CHAPTER - 6

ESTIMATION

6.1 INTRODUCTION

Consider a random sample of n observations, say $x_1, x_2, \ldots, x_n$ is selected from a population whose probability distribution function is $f(x; \theta_1, \theta_2, \ldots, \theta_k)$ where $\theta_1, \theta_2, \ldots, \theta_k$ are the population parameters. Then there will always be an infinite number of functions of sample values, called statistic, which may be proposed as estimates of one or more of the parameters. The statistic whose distribution concentrates as closely as possible near the true value of the parameter may be regarded as the best estimate. In other words, the problem of estimation can be written as "determining the functions of sample observations $\hat{\theta}_1(x_1, x_2, \ldots, x_n)$, $\hat{\theta}_2(x_1, x_2, \ldots, x_n)$, ..., $\hat{\theta}_k(x_1, x_2, \ldots, x_n)$ such that their distribution is concentrated as closely as possible near the true value of the parameter". These estimating functions are also referred to as estimators.

6.2 DEFINITIONS

Quantities appearing in distributions, such as p in the binomial distribution and μ and σ in the normal distribution are called parameters.

Estimate [JNTU (H) Nov. 2009, (A) 2010S, Nov. 2011 (Set No. 1), May 2012]

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

Estimator [JNTU (H) Nov. 2009, (A) 2010S, Nov. 2011 (Set No. 1), May 2012]

The procedure or rule to determine an unknown population parameter is called an estimator. For instance, sample mean is an estimator of population mean because sample mean is a method of determining the population mean. Remember that an estimator must be a statistic and it must depend only on the sample and not on the parameter to the estimated. So an estimator is a statistic which for all practical purposes, can be used in place of unknown parameter of the population.

A parameter can have one or two or many estimators.

Types of Estimation :

Basically, there are two kinds of estimates to determine the statistic of the population parameters namely, ***(a)* Point Estimation** and ***(b)* Interval Estimation.**

Statistical Estimation

W. A. Spur and C. P. Bonini defined ***Statistical Inference*** as "the process by which we draw a conclusion about some measure of a population based on a sample value. The measure might be a variable, such as the mean, S. D., etc. The purpose of sampling is to estimate some characteristics for the population from which the sample is selected".

Basically, there are two types of problems under statistical inference :

(1) Hypothesis testing - to test some hypothesis about parent population from which the sample is drawn.

(2) Estimation - to use the statistics obtained from the sample as estimate of the unknown parameter of the population from which the sample is drawn.

An important problem of Statistical Inference is the estimation of population parameters (*i.e.*, population mean, population S.D, etc.) from the corresponding sample statistics (*i.e.*, sample mean, sample S.D. etc)

Point Estimation and Interval Estimation **[JNTU(K) Nov. 2011 (Set No. 2)]**

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

Example : If the height of a student is measured as 162 cms, then the measurement gives a point estimation. But if the height is given as (163 ± 3.5) cms, then the height lies between 159.5 cms and 166.5 cms and the measurement gives an interval estimation.

The sample mean $\bar{x}$ is a point estimate of population mean μ, sample variance s^2 is a point estimate of population variance σ^2.

Definition : A point estimate of a parameter θ is a single numerical value, which is computed from a given sample and serves as an approximation of the unknown exact value of the parameter.

Definition : A point estimator is a statistic for estimating the population parameter θ and will be denoted by $\hat{\theta}$ (read as theta hat).

Properties of Estimation : An estimator is not expected to estimate the population parameter without error. An estimator should be close to the true value of unknown parameter.

Unbiased and Biased Estimates

A statistic is said to be an *unbiased estimator* of the corresponding parameter if the mean of the sampling distribution of the statistic is equal to the corresponding population parameter. Otherwise the statistic is called a *biased estimator* of the corresponding parameter. The values of statistics in the above two cases are called *unbiased* and *biased estimates* respectively.

If t be a statistic and θ be the correspondingparameter and $E(t) = \theta$, then t is an unbiased estimator of θ. Otherwise t is a biased estimator of θ and the bias is $E(t) - \theta$.

For example, sample mean $\bar{x}$ is an unbiased estimator of population mean μ [since $E(\bar{x}) = \mu$]. **[JNTU(A) Nov. 2011 (Set No. 2)]**

Proof : Let $x_1, x_2, ..., x_n$ be a random sample drawn from a given population with mean μ and variance σ^2. Then

$$E(\bar{x}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}E(x_1 + x_2 + \ldots. + x_n)$$

$$= \frac{1}{n}[E(x_1) + E(x_2) + \ldots. + E(x_n)]$$

$$= \frac{1}{n}[\mu + \mu + \ldots. + \mu] \quad [\because E(x_i) = \mu]$$

Hence the sample mean $\bar{x}$ is an unbiased estimator of the population mean μ.

Unbiased Estimator : Let $\hat{\theta}$ be an estimator of θ. The statistic θ is said to be an unbiased estimator, or its value an unbiased estimate, if, and only if the mean or expected value of $\hat{\theta}$ is equal to θ. This is equivalent to say that the mean of the probability distribution of $\hat{\theta}$ (or the mean of the sampling distribution of $\hat{\theta}$) is equal to θ. An estimator possessing this property is said to be unbiased.

Definition. *Unbiased estimator* : A statistic or point estimator $\hat{\theta}$ is said to be an ***unbiased*** estimator of the parameter θ if $E(\hat{\theta}) = \theta$.

In other words, if E (statistic) = parameter, then statistic is said to be an unbiased estimator of the parameter.

Example 1 : Show that S^2 is an unbiased estimator of the parameter σ^2.

Solution : Let us write

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} [(x_i - \mu) - (\bar{x} - \mu)]^2$$

$$= \sum_{i=1}^{n} (x_i - \mu)^2 - 2(\bar{x} - \mu)\sum_{i=1}^{n} (x_i - \mu) + n(\bar{x} - \mu)^2$$

$$= \sum_{i=1}^{n} (x_i - \mu)^2 - n(\bar{x} - \mu)^2 \quad \ldots. (1)$$

$$\text{Now } E(S^2) = E\left[\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(x_i - \mu)^2 - nE(\bar{x} - \mu)^2\right], \text{ using (1)}$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} \sigma_{X_i}^2 - n\sigma_{\bar{X}}^2\right)$$

However, $\sigma_{X_i}^2 = \sigma^2$ for $i = 1, 2, \ldots., n$ and $\sigma_{\bar{X}}^2 = \dfrac{\sigma^2}{n}$

$$\therefore \quad E(S^2) = \frac{1}{n-1}\left(n\sigma^2 - n.\frac{\sigma^2}{n}\right) = \sigma^2$$

Although S^2 is an unbiased estimator of σ^2, S, on the other hand, is a biased estimator of σ with the bias becoming insignificant for large samples. This example illustrates why we divide by $(n-1)$ rather than n when the variance is estimated.

