PBD-2802 Advanced Statistical Methods

Unit-2

Linear Estimation

INTRODUCTION:-

It is often of interest to learn about the characteristics of a large group of elements such as individuals, households, buildings, products, parts, customers, and so on. All the elements of interest in a particular study form the population. Because of time, cost, and other considerations, data often cannot be collected from every element of the population. In such cases, a subset of the population, called a sample, is used to provide the data. Data from the sample are then used to develop estimates of the characteristics of the larger population. The process of using a sample to make *inferences* about a population is called statistical inference.

In most statistical studies the population parameters are unknown and must be estimated from a sample because it is impossible to look at the entire population since it requires lot of time and expense.

One of the main objectives of the theory of statistics is to draw inferences about a population from the analysis of a sample drawn from the population. The two important aspects of statistical inference are:-

- **Theory of Point Estimation**
- **Theory of testing of Hypothesis.**

The theory of estimation was founded by Prof. R. A. Fisher in 1930.

The problem while estimating the parameter is to find the criteria for judging how well the given sample statistic estimates the population parameter. Mathematically these can be described as follows:-

Let us consider a random sample $x_1, x_2, \dots, x_n$ of size n from a population, the distribution of which has a known mathematical form say, $f(x; \theta_1, \theta_2, \dots, \theta_k)$ where $\theta_1, \theta_2, \dots, \theta_k$ are unknown population parameters.

Then there will always be an infinite number of functions of sample values which may be used to estimate one or more parameters. Out of these numbers of estimators available we would be interested to find the best estimator which will fall nearer to the true value of the parameter.

ESTIMATOR: - Any function of a random sample say $t = t(x_1, x_2, \dots, x_n)$ which is used to estimate the unknown parameter θ of the distribution it is called an estimator.

A specific value of the estimators is called its **"ESTIMATES".**

For e.g.:- If $\bar{x} = 10$ and $s^2 = 4.5$ then $\bar{x}$ is a estimator and its estimate is 10. Similarly s^2 is an estimator and 4.5 is its estimate.

Estimation of a parameter by a single value is called **POINT ESTIMATION** while estimation of a parameter by an interval is called **INTERVAL ESTIMATION.**

Example Let $X_1, X_2, \dots, X_n$ be a random sample from any distribution F_θ for which the mean exists and is equal to θ. We may want to estimate the mean θ of distribution. For this purpose, we may compute the mean of the observations $x_1, x_2, \dots, x_n$, i.e., say

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

This $\bar{x}$ can be taken as the point estimate of θ.

Example Let $X_1, X_2, \dots, X_n$ be a random sample from Poisson's distribution with parameter λ, i.e., $P(\lambda)$, where λ is not known. Then the mean of the observations $x_1, x_2, \dots, x_n$, i.e.,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

is a point estimate of λ.

Example Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with parameters μ and σ^2, i.e., $N(\mu, \sigma^2)$, where both μ and σ^2 are unknown. μ and σ^2 are the mean and variance respectively of the normal distribution. In this case, we may take a joint statistics $(\bar{x}, s^2)$ as a point estimate of $N(\mu, \sigma^2)$, where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \text{sample mean}$$

and

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_1 - \bar{x})^2 = \text{sample mean square.}$$

In this chapter we shall consider only unbiased estimators of estimable linear functions for the expectations of the random variables. We may have more than one unbiased estimator for estimating the same parameter. E.g. $\bar{y}$ and $\sum_i a_i y_i$ with $\sum_i a_i = 1$ are both unbiased linear estimators of population mean. So, we have the problem of selecting the one that may be taken to be the best, in some suitable sense.

One widely used criterion to make this selection is to choose that estimator which has smallest variance. Such an estimator is known as a minimum variance unbiased linear estimator. This approach of minimum variance linear unbiased estimator was given by Markov(1900).

*Minimum Variance Unbiased Estimator:(MVUE)

If a statistic $T = T(x_1, x_2, \ldots, x_n)$ *based on sample of size* n *is such that:*
(i) T is unbiased for $\gamma(\theta)$, *and*
(ii) It has the smallest variance among the class of all unbiased-estimators of $\gamma(\theta)$,

then Tis called the minimum variance unbiased estimator (MVUE) $\gamma(\theta)$.

