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

B.Sc Degree – Mathematics Major **18K4MAS3**

Allied course – Mathematical Statistics – III **Hrs:4**

Credit:3

Unit – I

Normal distribution – mean, median, mode, moments, β_1 and β_2, moment generating function and uses of Normal distribution, Binomial, Poisson and Chi – Square distribution tends to Normal distribution.

Unit – II

Continuous distributions – Rectangular, Exponential, Beta, gamma and their pdf, mgf, mean and variance.

Unit – III

Correlation – Definition and uses, Karl Pearson's coefficient of correlation, Spearman's rank correlation and their properties. Simple linear regression lines, regression coefficient and their properties.

Unit – IV

Tests of significance – Definition of Null hypothesis, alternative hypothesis, sampling distribution, standard error and critical region. Type I and Type II errors, one tailed and two tailed tests. Large sample test for single mean, difference between means, single proportion and difference between proportions.

Unit – V

Small sample tests – 't'- test for single mean, Difference between means. Paired 't' test, Chi – Square test for goodness of fit and independence of attributes.

Books for Study :

1. Fundamentals of Mathematical Statistics - S.C. Gupta & V.K. Kapoor.
2. Statistical Methods – S.P. Gupta.(Revised edition 2001)

UNIT - 1

NORMAL DISTRIBUTION :

A random variable x is said to have a normal distribution with parameters μ (called mean) and σ^2 (called 'variance') its p.d.f is given by the probability law :

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right\}$$

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

$; -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0.$

Mode of Normal Distribution :

Mode is the value of x for which $f(x)$ is maximum, i.e., mode is the solution of

$$f'(x) = 0 \quad \text{and} \quad f''(x) < 0$$

For normal distribution with mean μ and standared deviation σ,

$$\log f(x) = c - \frac{1}{2\sigma^2}(x-\mu)^2,$$

where $c = \log(1/\sqrt{2\pi}\,\sigma)$, is a constant, Differentiating with respect to x, we get

$$\frac{1}{f(x)} \cdot f'(x) = \frac{-1}{\sigma^2}(x-\mu) \Rightarrow f'(x) = -\frac{1}{\sigma^2}(x-\mu)f(x)$$

and $$f''(x) = -\frac{1}{\sigma^2}\left[1 \cdot f(x) + (x-\mu)f'(x)\right]$$

$$= -\frac{f(x)}{\sigma^2}\left[1 - \frac{(x-\mu)^2}{\sigma^2}\right] \quad \text{--- (1)}$$

$f'(x) = 0 \quad \Rightarrow x - \mu = 0 \Rightarrow x = \mu.$

At the point $x - \mu$,

we have from (i) :

$$f''(x) = \frac{1}{\sigma^2} \left[f(x) \right]_{x=\mu}$$

$$= \frac{1}{\sigma^2} \cdot \frac{1}{\sigma\sqrt{2\pi}} < 0$$

Hence $x = \mu$, is the mode of the normal distribution.

Medium of Normal Distribution :

If M is the median of the normal distribution, we have

$$\int_{-\infty}^{M} f(x)\,dx = \frac{1}{2} \Rightarrow \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{M} \exp\{-(x-\mu)^2/2\sigma^2\}\,dx = \frac{1}{2}$$

$$\Rightarrow \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu} \exp\{-(x-\mu)^2/2\sigma^2\}\,dx + \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} \exp\{-(x-\mu)^2/2\sigma^2\}\,dx$$

$$= 1/2 \quad \text{--- (1)}$$

But $\frac{1}{\sigma\sqrt{2\pi}} \int_{\infty}^{\mu} \exp\{(x-\mu)^2/2\sigma^2\}\,dx$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} \exp(-z^2/2)\,dz = \frac{1}{2}$$

$\therefore$ From (1) we have $\frac{1}{2} + \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} \exp\{-(x-\mu)^2/2\sigma^2\}\,dx = 1/2$.

$$\Rightarrow \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} \exp\{-(x-\mu)^2/2\sigma^2\}\,dx = 0$$

i.e., $\mu = M$

Hence, for the normal distribution, Mean = Median.