Example 2 : Prove that for a random sample of size n, $X_1, X_2, \ldots, X_n$ taken from an infinite population $s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is not unbiased estimator of the parameter σ^2 but $\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is unbiased. **[JNTU (H) Nov. 2010 (Set No. 2)]**

<u>**Solution :**</u> Let μ be the population mean. Then

$E(x_i) = \mu$ and $\text{var}(x_i) = E[(x_i - \mu)^2] = \sigma^2$ for $i = 1, 2, \ldots, n$

If s be the sample S. D., then $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ ... (1)

$$\text{Now} \quad s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

$$= \frac{1}{n}\sum x_i^2 - 2\bar{x}\cdot\frac{\sum x_i}{n} + \frac{1}{n}\sum \bar{x}^2$$

$$= \frac{\sum x_i^2}{n} - 2\bar{x}^2 + \bar{x}^2 = \frac{\sum x_i^2}{n} - \bar{x}^2$$

$$= \frac{\sum(x_i - \mu)^2}{n} - (\bar{x} - \mu)^2$$

$$\left[\because \frac{\sum(x_i - \mu)^2}{n} - (\bar{x} - \mu)^2 = \frac{\sum x_i^2}{n} - 2\mu\frac{\sum x_i}{n} + n.\frac{\mu^2}{n} - (\bar{x}^2 + \mu^2 - 2\bar{x}\mu) = \frac{\sum x_i^2}{n} - \bar{x}^2\right]$$

$$\therefore E(s^2) = E\left[\frac{\sum(x_i - \mu)^2}{n} - (\bar{x} - \mu)^2\right] = E\left[\frac{\sum(x_i - \mu)^2}{n}\right] - E(\bar{x} - \mu)^2$$

$$= \frac{\sum E(x_i - \mu)^2}{n} - E(\bar{x} - \mu)^2 = \frac{\sum \text{var}(x_i)}{n} - \text{var}(\bar{x})$$

$$= \frac{\sum \sigma^2}{n} - \frac{\sigma^2}{n} \quad \left(\because \bar{x} = \frac{\sum x_i}{n}\right)$$

$$= \frac{n\sigma^2}{n} - \frac{\sigma^2}{n} = \sigma^2\left(1 - \frac{1}{n}\right) = \frac{n-1}{n}\sigma^2$$

Thus $E(s^2) \neq \sigma^2$

Hence, s^2 is a biased estimator of σ^2

Let us consider $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n}{n-1}s^2$ [by (1)]

Then $E(S^2) = E\left(\frac{n}{n-1}s^2\right) = \frac{n}{n-1}E(s^2) = \frac{n}{n-1} \times \frac{n-1}{n}\sigma^2 = \sigma^2$

Hence $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is an unbiased estimator of σ^2.

Variance of a point estimator : If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators of the same population parameter θ, we would choose the estimator whose sampling distribution has the smaller variance. Hence, if $\sigma^2_{\hat{\theta}_1} < \sigma^2_{\hat{\theta}_2}$, we say that $\hat{\theta}_1$ is a more efficient estimator of θ than $\hat{\theta}_2$.

Most Efficient Estimator : If we consider all possible unbiased estimators of some parameter θ, the one with the smallest variance is called the most efficient estimator of θ.

[JNTU (A) Nov. 2011 (Set No. 2)]

In Figure, we illustrates the sampling distribution of three different estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ and $\hat{\theta}_3$ all estimating θ. It is clear that only $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased, since their distributions are centered at θ. The estimator $\hat{\theta}_1$ has a smaller variance than $\hat{\theta}_2$ and is therefore more efficient. Hence our choice for an estimator of θ, among the three considered would be $\hat{\theta}_1$.

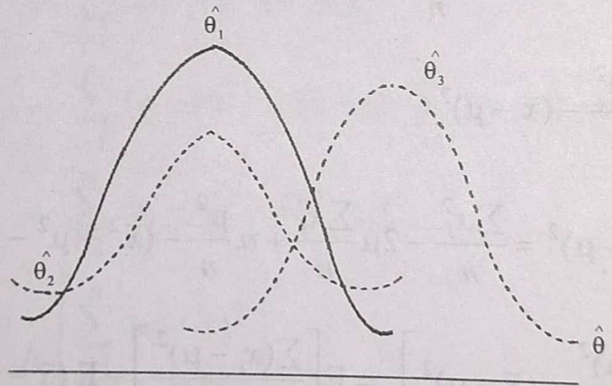

Fig. Sampling Distribution of Different Estimation of θ

Properties of Estimators : **[JNTU(K) Dec. 2013 (Set No. 3)]**

A good estimator is one which is as close to the true value of the parameter as possible. The important properties of a good estimator are :

(*i*) Consistency (*ii*) Unbiasedness (*iii*) Efficiency and (*iv*) Sufficiency

(*i*) An estimator $\hat{\theta}_n$ of a parameter θ is consistent if it converges to θ, as $n \to \infty$.

(*ii*) A statistic $\hat{\theta}$ is said to be an unbiased estimate of θ if $E(\hat{\theta}_n) = \theta$ for all θ.

(*iii*) A statistic $\hat{\theta}_1$ is said to be a more efficient unbiased estimator of the parameter θ than the statistic $\hat{\theta}_2$ if

(*a*) $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of θ

(*b*) $V(\hat{\theta}_1) < V(\hat{\theta}_2)$

(*iv*) An estimator is said to be sufficient for a parameter, if it contains all the information in the sample regarding the parameter.

6.3 INTERVAL ESTIMATION

[JNTU (A) Nov. 2011 (Set No. 3,4)]

Point estimates rarely coincide with quantities they are intended to estimate. So instead of point estimation where the quantity to be estimated is replaced by a single value a better way of estimation is interval estimation, which determines an interval in which the parameter lies.

Interval Estimate : Even the most efficient unbiased estimator cannot estimate the population parameter exactly. It is true that our accuracy increases with large samples. But there is still no reason why we should expect a **point estimate** from a given sample to be exactly equal to the population parameter, it is supposed to estimate.

Therefore in many situations it is preferable to determine an *interval* within which we would expect to find the value of the parameter. Such an interval is called an *interval estimate.* Thus an interval estimate is an interval (confidence interval) obtained from a sample.

Interval Estimation : An interval estimate of a population parameter θ 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 $\hat{\theta}$.

Since different samples will generally yield different values of $\hat{\theta}$ and, therefore, different values $\hat{\theta}_L$ and $\hat{\theta}_u$. These end points of the interval are values of corresponding random variables $\hat{\theta}_L$ and $\hat{\theta}_u$.

