INTRODUCTION TO DATA ANALYTICS

Class #12

Estimation

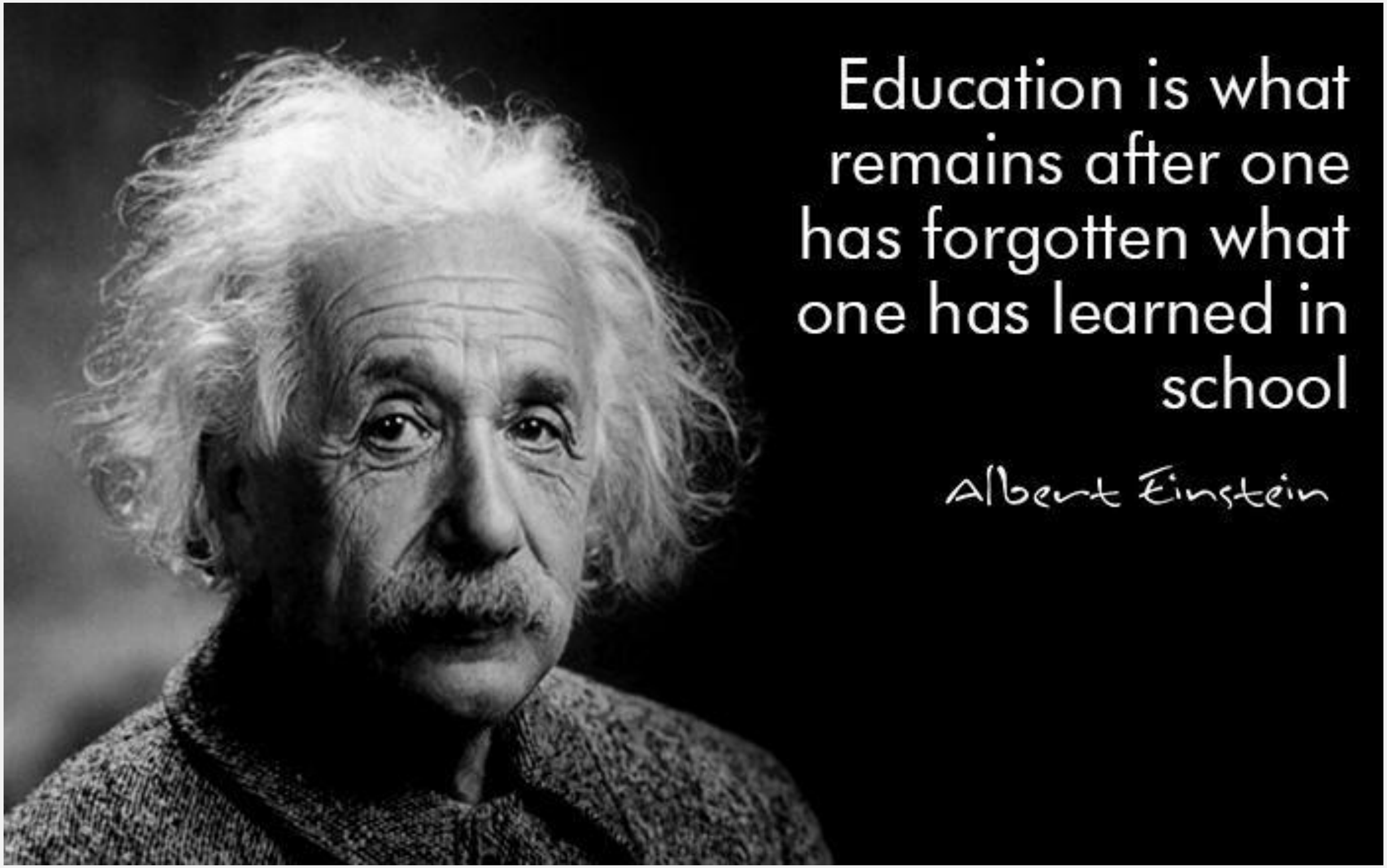
Dr. Sreeja S R

Assistant Professor

Indian Institute of Information Technology

IIIT Sri City

QUOTE OF THE DAY..



