Class Name - M. Sc. (Statistics)

Semester - IV

Title of Lecture - Linear Model (Two-Variable)

Teacher Name - Dr. Rajesh Tailor

Reader,

S. S. in Statistics,

V. U. Ujjain

Let Y and X be two variables such that
X is thought to influence the variable Y.
Here | Two-variable linear model]

$$Y = f(X) \qquad (1)$$

The simplest relationship between two variables is a linear one, namely

$$Y = \alpha + \beta X \qquad (2)$$

where α and β are unknown parameters indicating the intercept and slope of the function.

In practice the line $\alpha + \beta x$ is unknown. Instead we have a set of hypothesis

$$Y_i = \alpha + \beta x_i + u_i \qquad i = 1, 2, \ldots, n$$

$$E(u_i) = 0 \qquad \text{for all } i$$

$$E(u_i u_j) = \begin{cases} 0 & \text{for } i \neq j;\ i, j = 1, 2, \ldots n \\ \sigma_u^2 & \text{for } i = j;\ i, j = 1, 2, \ldots, n \end{cases}$$

We wish to estimate α β and σ_u^2 statistically on the basis of sample observations on X and Y..

* (2) may be written as

$$Y = \alpha + \beta X + u$$

where u denotes a variable which may take on positive or negative values.

Here u is a variable with a probability distribution centered at zero and having a finite variance σ_u^2. This is why u is referred to as a stochastic disturbance (or error) term. In view of many factors involved and their likely independence, an appeal to Central Limit Theorem would suggest a normal distribution for u.

Least-Squares Estimators :-

Let us consider the sample observations

$x_1, x_2, \ldots, x_n$ and

$Y_1, Y_2, \ldots, Y_n$ with means $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

Consider a two-variable linear model

$$Y = \alpha + \hat{\beta} x \quad (*)$$

We wish to pass a linear line through the observations as our estimate of the true line $\alpha+\beta x$. Denote the estimated line by $\hat{Y} = \hat{\alpha} + \hat{\beta}x$ --- (1), where $\hat{\alpha}$ and $\hat{\beta}$ are estimates of unknown parameters α and β and $\hat{Y}$ is the ordinate on the line for any given value of x. To fit such a line we must develop formulae for $\hat{\alpha}$ and $\hat{\beta}$ in terms of the sample observations.

For i^{th} sample observation $Y_i = \alpha+\beta x_i$ and $\hat{Y}_i = \hat{\alpha}+\hat{\beta}x_i$

Let $e_i = Y_i - \hat{Y}_i$

i.e. $\sum_{i=1}^{n} e_i^2 = f(\hat{\alpha}, \hat{\beta})$

The principal of least squares is that the $\hat{\alpha}$ and $\hat{\beta}$ values should be chosen so as to make Σe_i^2 as small as possible. A necessary condition is that the partial derivatives of the sum with respect to $\hat{\alpha}$ and $\hat{\beta}$ should be zero. We thus have

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$= \sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2$$

so that

$$\frac{\partial}{\partial \hat{\alpha}}\left(\sum_{i=1}^{n} e_i^2\right) = -2\sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta} X_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} Y_i = n\hat{\alpha} + \hat{\beta}\sum_{i=1}^{n} X_i \quad \text{---}(*)$$

and

$$\frac{\partial}{\partial \hat{\beta}}\left(\sum_{i=1}^{n} e_i^2\right) = -2\sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta} X_i)X_i = 0$$

$$\Rightarrow \sum_{i=1}^{n} Y_i X_i = \hat{\alpha}\sum_{i=1}^{n} X_i + \hat{\beta}\sum_{i=1}^{n} X_i^2 \quad \text{---}(3)$$

When the indicated values from the sample observations are inserted, they give two simultaneous equations, which may be solved for $\hat{\alpha}$ and $\hat{\beta}$

Devide by n, we have $\bar{Y} = \hat{\alpha} + \hat{\beta}\bar{X}$ --- (4)

Subtract (4) from 1

$$\hat{Y} - \bar{Y} = \hat{\beta}(X - \bar{X}) \quad \text{--}(5)$$

If $x_i = X_i - \bar{X}$ and $y_i = Y_i - \bar{Y}$ then $\hat{y}_i = \hat{Y}_i - \bar{Y}$

Now (5) i.e. least square equation can be written as

$$\hat{y} = \hat{\beta} x$$

Now e_i can be indicated by

$$e_i = y_i - \hat{y}_i = y_i - \hat{\beta} x_i$$

So that the some of squared residuals is

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta} x_i)^2 \quad \text{-----} \quad 6$$

Minimizing (6) with respect to β i.e

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}} = 0 \Rightarrow \sum_{i=1}^{n} (y_i - \hat{\beta} x_i) x_i = 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i x_i = \hat{\beta} \sum_{i=1}^{n} x_i^2$$

$$\Rightarrow \boxed{\hat{\beta} = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}}$$

and

$$\frac{\partial^2 (\Sigma e^2)}{\partial \hat{\beta}^2} = 2 \sum_{i=1}^{n} x_i^2$$ which is positive, so that

true minimum has been obtained. Thus least square estimate of β is $\hat{\beta} = \sum_{i=1}^{n} y_i x_i / \sum_{i=1}^{n} x_i^2$

The intercept $\hat{\alpha}$ is then obtained from the condition that the line pass through the point of means, namely $\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$ (Least square estimate of α)