From the sampling distribution of $\hat{\theta}$ we shall be able to determine $\hat{\theta}_L$ and $\hat{\theta}_u$ such that the $P(\hat{\theta}_L < \theta < \hat{\theta}_u)$ is equal to any positive fractional value we care to specify. If, for instance, we find $\hat{\theta}_L$ and $\hat{\theta}_u$ such that $P(\hat{\theta}_L < \theta < \hat{\theta}_u) = 1 - \alpha$ for $0 < \alpha < 1$, then we have a probability of $1 - \alpha$ of selecting a random sample that will produce an interval containing θ. The interval $\hat{\theta}_L < \theta < \hat{\theta}_u$, computed from the selected sample, is then called a $(1 - \alpha)$ 100% **confidence interval.** The fraction $1 - \alpha$ called the confidence coefficient or the degree of confidence, and the end points, $\hat{\theta}_L$ and $\hat{\theta}_u$, are called the lower and upper confidence limits.

Thus, when $\alpha = 0.05$, we have a 95% confidence interval, and when $\alpha = 0.01$, we obtain a wider 99% confidence interval.

6.4 CONFIDENCE INTERVAL ESTIMATES OF PARAMETERS

In an interval estimation of the population parameter θ, if we can find two quantities t_1 and t_2 based on sample observations drawn from the population such that the unknown parameter θ is included in the interval $[t_1, t_2]$ in a specified percentage of cases, then this interval is called a *confidence interval* for the parameter θ. The confidence interval has a specified confidence or probability of correctly estimating the true value of the population parameter. In this case, $P(t_1 \le \theta \le t_2) = c$ where c is a given (specified) probability.

The computation of confidence interval and confidence limits is based on the sampling distribution of a statistic. By central limit theorem, we know that the sampling distribution of the sample mean $\bar{x}$ of a random sample of size $n\,(n \ge 30)$ drawn from a population having mean μ and S. D. σ is approximately a normal distribution with mean μ and

S. D = S. E. of $\bar{x} = \dfrac{\sigma}{\sqrt{n}}$

$\therefore\ z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ has a standard normal distribution with mean $= 0$ and S. D. $= 1$.

From the standard normal curve, we know that 95.45% of the area lies between the ordinates $z = \pm 2$ *i.e.*, $P(-2 < z < 2) = 95.45\% = 0.9545$, where $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$.

This shows that the inequality $-2 \le \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le 2$ ... (1)

holds in 95.45% of cases and it does not hold in only 4.55 % of the cases.

Form (1), we have $\dfrac{-2\sigma}{\sqrt{n}} \le \bar{x}-\mu \le \dfrac{2\sigma}{\sqrt{n}}$

or $\bar{x} - \dfrac{2\sigma}{\sqrt{n}} \le \mu \le \bar{x} + \dfrac{2\sigma}{\sqrt{n}}$

or $\bar{x} - 2\,(\text{S.E. of } \bar{x}) \le \mu \le \bar{x} + 2\,(\text{S.E. of } \bar{x})$ $\left[\because \text{ S.E. of } \bar{x} = \dfrac{\sigma}{\sqrt{n}}\right]$

or $[\bar{x} - 2\,(\text{S.E. of } \bar{x}), \bar{x} + 2\,(\text{S.E. of } \bar{x})]$

This interval is called the 95.45% *confidence interval* for the population mean μ and $\bar{x} \pm 2\,(\text{S.E. of } \bar{x})$ are called the 95.45% *confidence limits* for μ. Here 2 is called the confidence coefficient or *z*-value.

Again 99.73% of the area under the standard normal curve lies between the ordinates $z = \pm 3$ *i.e.*, $P(-3 \le z \le 3) = 0.9973$. Proceeding as above, 99.73% *confidence interval* and *confidence limits* for the population μ are respectively

$[\bar{x} - 3\,(\text{S.E. of } \bar{x}), \bar{x} + 3\,(\text{S.E. of } \bar{x})]$ and $\bar{x} \pm 3\,(\text{S.E. of } \bar{x})$

Also 95% of the area under the standard normal curve lies between the ordinates $z = \pm 1.96$ *i.e.*, $P(-1.96 \le z \le 1.96) = 0.95$. Proceeding as above, 95% *confidence interval* and *confidence limits* for the population mean μ are respectively

$[\bar{x} - 1.96\,(\text{S.E. of } \bar{x}), \bar{x} + 1.96\,(\text{S.E. of } \bar{x})]$ and $\bar{x} \pm 1.96$ (S. E. of $\bar{x}$)

Similarly, using the table for the area under standard normal curve, we can find the confidence interval and confidence limits corresponding to any specified (or given) percentage.

By central limit theorem, the sampling distribution of the sample proportion p of a large random sample of size n $(n \ge 30)$ drawn from a population having population proportion P is approximately a normal distribution with mean P and S. D. = S. E. of p. Using this theorem, we can find the confidence interval and confidence limits for the population proportion P, proceeding as above.

For small samples $(n < 30)$ drawn from a normal population, $t-$distribution (discussed later) is used to find the confidence interval of the population mean if σ is not known.

The formulae for confidence limits of some well-known statistic for large random samples are given below.

I. Confidence limits for Population Mean μ

(*i*) 95% confidence limits are $\bar{x} \pm 1.96$ (S.E. of $\bar{x}$)

(*ii*) 99% confidence limits are $\bar{x} \pm 2.58$ (S.E. of $\bar{x}$)

(*iii*) 99.73% confidence limits are $\bar{x} \pm 3$ (S.E. of $\bar{x}$)

(*iv*) 90% confidence limits are $\bar{x} \pm 1.64$ (S.E. of $\bar{x}$)

II. Confidence limits for Population Proportion P

(*i*) 95% confidence limits are $p \pm 1.96$ (S.E. of p)

(*ii*) 99% confidence limits are $p \pm 2.58$ (S.E. of p)

(*iii*) 99.73% confidence limits are $p \pm 3$ (S.E. of p)

(*iv*) 90% confidence limits are $p \pm 1.64$ (S. E. of p)

III. Confidence limits for the difference $\mu_1 - \mu_2$ of two Population Means μ_1 and μ_2

(*i*) 95% confidence limits are $(\bar{x}_1 - \bar{x}_2) \pm 1.96$ (S.E. of $(\bar{x}_1 - \bar{x}_2)$)

(*ii*) 99% confidence limits are $(\bar{x}_1 - \bar{x}_2) \pm 2.58$ (S.E. of $(\bar{x}_1 - \bar{x}_2)$)