IN THIS PRESENTATION…

- POINT ESTIMATION
- GENERAL CONCEPTS OF POINT ESTIMATION
  - **Unbiased Estimators**
  - **Variance of a Point Estimator**
  - **Standard Error: Reporting a Point Estimate**
  - **Mean Squared Error of an Estimator**
- METHODS OF POINT ESTIMATION
  - **Method of Moments**
  - **Method of Maximum Likelihood**

POINT ESTIMATION

- Statistical inference is always focused on drawing conclusions about one or more parameters of a population.
- An important part of this process is obtaining estimates of the parameters.
- Suppose that we want to obtain a point estimate (a reasonable value) of a population parameter.
- Any function of the observation, or any statistic, is also a random variable. For example, the sample mean and the sample variance are statistics and they are also random variables.
- Since a statistic is a random variable, it has a probability distribution. We call the probability distribution of a statistic a sampling distribution.

POINT ESTIMATION

- When discussing inference problems, it is convenient to have a general symbol to represent the parameter of interest.
- We will use the Greek symbol θ (theta) to represent the parameter. The symbol θ can represent the mean μ, the variance σ^2, or any parameter of interest to us.
- The objective of point estimation is to select a single number, based on sample data, that is the most plausible value for θ.
- A numerical value of a sample statistic will be used as the point estimate. The statistic $\widehat{\Theta}$ is called a **point estimator** of θ.

POINT ESTIMATOR

Definition 12.1: **Point Estimator**

A **point estimate** of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic $\widehat{\Theta}$ .The statistic $\widehat{\Theta}$ is called the **point estimator.**

- **Example:**
- Suppose that the random variable X is normally distributed with an unknown mean μ. The sample mean is a point estimator of the unknown population mean μ.That is, $\hat{\mu} = \bar{X}$.

POINT ESTIMATOR

Estimation problems occur frequently in engineering. We often need to estimate

- The mean μ of a single population
- The variance σ^2 (or standard deviation σ) of a single population
- The proportion p of items in a population that belong to a class of interest.
- The difference in means of two populations, $\mu_1 - \mu_2$
- The difference in two population proportions, $p_1 - p_2$

POINT ESTIMATOR

Reasonable point estimates of these parameters are as follows:

- For μ, the estimate is $\hat{\mu} = \bar{X}$, the sample mean.
- For σ^2, the estimate is $\hat{\sigma}^2 = s^2$, the sample variance.
- For p, the estimate is $\hat{p} = x/n$, the sample proportion, where x is the number of items in a random sample of size n that belong to the class of interest.
- For $\mu_1 - \mu_2$, the estimate is $\hat{\mu}_1 - \hat{\mu}_2 = \hat{x}_1 - \hat{x}_2$, the difference between the sample means of two independent random samples.
- For $p_1 - p_2$, the estimate is $\hat{p}_1 - \hat{p}_2$, the difference between two sample proportions computed from two independent random samples.

CENTRAL LIMIT THEOREM

Definition 12.2: **Approximate Sampling Distribution of a Difference in Sample Means**

If we have two independent populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, and if $\bar{X}_1$ and $\bar{X}_2$ are the sample means of two independent random samples of sizes n_1 and n_2 from these populations, then the sampling distribution of

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

is approximately standard normal, if the conditions of the central limit theorem apply. If the two populations are normal, the sampling distribution of Z is exactly standard normal.

UNBIASED ESTIMATORS

An estimator should be "close" in some sense to the true value of the unknown parameter. Formally, we say that $\widehat{\Theta}$ is an unbiased estimator of θ if the expected value of $\widehat{\Theta}$ is equal to θ. This is equivalent to saying that the mean of the probability distribution of $\widehat{\Theta}$ (or the mean of the sampling distribution of $\widehat{\Theta}$) is equal to θ.

Definition 12.3: **Bias of an Estimator**

The point estimator $\widehat{\Theta}$ is an **unbiased estimator** for the parameter θ if