Moments of Normal Distribution :

Odd order moments about mean are given by:

$$\mu_{2n+1} = \int_{-\infty}^{\infty} (x-\mu)^{2n+1} f(x)\, dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^{2n+1} \exp\{-(x-\mu)^2/2\sigma^2\}\, dt$$

$$\mu_{2n+1} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n+1} \exp(-z^2/2)\, dz, \quad \left(z = \frac{x-\mu}{\sigma}\right)$$

$$= \frac{\sigma^{2n+1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n+1} \exp(-z^2/2)\, dz = 0 \quad \text{——} \quad (1)$$

Even Order moments about mean are given by :

$$\mu_{2n} = \int_{-\infty}^{\infty} (x-\mu)^{2n} f(x)\, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n} \exp(-z^2/2)\, dz.$$

$$= \frac{\sigma^{2n}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n} \exp(-z^2/2)\, dz$$

$$= \frac{\sigma^{2\pi}}{\sqrt{2\pi}}\, 2 \int_{0}^{\infty} z^{2n} \exp(-z^2/2)\, dz$$

(Since integral is an even function of z).

$$= \frac{2\sigma^{2\pi}}{\sqrt{2\pi}} \int_{0}^{\infty} (2t)^n e^{-t} \frac{dt}{\sqrt{2t}}, \quad (t = z^2/2)$$

$$\therefore \mu_{2n} = \frac{2^n \sigma^{2n}}{\sqrt{\pi}} \int_{0}^{\infty} e^{-t}\, t^{(n+1/2)-1}\, dt$$

$$\Rightarrow \mu_{2n} = \frac{2^n \sigma^{2n}}{\sqrt{\pi}} \cdot \Gamma(n+1/2)$$

changing n to $(n-1)$, we get

$$\mu_{2n-2} = \frac{2^{n-1} \cdot \sigma^{2n-2}}{\sqrt{\pi}} \Gamma(n-1/2)$$

$$\therefore \frac{\mu_{2n}}{\mu_{2n+2}} = 2\sigma^2 . \frac{\Gamma(n+1/2)}{(n+1/2)} = 2\sigma^2 (n-1/2)$$

$$\Rightarrow \mu_{2n} = \sigma^2 (2n-1)\mu_{2n-2} \quad \text{—— (1)} \qquad [\because \Gamma(r) = (r-1)\Gamma(r-1)]$$

which gives the recurrence relation for the moments of normal distribution.

From (1), we have

$$\mu_{2n} = [(2n-1)\sigma^2][(2n-3)\sigma^2]\mu_{2n-4}$$

$$= [(2n-1)\sigma^2][(2n-3)\sigma^2][(2n-5)\sigma^2]\mu_{2n-6}$$

$$= [(2n-1)\sigma^2][(2n-3)\sigma^2][(2n-5)\sigma^2]\ldots(3\sigma^2)(1\sigma^2)\mu_0$$

$$= 1.3.5 \ldots\ldots (2n-1)\sigma^{2n} \quad \text{—— (2)}$$

we conclude that for the normal distribution all odd order moments about mean vanish and even order moments about mean are given by

$1.3.5 \ldots (2n-1)\sigma^{2n}$

Moment Generating Function of Normal Distribution :

The m.g.f (about origin) is given by :

$$M_x(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} \exp\{-(x-\mu)^2/2\sigma^2\}\, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{t(\mu+\sigma z)\} \exp(-z^2/2)\, dz, \quad \left(z = \frac{x-\mu}{\sigma}\right)$$

$$= e^{\mu t} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}(z^2 - 2t\sigma z)\}\, dz$$

$$= e^{\mu t + t^2\sigma^2/2} \times \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}(z-\sigma t)^2\}\, dz$$

$$= e^{\mu t + t^2\sigma^2/2} \times \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-u^2/2)\, du$$

$$M_{(x)}(t) = e^{\mu t + t^2\sigma^2/2}$$

M.G.F. of Standard Normal Variate.