(*iii*) 99.73% confidence limits are $(\bar{x}_1 - \bar{x}_2) \pm 3$ (S.E. of $(\bar{x}_1 - \bar{x}_2)$)

(*iv*) 90% confidence limits are $(\bar{x}_1 - \bar{x}_2) \pm 1.64$ (S.E. of $(\bar{x}_1 - \bar{x}_2)$)

IV. Confidence limits for the difference $P_1 - P_2$ of two population proportions

(*i*) 95% confidence limits are $(p_1 - p_2) \pm 1.96$ (S.E. of $(p_1 - p_2)$)

(*ii*) 99% confidence limits are $(p_1 - p_2) \pm 2.58$ (S.E. of $(p_1 - p_2)$)

(*iii*) 99.73% confidence limits are $(p_1 - p_2) \pm 3$ (S.E. of $(p_1 - p_2)$)

(*iv*) 90% confidence limits are $(p_1 - p_2) \pm 1.64$ (S.E. of $(p_1 - p_2)$)

Note. If for a statistic, $P(-3 < z < 3)$ *i.e.*, confidence limits cover 99.73% of the area under the standard normal curve, then we say that the confidence limits for the statistic are almost sure limits without mentioning the degree of confidence.

In calculating S.E. if population parameters are unknown, corresponding sample statistics are used to find an approximate value of S. E. For example, if the population S. D., σ is not given, the S.E. of $\bar{x}$ is calculated by using sample S.D. s in place of σ.

6.5 DETERMINATION OF PROPER SAMPLE SIZE

Determination of proper sample size is important for testing of hypothesis in business problems. The size of sample should neither be too small nor too large. If the size of the sample is too small, then it may not give a valid conclusion. On the other hand, if the size of the sample is too large, then there may be loss of time and money without getting the required results.

1. Sample size for estimating population Mean

Let $\bar{x}$ be the mean of a random sample drawn from a population having mean μ and S.D. σ. Let the sampling distribution of the sample mean $\bar{x}$ be approximately a normal distribution with mean μ and S. D. σ. If E be the permissible sampling error, then $E = \bar{x} - \mu$.

The confidence interval for the population mean μ is $\bar{x} \pm z$ (S.E. of $\bar{x}$) $= \bar{x} \pm E$ where $z =$ confidence coefficient or z – value (which is 1.96 at 5% level of significance) and $E = z$ (S.E. of $\bar{x}$) $= z.\dfrac{\sigma}{\sqrt{n}}$, n being the sample size and $\sigma =$ population S.D.

Thus $E = \dfrac{2\sigma}{\sqrt{n}}$ or $\sqrt{n} = \dfrac{z\sigma}{E}$

$\therefore\ n = \left(\dfrac{z\sigma}{E}\right)^2$, which gives the required sample size for estimating population mean.

2. Sample size for Estimating Population Proportion

Let p be the sample proportion and P be the population proportion. If E be the permissible sampling error, then $E = p - P$ and the confidence interval for the population proportion P is $p \pm z$ (S.E. of p) $= p \pm \text{E}$ where $E = z$ (S.E. of p) $= z.\sqrt{\dfrac{PQ}{n}}$, n being the sample size, or , $E^2 = \dfrac{z^2 PQ}{n}$ or $n = \dfrac{z^2 PQ}{E^2}$ where $Q = 1 - P$.

Proposition : If $x_1, x_2, ..., x_n$ be a random sample from an infinite population with variance σ^2 and sample mean $\bar{x}$, then $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is a biased estimate of σ^2, but $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is an unbiased estimate of σ^2.

3. Maximum Error of Estimate E for Large Samples : **[JNTU (A) Dec. 2009 (Set No. 2)]**

Since the sample mean estimate very rarely equals to the mean of population μ, a point estimate is generally accompanied with a statement of error which gives difference between estimate and the quantity to be estimated, the estimator.

Thus error is $|\bar{x} - \mu|$.

For large n, the random variable $\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ is normal variate approximately.

Then $P(-z_{\alpha/2} < z < z_{\alpha/2}) = 1 - \alpha$ where $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

Hence $P\left(-z_{\alpha/2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha$

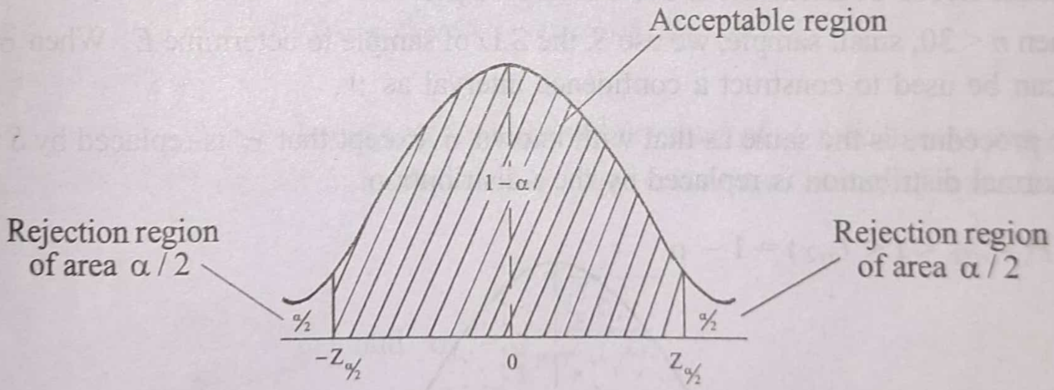

Fig. $P(-z_{\alpha/2} < z < z_{\alpha/2}) = 1 - \alpha$

Multiplying each term in the inequality by $\sigma/\sqrt{n}$, and then subtracting $\bar{x}$ from each term and multiplying by −1.

$$\boxed{\therefore\ P\left(\bar{x} - z_{\alpha/2}\cdot\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{\alpha/2}\cdot\left(\frac{\sigma}{\sqrt{n}}\right)\right) = 1 - \alpha}$$

Confidence interval for μ, σ known : **[JNTU (A) Nov. 2011 (Set No. 3)]**

If $\bar{x}$ is the mean of a random sample of size n from the population with known variance σ^2, $(1 - \alpha)$ 100% confidence interval for μ is given by

$$\bar{x} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

So, the maximum error of estimate E with $(1 - \alpha)$ probability is given by

$$E = z_{\alpha/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right)$$

Thus in the point estimation of population mean μ with sample mean $\bar{x}$ for a large random sample $(n \geq 30)$, one can assert with probability $(1 - \alpha)$ that the error $|\bar{x} - \mu|$ will not exceed $z_{\alpha/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right)$.

Sample size : When α, E, σ are known, the sample size n is given by

$$n = \left(\frac{z_{\alpha/2}\,\sigma}{E}\right)^2$$

When σ is unknown : In this case , σ is replaced by s, the standard deviation of sample to determine E.

