quantity tested is a ratio, hence dimensionless. Thus the original values can be coded by reducing the sample observations by subtracting a suitable constant to simply calculations without need for any sub sequent adjustment of the results.

Sign Test

The sign Test is used to test the null hypothesis that the medium of a distribution is equal to some value. It can be used

a. In place of a one sample t- test
b. In place of a paired t – test
c. for ordered categorical data where a numerical scale is in appropriate but where it is possible to rank the observations.

The sign Test is a statistical method to test for consistent difference between pair of observation. Such as the weight of subject before & after Treatment.

The sign Test can also test if the median of a collection of number is significantly greater than (or) less than a specified value.

The sign Test is a non-parametric test which makes very few assumptions about the nature of the distributions under test.

Median Test

In statistics, Median Test is a special case of pearson's Chi-squared Test. It is a non-parametric test that tests the null hypothesis that the medians of the populations from which two or more samples are drawn are identical. The data in each sample are assigned to two groups. One consisting of data whose values are higher than the median value. In the two groups combined, and the other consisting of data whose values are at the median (or) below. A pearson's chi-squared test is then used to determine whether the observed frequencies in each sample differ from expected frequencies derived from a distribution combining the two groups.

Fisher's Transformation Test

The T-test discussed earlier is applicable only for determining whether the computed r is significantly different from zero. When we want to test the sample correlation against any other theoretical value of r (or) if it is desired to test whether the two given samples have come from the same population (or) not, t-test cannot be used. Prof. Ronald fisher has shown that if it is changed to another statistic z by a suitable transformation then this testing is possible. He has derived that if z is calculated by z = 1.1513 log $\frac{1+r}{1-r}$ the distribution of z will be approximately normal.

The standard deviation of z distribution = $\sqrt{\frac{1}{N-3}}$ (where N is the size of sample)

The arithmetic mean of z distribution will be the 'z' corresponding to the population of r.

z transformation of r = $\frac{1}{2}\log_e \frac{1+r}{1-r} = 1.1513\log_{10}\frac{1+r}{1-r}$

z transformation of p (population) = $\frac{1}{2}\log_e \frac{1+p}{1-p} = 1.1513\log_{10}\frac{1+p}{1-p}$

Reference

1. D. Bose, An Introduction to Mathematical Methods, Himalaya publishing House, 2004.
2. Dr. Peer Mohamed & Dr. Shazuli Ibrahim, Business Mathematics, Pass publications, 2008.
3. V.K. Kapoor Modern Approach to Fundamentals of statistics for Business and Economics, Sultan channel & sons., 2015
4. S.C. Gupta and V.K. Kapoor, Elements of Mathematical Statistics, Sulton chanel & sons, 1979.
5. A.P Verma, Business Mathematics and statistics, Asian Books private limited, 2002.
6. Agarval B.L, Basic statistics, New Age International,
7. Mood. A.M, Graybill F.A and Bose D.C, Introduction to Theory of statistics, MC. Graw Hill publications.

Course Material Prepared by
Dr. VICTORIA
Department of Economics
S.T. Hindu College, Nagercoil.