If $X \sim N(\mu, \sigma^2)$, the standard normal variate is given by :

$$Z = (X-\mu)/\sigma.$$

$$M_Z(t) = e^{-\mu t/\sigma} M_X(t/\sigma)$$

$$= \exp(-\mu t/\sigma) \cdot \exp\{(\mu t/\sigma) + (t^2/\sigma^2)(\sigma^2/2)\}$$

$$= \exp(t^2/2)$$

UNIT - II

RECTANGULAR (OR UNIFORM) DISTRIBUTION :-

A random variable x is said to have a continuous rectangular (uniform) distribution over an interval (a, b), i.e., $(-\infty < a < b < \infty)$, if its p.d.f. is given by:

$$f(x\,;a,b) = \begin{cases} \dfrac{1}{b-a}, & \text{if } a < x < b \\ 0, & \text{otherwise} \end{cases}$$

Moments of Rectangular Distribution. Let $X \sim U[a,b]$.

$$\mu_r' = \int_a^b x^r f(x)\,dx$$

$$= \frac{1}{b-a}\int_a^b x^r\,dx = \frac{1}{b-a}\left(\frac{b^{r+1}-a^{r+1}}{r+1}\right) \qquad \text{—— (1)}$$

In particular

$$\text{Mean} = \mu_1' = \frac{1}{b-a}\left(\frac{b^2-a^2}{2}\right) = \left(\frac{b+a}{2}\right)$$

and

$$\mu_2' = \frac{1}{b-a}\left(\frac{b^3-a^3}{3}\right) = \frac{1}{3}(b^2+ab+a^2)$$

$$\therefore \text{Variance} = \mu_2' - \mu_1'^2 = \frac{1}{3}(b^2+ab+a^2) - \left\{\tfrac{1}{2}(b+a)\right\}^2$$

$$= \frac{1}{12}(b-a)^2$$

M.G.F. of Rectangular Distribution :

$$M_x(t) = \int_a^b e^{tx}\,dx = \frac{e^{ibt}-e^{iat}}{it(b-a)}, \; t \neq 0$$

Moment Generating function of Exponential Distribution:

$$M_X(t) = E(e^{tX}) = \theta\int_0^{\infty} e^{tx} e^{-\theta x} dx = \theta\int_0^{\infty} exp\{-(\theta-t)x\} dx$$

$$= \frac{\theta}{(\theta-t)} = \left(1-\frac{t}{\theta}\right)^{-1} = \sum_{r=0}^{\infty}\left(\frac{t}{\theta}\right)^r, \theta > t$$

$\therefore \mu_r' = E(X^r) =$ coefficient of $\frac{t^r}{r!}$ in $M_X(t) = \frac{r!}{\theta^r}$; $r = 1, 2, \ldots$

$\Rightarrow$ Mean $= \mu_1' = \frac{1}{\theta}$ and

Variance $= \mu_2 = \mu_2' - \mu_1'^2 = \frac{2}{\theta^2} - \frac{1}{\theta^2} = \frac{1}{\theta^2}$

Hence, if $X \sim exp(\theta)$, the Mean $= \frac{1}{\theta}$ and variance $= \frac{1}{\theta^2}$

Remark. Variance $= \frac{1}{\theta^2} = \frac{1}{\theta}\cdot\frac{1}{\theta} = \frac{\text{Mean}}{\theta}$

$\therefore$ Variance > Mean, if $0 < \theta < 1$

Variance = Mean, if $\theta = 1$

and Variance < Mean, if $\theta > 1$

Hence for the exponential distribution,

Variance >, =, or < Mean, for different values of the parameter.