Thus the maximum error estimate $E = t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$ with $(1-\alpha)$ probability.

4. Maximum Error of Estimate E for Small Samples :

When $n < 30$, small sample, we use S, the S.D of sample to determine E. When σ is not known, t can be used to construct a confidence interval as μ.

The procedure is the same as that with known σ except that σ is replaced by S and the standard normal distribution is replaced by the t-distribution.

$$P(-t_{\alpha/2} < t < t_{\alpha/2}) = 1 - \alpha$$

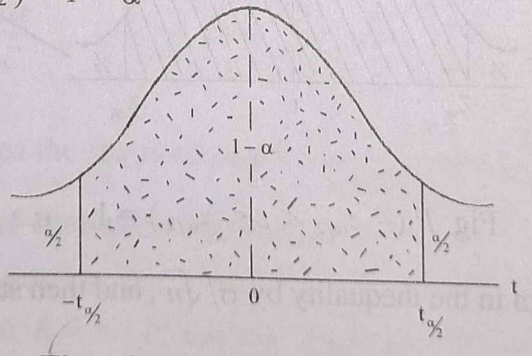

Fig. $P(-t_{\alpha/2} < t < t_{\alpha/2}) = 1 - \alpha$

Where $t_{\alpha/2}$ is the t-value within $(n - 1)$ degrees of freedom above which we find an area of $\alpha/2$. Because of symmetry, an equal area of $\alpha/2$ will fall to the left of $-t_{\alpha/2}$.

Substituting for t, we write $P\left(-t_{\alpha/2} < \dfrac{\bar{x}-\mu}{S/\sqrt{n}} < t_{\alpha/2}\right) = 1-\alpha$

Multiplying each term in the inequality by $S/\sqrt{n}$, and then subtracting $\bar{x}$ from each term and multiplying by -1, we obtain.

$$P\left(\bar{x} - t_{\alpha/2}\left(S/\sqrt{n}\right) < \mu < \bar{x} + t_{\alpha/2}\left(S/\sqrt{n}\right)\right) = 1-\alpha$$

Confidence interval for μ; σ unknown : **[JNTU (A) Nov. 2011 (Set No. 3)]**

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal population with unknown variance σ^2, a $(1-\alpha)100\%$ confidence interval for μ is

$$\bar{x} - t_{\alpha/2}\left(s/\sqrt{n}\right) < \mu < \bar{x} + t_{\alpha/2}\left(s/\sqrt{n}\right)$$

where $t_{\alpha/2}$ is the t - value with $\nu = n-1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

$\therefore$ The maximum error of estimate for small samples is given by

$$E = t_{\alpha/2} \cdot s/\sqrt{n}, \text{ where } n = \text{Sample size, } s = \text{S.D of sample}$$

6.6 BAYESIAN ESTIMATION **[JNTU (A), (H) Dec 2009, (A) Nov. 2011 (Set No.4)]**

The new concept introduced in Bayesian methods is personal or subjective probability. Also, parameters are considered as random variable in Bayesian method. To estimate the mean of a population, μ is treated as a random variable whose distribution is indicative of the "strong feelings" or assumption of a person about the possible value of μ. Let μ_0 and σ_0 be the mean and standard deviation of such a subjective "prior distribution".

Bayesian Estimation :

Combining the prior feelings about the possible values of μ with direct sample evidence, the "posterior" distribution of μ in Bayesian estimation is approximated by normal distribution with

$$\mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} \text{ and } \sigma_1 = \sqrt{\frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}}$$

where $n =$ sample size, $\bar{x} =$ sample mean and S is the standard deviation of sample.

Use $S = \sigma$.

Here μ_1 and σ_1 are known as the mean and standard deviation of the posterior distribution. In the computation of μ_1 and σ_1, σ^2 is assumed to be known, when σ^2 is unknown, which is generally the case, is replaced by sample variance S^2 provided $n \geq 30$ (large sample).

Bayesian interval for μ :

$(1-\alpha)$ 100% Bayesian interval for μ is given by

$$\mu_1 - z_{\alpha/2}.\sigma_1 < \mu < \mu_1 + z_{\alpha/2} + \sigma_1$$

SOLVED EXAMPLES

Example 1 : In a study of an automobile insurance a random sample of 80 body repair costs had a mean of Rs. 472.36 and the S.D of Rs. 62.35. If $\bar{x}$ is used as a point estimate to the true average repair costs, with what confidence we can assert that the maximum error doesn't exceed Rs. 10. **[JNTU Dec 2004S, (K) May 2010S (Set No. 2)]**

Solution : Size of a random sample, $n = 80$

The mean of random sample, $\bar{x}$ = Rs. 472.36

Standard deviation, σ = Rs. 62.35

Maximum error of estimate, E_{max} = Rs. 10

We have $\boxed{E_{max} = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}} \Rightarrow z_{\alpha/2} = \frac{E_{max} \cdot \sqrt{n}}{\sigma} = \frac{10\sqrt{80}}{62.35} = \frac{89.4427}{62.35} = 1.4345$

$\Rightarrow z_{\alpha/2} = 1.43$.

The area when $z = 1.43$ from tables is 0.4236.

$\therefore \frac{\alpha}{2} = 0.4236 \Rightarrow \alpha = 0.8472$

$\therefore$ Confidence $= (1 - \alpha)\,100\% = 84.72\,\%$

Hence we are 84.72% confidence that the maximum error is Rs. 10.

Example 2 : What is the size of the smallest sample required to estimate an unknown proportion to within a maximum error of 0.06 with atleast 95% confidence.

[JNTU Dec. 2004 S, April 2006, (K) Nov. 2009, (A) Nov. 2010 (Set No. 3)]

Solution : We are given

The maximum error, $E = 0.06$

Confidence limit = 95%

i.e. $(1 - \alpha)\,100 = 95$

$\Rightarrow \quad 1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$

$\therefore \quad z_{\alpha/2} = 1.96$

Here P is not given. So we take $P = \frac{1}{2}$. Thus $Q = \frac{1}{2}$

Hence $n = \left(\frac{z_{\alpha/2}}{E}\right)^2 (PQ)$

When P is unknown, sample size $n = \frac{1}{4}\left[\frac{z_{\alpha/2}}{E}\right]^2$

$\therefore \; n = \frac{1}{4}\left[\frac{1.96}{0.06}\right]^2 = 266.78 \approx 267$

Example 3 : If we can assert with 95% that the maximum error is 0.05 and P = 0.2, find the size of the sample. **[JNTU Dec 2005, (H) Nov. 2010 (Set No. 1)]**

