14. Explain the fitting of a straight line by Least Square Method. (Important)

**Solution :** Let us consider, $(x_i, y_j)$, $i = 1,2,3, \dots \dots, n$ are the $n$ sets of observations and suppose, the related relation is y = ax + b. Now we have to select $a$ and $b$ so that the straight line is the best fit to the data.
The residual at $x = x_i$ is

$$d_i = y_i - f(x_i) = y_i - (ax_i + b), i = 1,2,3, \dots, n$$

$$E = \sum_{i=1}^{n} {d_i}^2 = \sum_{i=1}^{n} [y_i - (ax_i + b)]^2$$

By the principle of least squares, E is minimum.

$$\frac{\partial E}{\partial a} = 0 \; and \; \frac{\partial E}{\partial b} = 0$$

i.e., $2\sum[y_i - (ax_i + b)]\,(-x_i) = 0 \; and \quad 2\sum[y_i - (ax_i + b)]\,(-1) = 0$

i.e., $\sum_{i=1}^{n}(x_i y_i - a{x_i}^2 - bx_i) = 0 \; and \quad \sum_{i=1}^{n}(y_i - ax_i - b) = 0$

i.e., $a\sum_{i=1}^{n} {x_i}^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i\, y_i$ ………. (1)

and, $a\sum_{i=1}^{n} x_i + nb = \sum_{i=1}^{n} y_i$ ……….. (2)

Since, $x_i, y_j$ are known, equations (1) & (2) give two equations in $a$ & $b$. Solving for $a$ & $b$ from (1) & (2) & we obtain the best fit $y = ax + b$.

**Important Note:**

(A) Equations (1) & (2) are called normal equations.

(B) Removing suffixes $i$ from (1) & (2), the normal equations are

$$a\sum x + nb = \sum y \;\; \& \;\; a\sum x^2 + b\sum x = \sum xy$$

These equations are obtained by taking $\sum on$ both sides of $y = ax + b$ and by taking $\sum on$ both sides after multiplying by $x$ on both sides of $y = ax + b$.

(C) Transformation like $X = \frac{x-a}{h}, Y = \frac{y-b}{h}$ reduce the linear equation $y = ax + b$ to the form Y = AX + B. Hence, a linear fit is another linear fit in both systems of co-ordinates.

15. By the method of least squares find the straight line to the data given below :

| | | | | | |
|---|---|---|---|---|---|
| x | 5 | 10 | 15 | 20 | 25 |
| y | 16 | 19 | 23 | 26 | 30 |

**Solution :** Let us consider, the straight line is, $y = ax + b$.

The normal equations are :

$a\sum x + 5b = \sum y$ …………(1)

$a\sum x^2 + b\sum x = \sum xy$ …………(2)

Now, we have to calculate $\sum x$, $\sum y$, $\sum x^2$, $\sum xy$ and so we form the following table

| | $x$ | $y$ | $x^2$ | $xy$ |
|---|---|---|---|---|
| | 5 | 16 | 25 | 80 |
| | 10 | 19 | 100 | 190 |
| | 15 | 23 | 225 | 345 |
| | 20 | 26 | 400 | 520 |
| | 25 | 30 | 625 | 750 |
| **Total** | 75 | 114 | 1375 | 1885 |

The normal equations are : 75a + 5b = 114 ………..(3)

1375a + 75b = 1885 ……(4)

Solving (3) and (4), we get, a = 0.7, b = 12.3

Hence, the best fitting line is y = 0.7$x$ + 12.3 ………… (P)

**By Second Form :**

Let us consider, $X = \frac{x-a}{h} = \frac{x-15}{5}$ where, $a$ = 15 (middle point of the column $x$)

$Y = \frac{y-b}{h} = \frac{y-23}{5}$ where, $b$ = 23 (middle point of the column $y$)

Let us take, the line in new variable is : $Y = AX + B$

| | $x$ | $y$ | $X$ | $X^2$ | $Y$ | $XY$ |
|---|---|---|---|---|---|---|
| | 5 | 16 | $-2$ | 4 | $-1.4$ | 2.8 |
| | 10 | 19 | $-1$ | 1 | $-0.8$ | 0.8 |
| | 15 | 23 | 0 | 0 | 0 | 0 |
| | 20 | 26 | 1 | 1 | 0.6 | 0.6 |
| | 25 | 30 | 2 | 4 | 1.4 | 2.8 |
| **Total** | | | 0 | 10 | $-0.2$ | 7 |

The normal equations are : $A\sum X + 5B = \sum Y$ ……….(5)

$A\sum X^2 + B\sum X = \sum XY$ ……(6)

Solving (5) and (6), we get, A = 0.7 and B = $-0.04$

∴ $Y = 0.7X - 0.04$

⇨ $\frac{y-23}{5} = 0.7 \times \frac{x-15}{5} - 0.04 = y - 23 = 0.7x - 10.5 - 0.2$

⇨ $y = 0.7x + 33.3$ ………………..…………….(Q)

Thus, the equations (P) and (Q) are same.

16. Fit a straight line to the data given below. Also estimate the value of y at x =2.5

| x | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| y | 1 | 1.8 | 3.3 | 4.5 | 6.3 |

**Solution :** Let us consider, the straight line is, $y = ax + b$.
The normal equations are :

$a\sum x + 5b = \sum y$ ……………….(1)

$a\sum x^2 + b\sum x = \sum xy$ ……………….(2)

| | $x$ | $y$ | $x^2$ | $xy$ |
|---|---|---|---|---|
| | 0 | 1 | 0 | 0 |
| | 1 | 1.8 | 1 | 1.8 |
| | 2 | 3.3 | 4 | 6.6 |
| | 3 | 4.5 | 9 | 13.5 |
| | 4 | 6.3 | 16 | 25.2 |
| Total | 10 | 16.9 | 30 | 47.1 |

From the above table, the equations can be re-written as

$$10a + 5b = 16.9 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (3)$$

$$30a + 10b = 47.1 \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

Solving (3) and (4), we get,

$$a = 1.33, \quad b = 0.72$$

Hence, the equation will be,

$$y = ax + b = 1.33x + 0.72$$

At $x = 2.5$, $\quad y = 1.33 \times 2.5 + 0.72 = 4.045$

[ These two problems (Q. Nos. 15 & 16) are very important.]