More precisely, T is MVUE of $\gamma(\theta)$ if

$$E_\theta(T) = \gamma(\theta) \quad \text{for all } \theta \in \Theta$$

and

$$\mathrm{Var}_\theta(T) \le \mathrm{Var}_\theta(T') \quad \text{for all } \theta \in \Theta$$

where T' is any other unbiased estimator of $\gamma(\theta)$.

* Gauss- Markov Linear Model:

Consider a set of n independent random variables $Y_1, Y_2, \ldots, Y_n$ with a common variance σ^2, whose expectations are linear functions with known coefficients(a_{ij}'s) of p unknown parameters $\beta_1, \beta_2, \ldots, \beta_p$ ($p < n$).

Thus,
$$\left.\begin{aligned} E(Y_i) &= a_{i1}\beta_1 + a_{i2}\beta_2 + \ldots.. + a_{ip}\beta_p \\ V(Y_i) &= \sigma^2 \text{ for } i = 1, 2, \ldots.., n \\ Cov(Y_i, Y_j) &= 0 \text{ for } i \neq j \end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots\ldots(1)$$

Result-(1) is called the **Gauss- Markov linear model**.

In matrix formulation the above model can be written as-

$$\underset{\sim}{Y} = A\underset{\sim}{\beta} + \underset{\sim}{e}$$

Where $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_n \end{pmatrix}_{nx1}$ is known, and $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \ldots \\ \beta_p \end{pmatrix}_{px1}$,which is unknown.

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1p} \\ a_{21} & a_{22} & \ldots & a_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{np} \end{pmatrix}_{n \text{ x p}}$$ is called data matrix which is known.

$$\underset{\sim}{e} = \begin{pmatrix} e_1 \\ e_2 \\ \ldots \\ e_n \end{pmatrix}_{n \text{ x } 1}$$ = Error or residual vector, and it is unknown.

With $E(\underset{\sim}{e}) = 0$ and $V(\underset{\sim}{e}) = \sigma^2$

In the above model given in (1), the unknown parameters β_1, β_2,….., β_p are called effects. In linear estimation the effects are all fixed quantities and such a model is where all the effects are unknown parameters are called a fixed effect model. Sometimes one of the β_j's is a constant, with a_{ij}'s are 'indicator variables' taking values 0 or 1 usually. Such a model is known as an 'Analysis of Variance' model. If a_{ij}'s are values taken not by indicator variables but by independent variable then we have a regression model. A model in which both types of a_{ij}'s are present is called analysis of co-variance model.

***Gauss- Markov Theorem: (Only Statement)**

For the p-variate general linear model $\underset{\sim}{Y} = A\underset{\sim}{\beta} + \underset{\sim}{e}$,

Where $\underset{\sim}{Y} : n \times 1, A : n \times p, \underset{\sim}{\beta} : p \times 1, \underset{\sim}{e} : n \times 1$

Let $\hat{\underset{\sim}{\beta}} = (A'A)^{-1} A'\underset{\sim}{Y}$ be linear estimator of $\underset{\sim}{\beta}$ with $V(\hat{\underset{\sim}{\beta}}) = (A'A)^{-1}\sigma^2$. Further if $\underset{\sim}{b} = [(A'A)^{-1} A + C]\underset{\sim}{Y}$ denote some other unbiased estimator of $\underset{\sim}{\beta}$ and $\underset{\sim}{C} : p \times p$ is any scalar matrix then for any scalar vector $\underset{\sim}{G}' = (G_1, G_2, ..., G_p) : 1 \times p$

we have $V(\underset{\sim}{G}'\hat{\beta}) \leq V(\underset{\sim}{G}'\underset{\sim}{b})$

This means that $\hat{\underset{\sim}{\beta}}$ is the best linear estimator.

*Basic Assumptions in the model:

(i) $E(\underset{\sim}{e}) = 0$

ii) $\rho(A) = p < n$, i.e. rank of matrix A is p.

iii) $(A'A)^{-1}$ exists.

iv) $E(\underset{\sim}{e}'\underset{\sim}{e}) = \sigma^2 I_n$

***Linear estimate of $\underset{\sim}{\beta}$:**

Let the p-variate general linear model $\underset{\sim}{Y} = A\underset{\sim}{\beta} + \underset{\sim}{e}$,

Where $\underset{\sim}{Y} : n\times 1, A : n\times p, \underset{\sim}{\beta} : p\times 1, \underset{\sim}{e} : n\times 1$ with usual assumptions.