Gamma Distribution:

A random variables X is said to have a gamma distribution with parameters $\lambda > 0$, if its p.d.f is given by:

$$f(x) = \begin{cases} \dfrac{e^{-x} x^{\lambda-1}}{\Gamma(\lambda)} ; & \lambda > 0, 0 < x < \infty \rightarrow (a) \\ 0, & \text{otherwise} \end{cases}$$

MGF of Gamma Distribution :

M.G.F about Origin is given by :

$$M_X(t) = E(e^{tX}) = \int_0^{\infty} e^{tx} f(x)\, dx$$

$$= \frac{1}{\Gamma(\lambda)} \int_0^{\infty} e^{tx} e^{-x} x^{\lambda-1}\, dx$$

$$= \frac{1}{\Gamma(\lambda)} \int_0^{\infty} e^{-(1-t)x} x^{\lambda-1}\, dx$$

$$= \frac{1}{\Gamma(\lambda)} \cdot \frac{\Gamma(\lambda)}{(1-t)^{\lambda}}, \quad |t|<1$$

$$M_X(t) = (1-t)^{-\lambda}, \quad |t|<1.$$

Cumulant Generating function of Gamma Distribution:

The cummulant generating function $K_X(t)$ is given by :

$$K_X(t) = \log M_X(t) = \log (1-t)^{-\lambda} = -\lambda \log(1-t); \; |t|<1$$

$$= \lambda\left(t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \cdots\right)$$

$\log(1) - \log(t)$

$-\lambda\left(0 - t^2 + \frac{t^2}{2!} + \cdots\right)$

$\therefore$ Mean $= K_1$ = co-efficient of t in $K_X(t) = \lambda$

Variance $= \mu_2 = K_2$ = co-efficient of $\frac{t^2}{2!}$ in $K_X(t) = \lambda$

Hence if $X \sim \gamma(\lambda)$, Mean = Variance = λ

$\mu_3 = K_3$ = coefficient of $\frac{t^3}{3!}$ in $K_X(t) = 2\lambda$

K_4 = Coefficient of $\frac{t^4}{4!}$ in $K_X(t) = 6\lambda$

$$\mu_4 = K_4 + 3K_2^2 = 6\lambda + 3\lambda^2$$

Hence $$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{4\lambda^2}{\lambda^3} = \frac{4}{\lambda}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{6}{\lambda}$$

Remark

Like poisson Distribution, the mean and Variance of the Gamma distribution are also equal. However, poisson distribution is discrete while Gamma distribution is continuous.

Beta distribution of first kind

A random variable x is said to have a beta distribution of first kind with parameters μ and v $(\mu > 0, v > 0)$ if its p.d.f is given by:

$$f(x) = \begin{cases} \frac{1}{B(\mu,v)} \cdot x^{\mu-1}(1-x)^{v-1} & ; (\mu,v) > 0, 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

[where $B(\mu,v)$ is the Beta function].

The random variables is known as a Beta variate of the first kind with parameters μ and v is referred to as $\beta_1(\mu,v)$ variate.

Constants of Beta Distribution of first kind

$$\mu_r' = \int_0^1 x^r f(x)\, dx$$

$$= \frac{1}{B(\mu,v)} \int_0^1 x^{\mu+r-1}(1-x)^{v-1} dx$$

$$= \frac{1}{B(\mu,v)} B(\mu+r, v)$$

$$\mu_r' = \frac{\Gamma(\mu+r)\,\Gamma(\mu+v)}{\Gamma(\mu+r+v)\,\Gamma(\mu)}$$

$r=1$ $[\because \Gamma(K) = K-1\,\Gamma(K-1)]$

$$\mu_1' = \frac{\Gamma(\mu+1)\,\Gamma(\mu+\nu)}{\Gamma(\mu+1+\nu)\,\Gamma(\mu)}$$

$$= \frac{\mu+1-1\,\Gamma(\mu+1-1)\,\Gamma(\mu+\nu)}{(\mu+\nu+1-1)\,\Gamma(\mu+\nu+1-1)\,\Gamma(\mu)} = \frac{\mu\,\cancel{\Gamma(\mu)}\,\cancel{\Gamma(\mu+\nu)}}{(\mu+\nu)\,\cancel{\Gamma(\mu+\nu)}\,\cancel{\Gamma(\mu)}}$$