$$E\left(\widehat{\Theta}\right) = \theta$$

If the estimator is not unbiased, then the difference

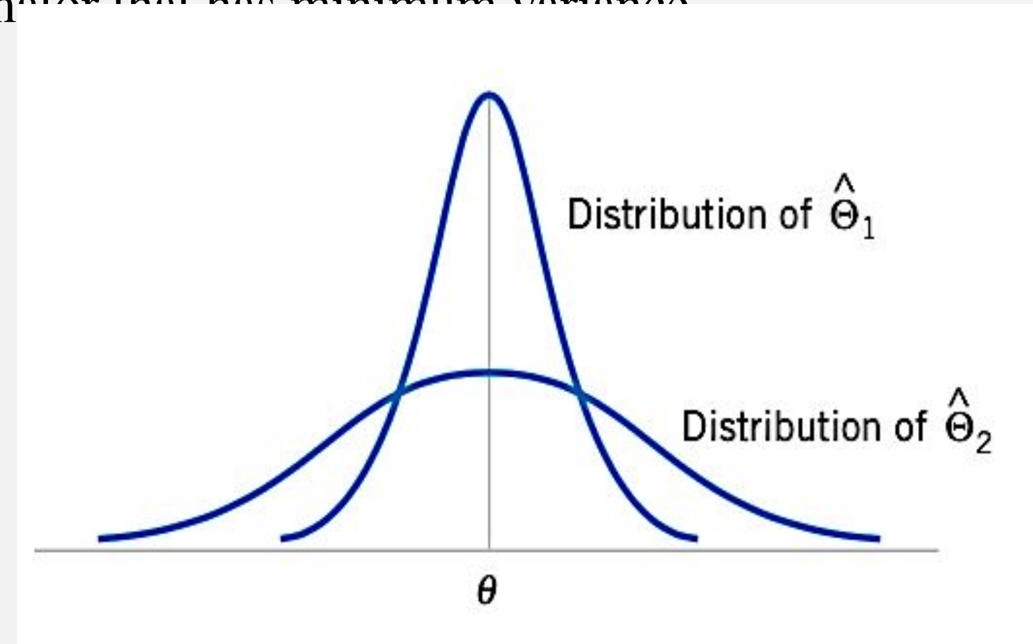
$$E\left(\widehat{\Theta}\right) - \theta$$

is called the **bias** of the estimator $\widehat{\Theta}$.

When an estimator is unbiased, the bias is zero; that is, $E\left(\widehat{\Theta}\right) - \theta = 0$.

VARIANCE OF A POINT ESTIMATOR

A logical principle of estimation, when selecting among several estimators, is to choose the estimator that has minimum variance.



If we consider all unbiased estimators of θ, the one with the smallest variance is called the **minimum variance unbiased estimator** (MVUE).

STANDARD ERROR

When the numerical value or point estimate of a parameter is reported, it is usually desirable to give some idea of the precision of estimation. The measure of precision usually employed is the standard error of the estimator that has been used.

The standard error of an estimator $\widehat{\Theta}$ is its standard deviation, given by $\sigma_{\widehat{\Theta}} = \sqrt{V(\widehat{\Theta})}$. If the standard error involves unknown parameters that can be estimated, substitution of those values into $\sigma_{\widehat{\Theta}}$ produces an **estimated standard error,** denoted by $\hat{\sigma}_{\widehat{\Theta}}$.

Sometimes the estimated standard error is denoted by $s_{\widehat{\Theta}}$ or $se(\widehat{\Theta})$.

MEAN SQUARED ERROR OF AN ESTIMATOR

Sometimes it is necessary to use a biased estimator. In such cases, the mean squared error of the estimator can be important. The **mean squared error** of an estimator $\widehat{\Theta}$ is the expected squared difference between $\widehat{\Theta}$ and θ.

The mean squared error of an estimator $\widehat{\Theta}$ of the parameter θ is defined as

$$MSE(\widehat{\Theta}) = E(\widehat{\Theta} - \theta)^2$$

The mean squared error can be rewritten as follows:

$$\begin{aligned} MSE(\widehat{\Theta}) &= E[\widehat{\Theta} - E(\widehat{\Theta})]^2 + [\theta - E(\widehat{\Theta})]^2 \\ &= V(\widehat{\Theta}) + (bias)^2 \end{aligned}$$

That is, the mean squared error of $\widehat{\Theta}$ is equal to the variance of the estimator plus the squared bias. If $\widehat{\Theta}$ is an unbiased estimator of θ, the mean squared error of $\widehat{\Theta}$ is equal to the variance o f $\widehat{\Theta}$.