Solution : Given $P = 0.2, E = 0.05$

We have $Q = 0.8$ and $z_{\alpha/2} = 1.96$ (for 95%)

We know that maximum error, $E = z_{\alpha/2}\sqrt{\left(\frac{PQ}{n}\right)}$

$$\Rightarrow 0.05 = 1.96\sqrt{\frac{0.2\times0.8}{n}}$$

$$\Rightarrow \text{Sample size}, n = \frac{0.2\times0.8\times(1.96)^2}{(0.05)^2} = 246$$

Example 4 : Assuming that $\sigma = 20.0$, how large a random sample be taken to assert with probability 0.95 that the sample mean will not differ from the true mean by more than 3.0 points ? **[JNTU Apr. 2005 , Dec 2005, (A) Nov. 2010 (Set No. 1), (K) May 2013 (Set No. 3)]**

Solution : Given maximum error $E = 3.0$ and $\sigma = 20.0$

We have $z_{\alpha/2} = 1.96$

We know that, $n = \left(\frac{z_{\alpha/2}.\sigma}{E}\right)^2$

$$\Rightarrow n = \left(\frac{1.96\times20}{3}\right)^2 = 170.74$$

$$\therefore n \simeq 171$$

Example 5 : It is desired to estimate the mean number of hours of continuous use until a certain computer will first require repairs. If it can be assumed that $\sigma = 48$ hours, how large a sample be needed so that one will be able to assert with 90% confidence that the sample mean is off by at most 10 hours.

[JNTU Dec 2004, April 2006, JNTU (K) Nov. 2009, (H) Nov.2010 (Set No. 3)]

Solution : It is given that

Maximum error, $E = 10$ hours

$\sigma = 48$ hours

and $z_{\alpha/2} = 1.645$ (for 90%)

$$\therefore n = \left(\frac{z_{\alpha/2}.\sigma}{E}\right)^2 = \left(\frac{1.645\times48}{10}\right)^2 = 62.3 \simeq 62$$

Hence sample size = 62

Example 6 : 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 population with $\sigma^2 = 2.56$. **[JNTU 2004S, (K) Nov. 2009, (H) May 2011 (Set No. 3)]**

Solution : Here $n = 64$

The probability $= 0.90$

$\sigma^2 = 2.56 \Rightarrow \sigma = \sqrt{2.56} = 1.6$

Confidence limit $= 90\%$

$\therefore\ z_{\alpha/2} = 1.645$

Hence maximum error $E = z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}} = 1.645 \times \dfrac{1.6}{\sqrt{64}} = 0.329.$

Example 7 : A random sample of size 100 has a standard deviation of 5. What can you say about the maximum error with 95% confidence. **[JNTU May 2005, (A) Apr. 2012 (Set No. 4)]**

Solution : Given $s = 5$, $n = 100$

$z_{\alpha/2}$ for 95% confidence $= 1.96$

We know that

$$\text{Maximum error, } E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\therefore\ E = (1.96) \cdot \frac{5}{\sqrt{100}} = 0.98$$

Example 8 : The efficiency expert of a computer company tested 40 engineers to estimate the average time it takes to assemble a certain computer component, getting a mean of 12.73 minutes and S.D of 2.06 minutes.

(*a*) If $\overline{x} = 12.73$ is used as a point estimate of the actual average time required to perform the task, determine the maximum error with 99% confidence.

(*b*) Construct 98% confidence intervals for the true average time it takes to do the job.

(*c*) With what confidence can we assert that the sample mean does not differ from the true mean by more than 30 seconds. **[JNTU (K) Nov. 2009 (Set No. 1)]**

(OR)

To estimate the average time it takes to assemble a certain computer component, the industrial engineer at an electronic firm timed 40 technicians in the performance of the task, getting a mean of 12.73 min. and a S. D. of 2.06 min.

(*i*) What can we say with 99% confidence about the maximum error if $\overline{x} = 12.73$ is used as a point estimate of the actual average time required to do the job ?

(*ii*) Use the given data to construct 98% confidence interval.

(*iii*) With what confidence we can assert that the sample mean does not differ from the true mean by more than 30 sec. **[JNTU (A) Apr. 2012 (Set No. 1)]**

Solution : Here $\overline{x} = 12.73$, $S = 2.06$, $n = 40$. For 99%, $z_{\alpha/2} = 2.575$

(*a*) Maximum error of estimate $E = z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$

$$\therefore\ E = (2.575) \cdot \frac{(2.06)}{\sqrt{40}} = 0.8387$$

(*b*) For 98% confidence, $E = (2.33) \cdot \dfrac{(2.06)}{\sqrt{40}} = 0.758915$

98% confidence interval limits are

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = \bar{x} \pm E = 12.73 \pm 0.7589$$

i.e., confidence interval is (11.97, 13.4889)

(c) $\frac{30}{60}$ minutes $= \frac{1}{2}$ minutes $\Rightarrow E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = z_{\alpha/2} \cdot \frac{2.06}{\sqrt{40}}$

$\therefore \quad z_{\alpha/2} = 1.5350.$

From normal distribution table, the area corresponding to $z_{\alpha/2} = 1.5350$ is 0.4370.

Then the area between $z_{-\alpha/2}$ to $z_{\alpha/2}$ is 2 (0.4370) = 0.8740

Thus we have 87.4% confidence.

Example 9 : The mean and standard deviation of a population are 11,795 and 14,054 respectively. What can one assert with 95% confidence about the maximum error if $\bar{x} = 11,795$ and $n = 50$. And also construct 95% confidence interval for the true mean.

[JNTU (A) Apr. 2012(Set No. 4)]

(OR)

The mean and the standard deviation of a population are 11,795 and 14054 respectively. If $n = 50$, find 95% confidence interval for the mean.

[JNTU 2006, 2008S, 2008 (Set No. 3), (K) May 2013 (Set No. 2)]

Solution : Here Mean of population, $\mu = 11795$

S.D of population, $\sigma = 14054$

$\bar{x} = 11795$

n = sample size = 50, maximum error $= z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$z_{\alpha/2}$ for 95% confidence = 1.96

Max. error, $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{(14054)}{\sqrt{50}} = 3899$

$\therefore$ Confidence interval $= \left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$

$= (11795 - 3899, 11795 + 3899)$

$= (7896, 15694)$

Example 10 : A sample of 10 cam shafts intended for use in gasoline engines has an average eccentricity of 1.02 and a standard deviation of 0.C14 inch. Assuming the data may be treated a random sample from a normal population, determine a 95% confidence interval for the actual mean eccentricity of the cam shaft ? **[JNTU 2004, (K) May 2010S (Set No. 4)]**

Solution : We know that confidence interval is $\bar{x} \pm E$ where $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