$$\mu_1' = \frac{\mu}{(\mu+\nu)}$$

$r=2$

$$\mu_2' = \frac{\Gamma(\mu+2)\,\Gamma(\mu+\nu)}{\Gamma(\mu+2+\nu)\,\Gamma(\mu)} = \frac{\mu+2-1\,\Gamma(\mu+2-1)\,\Gamma(\mu+\nu)}{\mu+\nu+2-1\,\Gamma(\mu+\nu+2-1)\,\Gamma(\mu)}$$

$$= \frac{\mu+1\,\Gamma(\mu+1)\,\Gamma(\mu+\nu)}{\mu+\nu+1\,\Gamma(\mu+\nu+1)\,\Gamma(\mu)} = \frac{(\mu+1)(\mu+1-1)\,\Gamma(\mu+1-1)\,\Gamma(\mu+\nu)}{(\mu+\nu+1)(\mu+\nu+1-1)\,\Gamma(\mu+\nu+1-1)\,\Gamma(\mu)}$$

$$= \frac{(\mu+1)(\mu)\,\cancel{\Gamma(\mu)}\,\cancel{\Gamma(\mu+\nu)}}{(\mu+\nu+1)(\mu+\nu)\,\cancel{\Gamma(\mu+\nu)}\,\cancel{\Gamma(\mu)}} = \frac{(\mu+1)(\mu)}{(\mu+\nu+1)(\mu+\nu)}$$

Variance

$$\mu_2 = \mu_2' - \mu_1'^2$$

$$= \frac{(\mu+1)(\mu)}{(\mu+\nu+1)(\mu+\nu)} - \left(\frac{\mu}{\mu+\nu}\right)^2$$

$$= \frac{\mu}{(\mu+\nu)^2(\mu+\nu+1)}\left[(\mu+\nu)(\mu+1) - \mu(\mu+\nu+1)\right]$$

$$= \frac{\mu}{(\mu+\nu)^2(\mu+\nu+1)}\left[\cancel{\mu^2}+\cancel{\mu}+\cancel{\mu\nu}+\nu-\cancel{\mu^2}-\cancel{\mu\nu}-\cancel{\mu}\right]$$

$$\mu_2 = \frac{\mu\nu}{(\mu+\nu)^2(\mu+\nu+1)}$$

Similarly we have

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$

$$= \mu_3' - 3\left(\frac{\mu(\mu+1)}{(\mu+\nu)(\mu+\nu+1)}\right)\left(\frac{\mu}{\mu+\nu}\right) + 2\left(\frac{\mu}{\mu+\nu}\right)^3$$

$$\mu_3' = \frac{(\mu+2)(\mu+1)(\mu)}{(\mu+\nu+2)(\mu+\nu+1)(\mu+\nu)}$$

$$\mu_3 = \frac{(\mu+2)(\mu+1)(\mu)}{(\mu+\nu+2)(\mu+\nu+1)(\mu+\nu)} - 3\frac{\mu(\mu+1)}{(\mu+\nu)(\mu+\nu+1)}\left(\frac{\mu}{\mu+\nu}\right) + 2\left(\frac{\mu}{\mu+\nu}\right)^3$$

On simplification

$$\mu_3 = \frac{2\mu\nu(\nu-\mu)}{(\mu+\nu)^3(\mu+\nu+1)(\mu+\nu+2)}$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'$$

On simplification

$$= \frac{3\mu\nu\{\mu\nu(\mu+\nu-6) + 2(\mu+\nu)^2\}}{(\mu+\nu)^4(\mu+\nu+1)(\mu+\nu+2)(\mu+\nu+3)}$$

$$\therefore \beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

$$= \left(\frac{2\mu\nu(\nu-\mu)}{(\mu+\nu)^3(\mu+\nu+1)(\mu+\nu+2)}\right)^2 \times \left(\frac{(\mu+\nu)^2(\mu+\nu+1)}{\mu\nu}\right)^3$$

$$= \frac{4\cancel{\mu^2}\cancel{\nu^2}(\nu-\mu)^2}{\cancel{(\mu+\nu)^6}\cancel{(\mu+\nu+1)^2}(\mu+\nu+2)^2} \times \left(\frac{\cancel{(\mu+\nu)^6}(\mu+\nu+1)^3}{\mu^{\cancel{3}}\nu^{\cancel{3}}}\right)$$

