# UNIT - III

## Estimation and Test of Hypothesis

### Introduction:

A statistical inference, we use the sample and apply a suitable statistical method for drawing conclusions about the unknown properties of the population so that, we obtain the answer to a problem.

There are two types of problems under statistical inference

1) Estimation
2) Testing of Hypothesis

**Estimate:** An estimate is a statement made to find an unknown population parameter.

**Estimator:** The procedure (or) rule to determine an unknown population parameter is called an Estimator.

**Eg:** Sample mean is an estimator of Population mean.

**Estimation:** To use the statistics, obtained from the samples as an estimate of the unknown Parameter of the population from which the sample is drawn.

There are two types of estimation

1) Point estimation
2) Interval estimation

* Estimation means approximating a quantity to the required accuracy.

**Point Estimation:** If an estimate of the population parameter is given by a single value, then the estimate is called a point estimation of the parameter.

**Eg:** If the height of a student is measured as 162 cm then the measurement gives a point estimation.

* A point estimate of a parameter ($\theta$) is a single numerical value which is computed from a given sample and acts as an approximation of the unknown exact value of the parameter.

* A point estimator is denoted by '$\hat{\theta}$'.

**Interval Estimation:** An interval estimate of a population parameter ($\theta$) is an interval of the form $\hat{\theta}_L < \theta < \hat{\theta}_u$ where $\hat{\theta}_L$ and $\hat{\theta}_u$ depend on the value of the statistic $\hat{\theta}$ for a particular sample and also on the sampling distribution of $\hat{\theta}$.

* If an estimate of a population parameter is given by two different values between which the parameter may be considered to lie, then the estimate is called an interval estimation of the parameter.

**Eg:** If the height of the student is given as $(163 \pm 3.5)$ cm Then the height lies between 159.5 cms and 166.5 cms and the measurement gives an interval estimation.

**Unbiased Estimator:** A point estimator ($\hat{\theta}$) is said to be unbiased estimator of the parameter ($\theta$) if $\boxed{E(\hat{\theta}) = \theta}$

In other words, $\boxed{E(\text{statistic}) = \text{parameter}}$ then the statistics is said to be unbiased estimator of the parameter.

**Biased Estimator:** A point estimator ($\hat{\theta}$) is said to be biased estimator if $\boxed{E(\hat{\theta}) \neq \theta}$ i.e

$\boxed{E(\text{Statistic}) \neq \text{Parameter}}$

Parameter → Population → $\mu, \sigma^2$

Statistic → Sample → $\bar{x}, s^2$

**Properties of estimators:**

* A good estimator is one which is closed to the true value of the parameter.

* The important properties are

1) Consistency
2) Unbiasedness
3) Efficiency
4) Sufficiency

**Theorem:** The sample mean ($\bar{x}$) is an unbiased estimator of the population mean ($\mu$)

**Proof:** Let $x_1, x_2, \ldots x_n$ be a ~~[illegible]~~ random sample drawn from a given population with mean $\mu$

To prove that $E(\bar{x}) = \mu$

Consider, $E(\bar{x}) = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right]$

$$= E.\frac{1}{n}\left[\sum_{i=1}^{n} x_i\right]$$

$$= \frac{1}{n} E[x_1 + x_2 + \cdots + x_n]$$

$$= \frac{1}{n}\left[E(x_1) + E(x_2) + \cdots + E(x_n)\right]$$

$$= \frac{1}{n}[\mu + \mu + \mu + \cdots + \mu]$$

$$= \frac{1}{\cancel{n}}(\cancel{n}\mu)$$

$$E(\bar{x}) = \mu$$

Hence Proved

$\boxed{15/11/21}$

**Example:** The mean of a random sample is an unbiased estimate of the mean of the population 3, 6, 9, 15, 27

a) List all possible samples of size 3 that can be taken without replacement from the finite Population

b) Calculate the mean of each sample & assign each sample a probability of $\frac{1}{10}$. Verify that the mean of the sample $\bar{x} = 12$. Which is equal to the mean of the population $\theta$ i.e $E(\bar{x}) = \theta$

Prove that $\bar{x}$ is an unbiased estimate of $\theta$.