Given $n = 10$, $z_{\alpha/2} = 1.96$, $\sigma = 0.044$, $\bar{x} = 1.02$

Confidence interval is $1.02 \pm (1.96)\left(\frac{0.044}{\sqrt{10}}\right) = 1.02 \pm 0.027$

$\therefore$ Confidence interval = (0.993, 1.047)

Example 11 : A random sample of size 81 was taken whose variance is 20.25 and mean is 32, construct 98% confidence interval. **[JNTU Dec 2005 (Set No. 2)]**

Solution : Given $\bar{x}$ = sample mean = 32, n = 81,

$$\sigma^2 = 20.25 \Rightarrow \sigma = 4.5$$

and $z_{\alpha/2} = 2.33$ (for 98%)

We know that 98% confidence interval is $\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

Now $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = (2.33)\frac{4.5}{\sqrt{81}} = 1.165$

$\therefore$ Confidence interval = (32 – 1.29, 32 + 1.29)

= (30.835, 33.165)

Example 12 : A research worker wants to determine the average time it takes a mechanic to rotate the tyres of a car and he wants to be able to assert with 95%. Confidence that the mean of his sample is off by atmost 0.5 minutes. If he can presume from past experience that σ = 1.6 minutes, how large a sample will have to take? **[JNTU 1999]**

Solution : Max. error, $E = 0.5$

n = sample size = ?

σ = standard deviation of the population = 1.6 minutes

$z_{\alpha/2} = 1.96$

$$\text{Hence } n = \left[\frac{z_{\alpha/2}.\sigma}{E}\right]^2 = \left[\frac{1.96 \times 1.6}{0.5}\right]^2 \approx 40$$

$\therefore$ 40 mechanics will have to perform the task.

Example 13 : It is desired to estimate the mean time of continuous use until an answering machine will first require service. If it can be assumed that σ = 60 days, how large a sample is needed so that one will be able to assert with 90% confidence that the sample mean is off by at most 10 days.

Solution : We have

Maximum error = 10 days = E, σ = 60 days and $z_{\alpha/2} = 1.645$

$$\therefore n = \left[\frac{z_{\alpha/2}.\sigma}{E}\right]^2 = \left[\frac{1.645 \times 60}{10}\right]^2 = 97$$

Example 14 : The dean of a college wants to use the mean of a random sample to estimate the average amount of time students take to get from one class to the next and she wants to be able to assert with 99% confidence that the error is at most 0.25 minute. If it can be presumed from experience that σ = 1.40 minutes. How large a sample will she have to take ? **[JNTU 2000]**

Solution : We are given

Maximum error, $E = 0.25$ minutes

σ = standard deviation = 1.40 minutes

$z_{\alpha/2} = 2.575$,

$$\therefore \text{ Sample size, } n = \left[\frac{z_{\alpha/2}.\sigma}{E}\right]^2 = \left[\frac{2.575 \times 1.4}{0.25}\right]^2 = 208$$

Example 15 : A random sample of size 100 is taken from a population with $\sigma = 5.1$. Given that the sample mean is $\bar{x} = 21.6$. Construct a 95% confidence interval for the population mean μ. **[JNTU 2001]**

Solution : Given $\bar{x}$ = sample mean = 21.6,

$z_{\alpha/2} = 1.96$, n = sample size = 100, $\sigma = 5.1$

$\therefore$ Confidence interval = $\left(\bar{x} - z_{\alpha/2}.\sigma/\sqrt{n},\ \bar{x} + z_{\alpha/2}.\sigma/\sqrt{n}\right)$

Now $\bar{x} - z_{\alpha/2}.\dfrac{\sigma}{\sqrt{n}} = 21.6 - \dfrac{1.96 \times 5.1}{10} = 20.6$

and $\bar{x} + z_{\alpha/2}.\dfrac{\sigma}{\sqrt{n}} = 22.6$

Hence (20.6, 22.6) is the confidence interval for the population mean μ.

Example 16 : The mean of random sample is an unbiased estimate of the mean of the population 3, 6, 9, 15, 27.

(*i*) List of all possible samples of size 3 that can be taken without replacement from the finite population.

(*ii*) Calculate the mean of each of the samples listed in (*a*) and assigning each sample a probability of 1/10. Verify that the mean of these $\bar{x}$ is equal to 12. Which is equal to the mean of the population θ *i.e* E $(\bar{x}) = \theta$ *i.e.*, prove that $\bar{x}$ is an unbiased estimate of θ. **[JNTU 2004S, 2008S, (K) Nov. 2009, (A) Nov. 2010 (Set No. 2)]**

Solution : (*i*) The possible samples of size 3 taken from 3, 6, 9, 15, 27 without replacement, are ${}^5C_3 = 10$ samples *i.e.*,(3,6,9), (3,6,15), (3,6,27), (6,9,15), (6, 9, 27), (3, 9, 15), (3, 9, 27), (9, 15, 27), (6, 15, 27), (3, 15, 27).

(*ii*) Mean of the population $\theta = \dfrac{3+6+9+15+27}{5} = 12$

Means of the samples are 6, 8, 12, 10, 14, 9, 13, 17, 16, 15.

Probability assigned to each one is $\dfrac{1}{10}$ each.

$\bar{x}$	6	8	12	10	14	9	13	17	16	15
$P(\bar{x})$	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10

$$E(\bar{x}) = 6.\frac{1}{10} + 8.\frac{1}{10} + 12.\frac{1}{10} + 10.\frac{1}{10} + \frac{14}{10} + 9.\left(\frac{1}{10}\right) + 13.\frac{1}{10}$$

$$+ 17.\frac{1}{10} + 16.\frac{1}{10} + 15.\frac{1}{10} = \frac{1}{10} \times 120 = 12 = \theta$$

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

$\therefore$ $\bar{x}$ is an unbiased estimate of θ.

i.e., the mean of a random sample is an unbiased estimator of the mean of the population.

Example 17 : Suppose that we observe a random variable having the binomial distribution and get x successes in n trials.

(*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 the binomial parameter p.

<u>**Solution :**</u> $\hat{\theta}$ is an unbiased estimator of θ if $E(\bar{\theta}) = \theta$

(*a*) $$E\left(\frac{x}{n}\right) = \frac{1}{n}E(x) = \frac{np}{n} = p \quad [\because E(x) = np]$$

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

(*b*) $$E\left(\frac{x+1}{n+2}\right) = E\left(\frac{x}{n+2} + \frac{1}{n+2}\right) = \frac{1}{n+2}E(x) + \frac{1}{n+2} \quad [\because E(ax+b) = aE(x) + b]$$

$$= \frac{np}{n+2} + \frac{1}{n+2} \quad [\because E(x) = p]$$

$$= \frac{np+1}{n+2} \neq p$$

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

Example 18 : Find 95% confidence limits for the mean of a normality distributed population from which the following sample was taken 15, 17, 10, 18, 16, 9, 7, 11, 13, 14.