$$\beta_1 = \frac{4(\nu-\mu)^2(\mu+\nu+1)}{\mu\nu(\mu+\nu+2)^2}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

$$= \frac{3\cancel{\mu\nu}\{\mu\nu(\mu+\nu-6)+2(\mu+\nu)^2\}}{\cancel{(\mu+\nu)^4}\cancel{(\mu+\nu+1)}(\mu+\nu+2)(\mu+\nu+3)} \times \frac{\cancel{(\mu+\nu)^4}(\mu+\nu+1)^2}{\mu^2\nu^2}$$

$$\beta_2 = \frac{3(\mu+\nu+1)\{\mu\nu(\mu+\nu-6)+2(\mu+\nu)^2\}}{\mu\nu(\mu+\nu+2)(\mu+\nu+3)}$$

The harmonic mean H is given by

$$\frac{1}{H} = \int_0^1 \frac{1}{x} f(x)\,dx$$

$$= \frac{1}{B(\mu,\nu)} \int_0^1 x^{\mu-2}(1-x)^{\nu-1}\,dx$$

$$= \frac{1}{B(\mu,\nu)} B(\mu-1,\nu)$$

$$= \frac{\Gamma(\mu-1)\Gamma(\nu)}{\Gamma(\mu+\nu-1)} \cdot \frac{\Gamma(\mu+\nu)}{\Gamma(\mu)\Gamma(\nu)}$$

$$= \frac{\cancel{\Gamma(\mu-1)}(\mu+\nu-1)\cancel{\Gamma(\mu+\nu-1)}}{\cancel{\Gamma(\mu+\nu-1)}(\mu-1)\cancel{\Gamma(\mu-1)}}$$

$$= \frac{\mu+\nu-1}{\mu-1}$$

$$H = \frac{\mu-1}{\mu+\nu-1}$$

UNIT - III

Correlation :

According to ya Lun chou, "correlation analysis attempts to determine the degree of relationship between variables".

Types of correlation :

Correlation is classified into many types, but the important are

1. positive and Negative
2. Simple and Multiple
3. Partial and total
4. Linear and Non-linear

Karl pearson's coefficient of correlation :

Karl pearson, a great biometrician and statistician, suggested a mathematical method for measuring the magnitude of linear relationship between two variables. Karl pearson's method is the most widely used method in practice and is known as pearsonian coefficient of correlation.

(1) $r = \frac{\text{covariance of } xy}{\sigma x \times \sigma y}$ (2) $r = \frac{\Sigma xy}{N \sigma x \times \sigma y}$ (3) $\pm r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$

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

σx = Standard deviation of series x

σy = standard deviation of series y

when the deviation of items are taken from the actual mean, we can apply any one of these methods; but the simplest formula is the third one.

Properties of coefficient of correlation :

1. The measure of correlation, called coefficient of correlation, summarizes in one figure, the direction and the degree of correlation

2. The value of the coefficient of correlation shall always lie between +1 and -1

3. When $r = +1$, then there is perfect positive correlation between the variables

4. When $r = -1$, then there is perfect Negative correlation between the variables.

5. When $r = 0$, then there is no relation between the variables.

Theoretically, We get values which lie between +1 and -1; but normally the values lies between -0.8 and -0.5;

$r = +0.8$ means there is positive correlation, because r is positive and the magnitude of correlation is 0.8 -0.5 means that the correlation is negative and magnitude of correlation is 0.5. Thus, the coefficient of correlation describes the magnitude and the direction of correlation.

The third formula given above, that, $r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$ is easy to calculate, and it is not necessary to calculate the standard deviation of x and y series separately.

Illustration 12.2

calculate coefficient of correlation from the following data.

