Variance and Standard Deviation: In **probability** theory and statistics, **variance** is the expectation of the squared deviation of a random variable from its mean. Informally, it measures how far a set of numbers is spread out from their average value and is denoted by σ^2.

$$\sigma^2 = E[(x-\mu)^2]$$

The positive square root of the variance is called the standard deviation and is given by,

$$\sigma = \sqrt{E[(x-\mu)^2]}$$

If x is discrete: $\sigma^2 = E[(x-\mu)^2] = \sum_{i=1}^{n} (x_i-\mu)^2 \, P(x)$

If x is continuous: $\sigma^2 = E[(x-\mu)^2] = \int (x-\mu)^2 \, f(x)dx$

Ex. 1) Find the variance and standard deviation of a random variable x having p.d.f,

$$\mathbf{f(x)} = \begin{cases} \dfrac{\mathbf{x}}{\mathbf{2}} & ; \quad \mathbf{0 < x < 2} \\ \mathbf{0} & ; \quad \textbf{otherwise} \end{cases}$$

Solution: The expected value of x is,

$$\mu = E(x) = \int_x x\,f(x)\,dx = \int_0^2 x\frac{x}{2}\,dx = \frac{1}{2}\int_0^2 x^2\,dx = \frac{1}{2}\left[\frac{x^3}{3}\right]_0^2 = \frac{1}{2}\times\frac{8}{3} = \frac{4}{3}$$

Then the variance is,

$$\sigma^2 = E[(x-\mu)^2] = E\left[\left(x-\frac{4}{3}\right)^2\right]$$

$$= \int_0^2 \left(x-\frac{4}{3}\right)^2 \frac{x}{2}\,dx = \int_0^2 \left(x^2 - 2.x.\frac{4}{3} + \frac{16}{9}\right)\frac{x}{2}\,dx$$

$$= \frac{1}{2}\int_0^2 \left(x^3 - \frac{8}{3}x^2 + \frac{16}{9}x\right)dx = \frac{1}{2}\left[\frac{x^4}{4} - \frac{8}{3}.\frac{x^3}{3} + \frac{16}{9}.\frac{x^2}{2}\right]_0^2$$

$$= \frac{1}{2}\left(4 - \frac{64}{9} + \frac{32}{9}\right) = \frac{2}{9}$$

So, the standard deviation is

$$\sigma = \sqrt{\frac{2}{9}} = \frac{\sqrt{2}}{3}$$

Characteristic Functions:

If x is discrete: $\emptyset(t) = E(e^{itx}) = \sum_{i=1}^{n} e^{itx}\ P(x)$

If x is continuous: $\emptyset(t) = E(e^{itx}) = \int e^{itx}\ f(x)dx$

Ex. 2) Find the characteristic function of a random variable x having pdf,

$$\mathbf{f(x)} = \begin{cases} \mathbf{\frac{1}{2a}} & ; \quad \mathbf{|x| < a} \\ \mathbf{0} & ; \quad \mathbf{otherwise} \end{cases}$$

Solution: The characteristic function is given by,

$$\emptyset(t) = E(e^{itx}) = \int e^{itx}\ f(x)dx = \int_{-a}^{a} e^{itx}.\frac{1}{2a}\ dx$$

$$= \frac{1}{2a}\int_{-a}^{a} e^{itx}\ dx = \frac{1}{2a}\left[\frac{e^{itx}}{it}\right]_{-a}^{a} = \frac{1}{2ait}\left[e^{ita} - e^{-ita}\right]$$

$$= \frac{1}{2ait}[(\cos at + i\sin at) - (\cos at - i\sin at)] \qquad [\because e^{i\theta} = \cos\theta + i\sin\theta \text{ and } e^{-i\theta} = \cos\theta - i\sin\theta]$$

$$= \frac{2i\sin at}{2ait} = \frac{\sin at}{at}$$

Bivariate Distribution: The distribution in which we consider two variables simultaneously for each item of the series is known as bivariate distribution.

Linear Regression: If the variable x and y in a bivariate distribution are related, we will find that the points in a scatter diagram will cluster around a curve called regression curve.

If the curve is a straight line, it is called the line of regression and the regression is known as Linear Regression.

In the case of bivariate distribution, the co-efficient of regression of y on x is denoted by β_{yx} or b_{yx} and x on y is denoted by β_{xy} or b_{xy}.

Therefore, the formula for least square regression is,

$$b = \beta_{yx} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i{}^2 - \frac{(\sum x_i)^2}{n}} \quad \text{and } b = \beta_{xy} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum y_i{}^2 - \frac{(\sum y_i)^2}{n}}$$

Then, $a = \bar{y} - b\bar{x}$

For regression equation, $y = a + bx$ [when, $x = 0, y = a$]

Correlation: A measure of intensity or degree of linear relationship between two variables is called coefficient of correlation.

Coefficient of correlation between two random variables is denoted by,

$$\rho_{xy} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left\{\sum x_i{}^2 - \frac{(\sum x_i)^2}{n}\right\}\left\{\sum y_i{}^2 - \frac{(\sum y_i)^2}{n}\right\}}}$$

Ex. 3) Per week weight (in gm)of a calf from its birth is given below:

Age (x)	1	2	3	4	5	6	7	8	9	10
Weight (y)	52.5	58.7	65	70.2	75.4	81.1	87.2	95.5	102.2	108

Estimate the least square regression of weight on age and also estimate the weight when the age is 6.5 weeks.

Solution: Estimation of regression equation:

x	y	x^2	y^2	xy
1	52.5	1	2756.25	52.5
2	58.7	4	3445.69	117.4
3	65	8	4225	195
4	70.2	16	4928.04	280.8
5	75.4	25	5685.16	377
6	81.1	36	6577.21	486.6
7	87.2	49	7603.84	610.4
8	95.5	64	9120.25	764
9	102.2	81	10444.84	919.8
10	108	100	11664	1080
$\sum x = 55$	$\sum y = 795.8$	$\sum x^2 = 385$	$\sum y^2 = 66450.28$	$\sum xy = 4883.5$

Now, $n = 10$

$$\bar{x} = \frac{\sum x}{n} = \frac{55}{10} = 5.5 \text{ and } \bar{y} = \frac{\sum y}{n} = \frac{795.8}{10} = 79.58$$

Therefore, least square regression of weight on age that is y on x is,

$$b = \beta_{yx} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i{}^2 - \frac{(\sum x_i)^2}{n}} = \frac{4883.5 - \frac{55 \times 795.8}{10}}{385 - \frac{(55)^2}{10}} = \frac{506.6}{82.5} = 6.14$$

and, $a = \bar{y} - b\bar{x} = 79.58 - 6.14 \times 5.5 = 45.81$

$\therefore$ Regression equation of weight on age is, $y = a + bx = 45.81 + 6.14x$

Now, estimated weight at the age of 6.5 week is,

$y_{6.5} = 45.81 + 6.14 \times 6.5 = 85.72$ gm

H.W:

Following marks were obtained out of 100 by 7 students:

Marks in Statistics (x):	70	66	68	71	69	65	67
Marks in Mathematics (y):	72	68	69	69	72	67	66

Estimate the least square regression of x on y and also estimate the marks in mathematics when the mark in statistics is 73. Hence find the correlation between x and y.