MEAN SQUARED ERROR OF AN ESTIMATOR

The mean squared error is an important criterion for comparing two estimators. Let $\widehat{\Theta}_1$ and $\widehat{\Theta}_2$ be two estimators of the parameter θ, and let $MSE\left(\widehat{\Theta}_1\right)$ and $MSE\left(\widehat{\Theta}_2\right)$ be the mean squared errors of $\widehat{\Theta}_1$ and $\widehat{\Theta}_2$. Then the **relative efficiency** of $\widehat{\Theta}_2$ to $\widehat{\Theta}_1$ is defined as

$$\frac{MSE\left(\widehat{\Theta}_1\right)}{MSE\left(\widehat{\Theta}_2\right)}$$

If this relative efficiency is less than 1, we would conclude that $\widehat{\Theta}_1$ is a more efficient estimator of θ than $\widehat{\Theta}_2$, in the sense that it has a smaller mean squared error.

METHODS OF POINT ESTIMATION

The definitions of unbiasedness and other properties of estimators do not provide any guidance about how good estimators can be obtained.

Methods for obtaining point estimators:

- The method of moments
- The method of maximum likelihood.

METHOD OF MOMENTS

The general idea behind the method of moments is to equate **population moments,** which are defined in terms of expected values, to the corresponding **sample moments.** The population moments will be functions of the unknown parameters. Then these equations are solved to yield estimators of the unknown parameters.

Definition 12.4: Moments

Let $X_1, X_1, \dots, X_n$ be a random sample from the probability distribution $f(x)$, where $f(x)$ can be a discrete probability mass function or a continuous probability density function. The kth **population moment (or distribution moment)** is $E\left(X^k\right), k = 1,2,\dots$. The corresponding kth sample moment is $\left(\frac{1}{n}\right)\sum_{i=1}^{n} X_i^k, k = 1,2,\dots$.

METHOD OF MOMENTS

To illustrate, the first population moment is $E(X) = \mu$, and the first sample moment is $\left(\frac{1}{n}\right)\sum_{i=1}^{n} X_i = \bar{X}$. Thus by equating the population and sample moments, we find that $\hat{\mu} = \bar{X}$. That is, the sample mean is the **moment estimator** of the population mean. In the general case, the population moments will be functions of the unknown parameters of the distribution, say, $\theta_1, \theta_2, \ldots, \theta_m$.

Definition 12.5: **Moment Estimators**

Let $X_1, X_2, \ldots, X_n$ be a random sample from either a probability mass function or a probability density function with m unknown parameters $\theta_1, \theta_2, \ldots, \theta_m$. The **moment estimators** $\widehat{\Theta}_1, \widehat{\Theta}_2, \ldots, \widehat{\Theta}_m$ are found by equating the first m population moments to the first m sample moments and solving the resulting equations for the unknown parameters.

METHOD OF MAXIMUM LIKELIHOOD

One of the best methods of obtaining a point estimator of a parameter is the method of maximum likelihood. This technique was developed in the 1920s by a famous British statistician, Sir R. A. Fisher. As the name implies, the estimator will be the value of the parameter that maximizes the likelihood function.

Definition 12.6: **Maximum Likelihood Estimator**

Suppose that X is a random variable with probability distribution $f\,(x;\,\theta)$, where θ is a single unknown parameter. Let $x_1, x_2, \dots, x_n$ be the observed values in a random sample of size *n*. Then the **likelihood function** of the sample is

$$L(\theta) = f\,(x_1;\,\theta) \cdot f\,(x_2;\,\theta) \dots \cdot f\,(x_n;\,\theta)$$

Note that the likelihood function is now a function of only the unknown parameter θ . The **maximum likelihood estimator** (MLE) of θ is the value of θ that maximizes the likelihood function $L(\theta)$.

METHOD OF MAXIMUM LIKELIHOOD

In the case of a discrete random variable, the interpretation of the likelihood function is simple. The likelihood function of the sample $L(\theta)$ is just the probability

$$P\,(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

That is, $L(\theta)$ is just the probability of obtaining the sample values $x_1, x_2, \dots, x_n$. Therefore, in the discrete case, the maximum likelihood estimator is an estimator that maximizes the probability of occurrence of the sample values.

PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATOR

The method of maximum likelihood is often the estimation method that mathematical statisticians prefer, because it produces estimators with good statistical properties.

Under very general and not restrictive conditions, when the sample size n is large and if $\widehat{\Theta}$ is the maximum likelihood estimator of the parameter θ,

1. $\widehat{\Theta}$ is an approximately unbiased estimator for θ, $[\mathrm{E}(\widehat{\Theta}) \approx \theta]$
2. the variance of $\widehat{\Theta}$ is nearly as small as the variance that could be obtained with any other estimator, and
3. $\widehat{\Theta}$ has an approximate normal distribution.

Any question?