Sol: Given $N = 5$ and $n = 3$

a) The total no. of samples without replacement is given by $^{N}C_{n} = {^{5}C_{3}} = 10$ samples

$$\left\{ \begin{array}{lll} (3,6,9) & (3,6,15) & (3,6,27) \\ (6,9,15) & (6,9,27) & (9,15,27) \\ (3,9,15) & (3,9,27) & (6,15,27) \\ (3,15,27) & & \end{array} \right\} \longrightarrow (1)$$

b) To prove that $\bar{x}$ is an unbiased estimate of Population $\theta$ i.e $E(\bar{x}) = \theta$

The population mean $\theta = \dfrac{3+6+9+15+27}{5}$

$$= \frac{60}{5}$$

$$\theta = 12$$

Find mean of each sample of (1)

$$\left\{ \begin{array}{lll} 6 & 8 & 12 \\ 10 & 14 & 17 \\ 9 & 13 & 16 \\ 15 & & \end{array} \right\}$$

Now assign each sample a probability of $\frac{1}{10}$ i.e

| $x_i$ | 6 | 8 | 12 | 10 | 14 | 17 | 9 | 13 | 16 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|
| $P(x_i)$ | $\frac{1}{10}$ | $\frac{1}{10}$ | $\frac{1}{10}$ | $\frac{1}{10}$ | $\frac{1}{10}$ | $\frac{1}{10}$ | $\frac{1}{10}$ | $\frac{1}{10}$ | $\frac{1}{10}$ | $\frac{1}{10}$ |

To find,

$$E(\bar{x}) = \sum_{i=1}^{n} x_i P_i$$

$$= 6\left(\frac{1}{10}\right) + 8\left(\frac{1}{10}\right) + 12\left(\frac{1}{10}\right) + 14\left(\frac{1}{10}\right) + 10\left(\frac{1}{10}\right) + 17\left(\frac{1}{10}\right) + 9\left(\frac{1}{10}\right) + 13\left(\frac{1}{10}\right) + 16\left(\frac{1}{10}\right) + 15\left(\frac{1}{10}\right)$$

$$= \frac{120}{10}$$

$$= 12$$

$\therefore E(\bar{x}) = \theta$

$\therefore$ $\bar{x}$ is an unbiased estimate of population parameter $\theta$

(2) Suppose that we observe a random variable having the binomial distribution

a) Show that $\frac{x}{n}$ is an unbiased estimate of the binomial parameter p.

b) Show that $\frac{x+1}{n+2}$ is not an unbiased estimate of binomial parameter p.

Sol:

a) To show that, $E\left(\frac{x}{n}\right) = P$

Consider $E\left(\frac{x}{n}\right) = \frac{1}{n} E(x)$

$$= \frac{1}{n}[\mu]$$

$$= \frac{1}{\cancel{n}}(\cancel{n}P) \qquad [\because \text{In B.D } \mu = nP]$$

$\therefore E\left(\frac{x}{n}\right) = P$

$\therefore \frac{x}{n}$ is an unbiased estimate of p.

b) To show that, $E\left(\frac{x+1}{n+2}\right) \neq P$

Consider $E\left(\frac{x+1}{n+2}\right) = \frac{1}{n+2} E(x+1)$

$$= \frac{E(x)}{n+2} + \frac{E(1)}{n+2}$$

$$= \frac{1}{n+2}(np) + \frac{1}{n+2}(1)$$

$$E\left(\frac{x+1}{n+2}\right) = \frac{np+1}{n+2} \neq P$$

$\therefore \frac{x+1}{n+2}$ is not an unbiased estimate of p.

## Formulae:

* To find the size of sample

$$\rightarrow n = \left(\frac{Z_{\alpha/2}}{E}\right)^2 PQ \qquad [\because \text{If proportion is given}]$$

$$\rightarrow n = \left(\frac{Z_{\alpha/2}\,\sigma}{E}\right)^2 \qquad [\because \text{If } \sigma \text{ is given}]$$

The max error is $E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

→ The confidence interval for mean is

$$\left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

## Confidence Limits:

* For 95% confidence limit is 1.96
* For 99% confidence limit is 2.58
* For 90% confidence limit is 1.645
* For 98% confidence limit is 2.33

③ What is the maximum error one can expect to make with probability 0.90 when using the mean of a random sample of size $n = 64$ to estimate the mean of the population with $\sigma^2 = 2.56$

Sol: Given $n = 64, \sigma^2 = 2.56 \Rightarrow \sigma = 1.6$

$Z_{\alpha/2} = 0.90 = 90\%$

$Z_{\alpha/2} = 1.645$

Maximum error $E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

$= 1.645 \times \frac{1.6}{\sqrt{64}}$

$= \frac{2.632}{8}$

$= 0.329$

Confidence Interval :- A confidence interval displays the Prob. that a parameter will fall between a pair of values around the mean. Confidence interval is used to measure uncertainty in a sample variable. The confidence level represents the frequency of acceptable confidence intervals that contain the true value of unknown parameter

(4) Assuming that $\sigma = 20$, how large a random sample is taken to assert with probability 0.95 that the sample will not differ from the true mean by more than three points error.