X	12	9	8	10	11	13	7
Y	14	8	6	9	11	12	3

Soln

Computation of coefficient of correlation

X	Y	x^2	y^2	xy
12	14	144	196	168
9	8	81	64	72
8	6	64	36	48
10	9	100	81	90
11	11	121	121	121
13	12	169	144	156
7	3	49	9	21
$\Sigma x = 70$	$\Sigma y = 63$	$\Sigma x^2 = 728$	$\Sigma y^2 = 651$	$\Sigma xy = 676$

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

$\Sigma xy = 676$; $\Sigma x = 70$; $\Sigma y = 63$; $\Sigma x^2 = 728$; $\Sigma y^2 = 651$; $N = 7$

$$r = \frac{(676 \times 7) \times (70 \times 63)}{\sqrt{(728 \times 7 - (70)^2) \times (651 \times 7 - (63)^2)}}$$

$$= \frac{4732 - 4410}{\sqrt{5096 - 4900} \cdot \sqrt{4557 - 3969}}$$

$$= \frac{322}{\sqrt{196 \times 588}}$$

$$= \frac{322}{339.48}$$

$$= +0.95$$

$$\boxed{r = +0.95}$$

Regression :

According to Taro Yamane, "One of the most frequently used techniques in economics and business research, to find a relation between two or more variable that are related casually, is regression analysis.

Uses of Regression Analysis :

- Regression Analysis predicts the value of dependent variables from the values of independent variables.
- We can calculate coefficient of correlation (r) and coefficient of determination (r^2) with the help of regression coefficient.

Mathematical properties :

1. The geometric mean between regression coefficient is the coefficient of correlation, Symbolically :

$$r = \sqrt{bxy \times byx}$$

It can be proved as

$$bxy = r \frac{\sigma x}{\sigma y} \; ; \; byx = r \frac{\sigma y}{\sigma x}$$

$$bxy \times byx = r \frac{\sigma x}{\sigma y} \times r \frac{\sigma y}{\sigma x} = r^2$$

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

2. Arithmetic mean of bxy and byx is equal to or greater than r.

$$\frac{bxy + byx}{2} \geq r .$$

3. Regression coefficient are independent of change of origin but not of scale.

Problem:

Determine the equation of a straight line which best fits the data

x	10	12	13	16	17	20	25
y	10	22	24	27	29	33	37

Soln

straight line $y = a + bx$

The two normal equations are:

$$\Sigma y = b\,\Sigma x + Na$$

$$\Sigma xy = b\,\Sigma x^2 + a\,\Sigma x$$

X	x^2	Y	XY
10	100	10	100
12	144	22	264
13	169	24	312
16	256	27	432
17	289	29	493
20	400	33	660
25	625	37	925
$\Sigma x = 113$	$\Sigma x^2 = 1983$	$\Sigma y = 182$	$\Sigma xy = 3186$

substituting the values,

$$\Sigma y = b\,\Sigma x + Na$$

$$\Sigma y = 182,\ \Sigma x = 113,\ N = 7 \qquad \text{—— (1)}$$

$$113b + 7a = 182$$

$$\Sigma xy = b\,\Sigma x^2 + a\,\Sigma x$$

$$\Sigma xy = 3186,\ \Sigma x^2 = 1983,\ \Sigma x = 113 \qquad \text{—— (2)}$$

$$1983b + 113a = 3186$$

Multiplying (1) by 113

$$12769b + 791a = 20566 \qquad \text{—— (3)}$$

Multiplying (2) by 7

$$13881b + 791a = 22302 \qquad \text{—— (4)}$$

subtracting (4) from (3)

$$-1,112b = -1736$$

$$1112b = 1736$$

$$b = \frac{1736}{1112} = 1.56$$

$$\boxed{b = 1.56}$$

when $b = 1.56$

$$113b + 7a = 182$$

$$113 \times 1.56 + 7a = 182$$

$$176.28 + 7a = 182$$

$$7a = 182 - 176.28$$

$$7a = 5.72$$

$$a = \frac{5.72}{7}$$

$$\boxed{a = 0.82}$$

The equation of straight line is

$$y = a + bX$$

$$a = 0.82, \quad b = 1.56$$

$$\therefore \quad y = 0.82 + 1.56X$$

The equation of the required straight line is

$$y = 0.82 + 1.56y$$

This is called regression equation of y on x.