According to the principle of least squares we obtain $\hat{\underset{\sim}{\beta}}$ such that error sum of squares is minimum.

i.e. find $\hat{\underset{\sim}{\beta}}$ such that $S^2 = \sum_{i=1}^{n} e_i^2 = \underset{\sim}{e}'\underset{\sim}{e}$ is minimum.

Now, $\underset{\sim}{e} = \underset{\sim}{Y} - A\underset{\sim}{\beta}$

$$\underset{\sim}{e}'\underset{\sim}{e} = (\underset{\sim}{Y} - A\underset{\sim}{\beta})'(\underset{\sim}{Y} - A\underset{\sim}{\beta})$$

$$S^2 = \sum_{i=1}^{n} e_i^2 = \underset{\sim}{e}'\underset{\sim}{e}$$

$$= (\underset{\sim}{Y} - A\underset{\sim}{\beta})'(\underset{\sim}{Y} - A\underset{\sim}{\beta})$$

$$= (\underset{\sim}{Y}' - \underset{\sim}{\beta}'A')(\underset{\sim}{Y} - A\underset{\sim}{\beta})$$

$$= \underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{Y}'A\underset{\sim}{\beta} - \underset{\sim}{\beta}'A'\underset{\sim}{Y} + \underset{\sim}{\beta}'A'A\underset{\sim}{\beta}$$

$$= \underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{\beta}'A'\underset{\sim}{Y} - \underset{\sim}{\beta}'A'\underset{\sim}{Y} + \underset{\sim}{\beta}'A'A\underset{\sim}{\beta}$$

Differentiating w.r.to $\underset{\sim}{\beta}$ we get $\frac{\partial S^2}{\partial \beta} = 0$

$$\Rightarrow -2A'\underset{\sim}{Y} + 2(A'A)\underset{\sim}{\beta} = 0$$

$\Rightarrow \hat{\underset{\sim}{\beta}} = (A'A)^{-1}A'\underset{\sim}{Y}$ Where $|A'A| \neq 0$ i.e. $(A'A)^{-1}$ exists.

$$= (A'A)^{-1}A'(A\beta + \underset{\sim}{e})$$

$$= (A'A)^{-1}A'A\beta + (A'A)^{-1}A'\underset{\sim}{e}$$

$$= \beta + (A'A)^{-1} A'\underset{\sim}{e}$$

$$\underset{\sim}{\hat{\beta}} - \beta = (A'A)^{-1} A'\underset{\sim}{e} \text{ ....[2]}$$

Now, $E(\underset{\sim}{\hat{\beta}}) = E[(A'A)^{-1} A'\underset{\sim}{Y}]$

$$= (A'A)^{-1} A'E(\underset{\sim}{Y})$$

$$= (A'A)^{-1} A'E(A\underset{\sim}{\beta} + \underset{\sim}{e})$$

$$= (A'A)^{-1} A'A\underset{\sim}{\beta} + E(\underset{\sim}{e})$$

$$= \underset{\sim}{\beta}$$

Now, V($\underset{\sim}{\hat{\beta}}$) = $E[(\underset{\sim}{\hat{\beta}} - E(\underset{\sim}{\hat{\beta}}))(\underset{\sim}{\hat{\beta}} - E(\underset{\sim}{\hat{\beta}}))']$

$$= E[(A'A)^{-1} A'\underset{\sim}{e})(A'A)^{-1} A'\underset{\sim}{e})']$$

$$= E[(A'A)^{-1} A'A\underset{\sim}{e}\underset{\sim}{e}'(A'A)^{-1}]$$

$$= (A'A)^{-1} A'AE(\underset{\sim}{e}\underset{\sim}{e}')(A'A)^{-1}$$

$$= \sigma^2 I_n (A'A)^{-1}$$

Properties of $\underset{\sim}{\hat{\beta}}$:

1) It is a linear function of the observations.
2) It is an unbiased estimator of $\underset{\sim}{\beta}$
3) $V(\underset{\sim}{\hat{\beta}}) = \sigma^2 I_n (A'A)^{-1}$
4) $\underset{\sim}{\hat{\beta}}$ is BLUE(Best Linear Unbiased Estimator)