sol: Given $\sigma = 20$, $E = 3$, $Z_{\alpha/2} = 95\%$

$= 1.96$

Size of the sample, $n = \left(\frac{Z_{\alpha/2}\sigma}{E}\right)^2$

$$= \left(\frac{1.96 \times 20}{3}\right)^2$$

$$= (13.06)^2$$

$$= 170.737$$

$$\cong 171$$

(5) The dean of a college wants to use the mean of a random sample to estimate the average amount of time students to get from one class to next class and he want to be able to assert with 99% confidence that the error is atmost 0.25 min. If it can be preresumed from experience that $\sigma = 1.40$ min. How large a sample will he have to take.

sol: Given $E = 0.25$ min

$\sigma = 1.40$ min

$Z_{\alpha/2} = 99\%$

$= 2.58$

Size of Sample $n = \left(\frac{Z_{\alpha/2}\,\sigma}{E}\right)^2$

$= \left(\frac{2.58 \times 1.40}{0.25}\right)^2$

$= (14.448)^2$

$= 208.74$

$\cong 209$

⑥ What is the size of the sample required to estimate an unknown proportion with a maximum error of 0.06 with atleast 95% confidence.

Sol: Given $E = 0.06$

Given, proportion i.e $P = \frac{1}{2}$

$q = \frac{1}{2}$

$Z_{\alpha/2} = 95\% = 1.96$

Size of the sample $n = \left(\frac{Z_{\alpha/2}}{E}\right)^2 pq$

$= \left(\frac{1.96}{0.06}\right)^2 \frac{1}{2} \times \frac{1}{2}$

$= \frac{1067.11}{4}$

$= 266.77$

$\cong 267$

16/11/21

⑦ The mean and S.D of a population are 11.795 and 14.054. What can one assert with 95% confidence about the maximum error if $\bar{x} = 11.795$ and $n = 50$. Construct 95% confidence interval for the true mean.

Sol: Given $\mu = 11.795$, $\sigma = 14.054$, $\bar{x} = 11.795$ $n = 50$

The confidence interval for mean is

$$\left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

$$= \left[11.795 - \frac{1.96(14.054)}{\sqrt{50}}, 11.795 + \frac{1.96(14.054)}{\sqrt{50}}\right]$$

$$= \left[11.795 - 3.895, 11.795 + 3.895\right]$$

$$= \left[7.89, 15.69\right]$$

$$= (7.89, 15.69)$$

⑧ A random sample of size 81 was taken with variance 20.25 and mean is 32, Construct 98% confidence interval.

Sol: Given $\bar{x} = 32$, $n = 81$, $\sigma^2 = 20.25 \Rightarrow \sigma = 4.5$

$$Z_{\alpha/2} = 2.33$$

Confidence interval is $\left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$

$$= \left[32 - \frac{2.33(4.5)}{\sqrt{81}}, 32 + \frac{2.33(4.5)}{\sqrt{81}}\right]$$

$= [32 - 1.165, 32 + 1.165]$

$= [30.835, 33.165]$

⑨ A random sample of 500 points on a heated plate result in an avg. temperature of 73.54 degrees with a S.D of 2.79 degrees. Find a 99% confidence interval for the average temperature of the plate.

Sol: $\bar{x} = 73.54$, $\sigma = 2.79$, $Z_{\alpha/2} = 2.58$, $n = 500$

Confidence interval is $\left(\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

$$= \left(73.54 - \frac{(2.79)(2.58)}{\sqrt{500}}, 73.54 + \frac{(2.58)(2.79)}{\sqrt{500}}\right)$$

$= (73.54 - 0.32, 73.54 + 0.32)$

$= (73.22, 73.86)$

Estimation: To use the statistic obtained by the samples as an estimate to predict the unknown parameter of the population from which the sample is drawn.

# Applications of Normal Distribution:

① Data obtained from Psychological, Physical and Biological measurements approximately follow Normal Distribution. Eg: Scores, Heights & Weights

② Most of the distributions like Binomial, Poisson can be approximated to Normal Distribution.

- If $n$ is very large and neither $p$ nor $q$ is very small, Binomial Distribution tends to Normal Distribution.
- If the parameter $\lambda \to \infty$, then Poisson Distribution tends to Normal Distribution.

③ Normal Distribution is applicable to many applied Problems in Kinetic Theory of Gases & Fluctuations in magnitude of an electric current.

④ For large samples, any statistic (sample mean, sample S.D etc) approximately follows Normal Distribution.

⑤ Normal curve is used to find confidence limits of population parameters.

⑥ Normal distribution finds large application in Statistical Quality Control in industry for finding control limits.