[JNTU (H) Nov. 2009 (Set No. 4), (A) Nov. 2010 (Set No. 4), (K) May 2010S, May 2013 (Set No. 4)]

<u>**Solution :**</u> We have $\bar{x} = \frac{15+17+10+18+16+9+7+11+13+14}{10} = 13$

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$$

$$= \frac{1}{9}[(15-13)^2 + (17-13)^2 + (10-13)^2 + (18-13)^2 + (16-13)^2$$

$$+ (9-13)^2 + (7-13)^2 + (11-13)^2 + (13-13)^2 + (14-13)^2] = \frac{40}{3}$$

Since $z_{\alpha/2} = 1.96$, we have

$$z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 1.96 \,.\, \frac{\sqrt{40}}{\sqrt{10}.\sqrt{3}} = 2.26$$

$\therefore$ Confidence limits are $\overline{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 13 \pm 2.26 = (10.74, 15.26)$

Example 19 : Determinne a 95% confidence interval for the mean of a normal distribution with variance 0.25, using a sample of $n = 100$ values with mean 212.3.

[JNTU(K) March. 2014 (Set No. 1)]

Solution : We have $n = 100$, $\overline{x} = 212.3$, S.D. $= \sigma = \sqrt{0.25}$ and $z_{\alpha/2} = 1.96$ (for 95%)

We know that 95% confidence interval is

$$\left(\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Now $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = \frac{1.96 \times \sqrt{0.25}}{\sqrt{100}} = \frac{1.96 \times 0.5}{10} = \frac{0.98}{10} = 0.098$

$\therefore$ Confidence interval $= (212.3 - 0.098, 212.3 + 0.098)$

$= (212.202, 212.398)$.

Example 20 : A professor's feelings about the mean mark in the final examination in "Probability" of a large group of students is expressed subjectively by normal distribution with $\mu_0 = 67.2$ and $\sigma_0 = 1.5$.

(*a*) If the mean mark lies in the interval (65.0, 70.0) determine the prior probability the professor should assign to the mean mark.

(*b*) Find the professor mean μ_1 and the posterior S.d σ_1 if the examinations are conducted on a random sample of 40 students yielding mean 74.9 and S.D. 7.4. Use S = 7.4 as an estimate σ.

(*c*) Determine the posterior probability which he will thus assign to the mean mark being in the interval (65.0, 70.0) using results obtained in (b).

(*d*) Construct a 95% Bayesian interval for μ. **[JNTU(K) Nov. 2011 (Set No. 4)]**

Solution : (*a*) Here $\mu_0 = 67.2$, $\sigma_0 = 1.5$ and $n = 40$

Standard variable corresponding to 65.0 is

$$z_1 = \frac{65.0 - 67.2}{1.5} = -1.466$$

$$\text{Similarly, } z_2 = \frac{70 - 67.2}{1.5} = 1.866$$

Let X be the mean mark obtained in the final examination

Prior probability $= P(65 < x < 70) = P(-1.47 < z < 1.87)$

$= 0.4292 + 0.4693 = 0.8985$

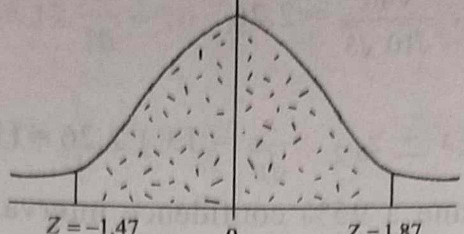

Fig.

(b) Here $\overline{x} = 74.9$, $\sigma = S = 7.4$

Posterior mean, $\mu_1 = \dfrac{n\bar{x}\,\sigma_0^2 + \mu_0\,\sigma^2}{n\sigma_0^2 + \sigma^2}$

i.e., $\mu_1 = \dfrac{40\,(74.9)\,(1.5)^2 + (67.2)\,(7.4)^2}{40\,(1.5)^2 + (7.4)^2} = 71.987 \approx 72$

Posterior S.D, $\sigma_1 = \sqrt{\dfrac{\sigma^2\,\sigma_0^2}{n\,\sigma_0^2 + \sigma^2}} = \sqrt{\dfrac{(7.4)^2\,(1.5)^2}{40\,(1.5)^2}} = 0.922568 \approx 0.923$

(c) Here $\mu_1 = 72$, $\sigma_1 = 0.923$

Standard variable corresponding to 65 is

$$Z_1 = \frac{65 - 72}{0.923} = -7.5839$$

Similarly, corresponding to 70.0 is

$$Z_2 = \frac{70 - 72}{0.923} = -2.16684$$

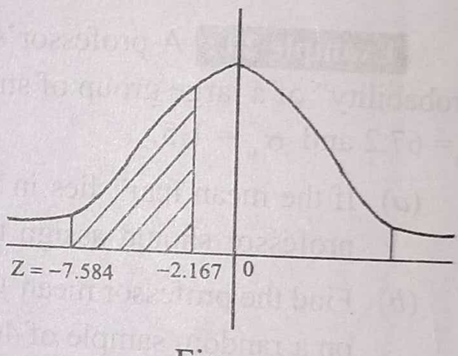

Fig.

Posterior probability $= P(65 < x < 70)$

$= P(-7.584 < Z < -2.167)$

$= 0.5 - 0.4850 = 0.0150$

(d) 95% Bayesian Interval limits are

$$\mu_1 \pm Z_{\alpha/2} \cdot \sigma_1 = 71.987 \pm (1.96)\,(0.922568)$$

Thus the Bayesian interval is $(70.17876, 72.909568)$

Example 21 : A random sample of 100 teachers in a large metropolitan area revealed a mean weekly salary of Rs. 487 with a standard deviation Rs 48. With what degree of confidence can we assert that the average weekly salary of all teachers in the metropolitan area is between 472 to 502 ? **[JNTU 2006S (Set No.2), 2007, (A) Nov. 2010, Apr. 2012, (K) Nov. 2011 (Set No. 3)]**

Solution : Given $\mu = 487$, $\sigma = 48$, $n = 100$

$$z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\overline{x} - 487}{\frac{48}{\sqrt{100}}} = \frac{\overline{x} - 487}{4.8}$$

Standard variable corresponding to Rs. 472 is

$$z_1 = \frac{472-487}{4.8} = -3.125$$

Standard variable corresponding to Rs. 502 is

$$z_2 = \frac{502-487}{4.8} = 3.125$$

Let X be the mean salary of teacher. Then

$$P(472 < x < 502) = P(-3.125 < z < 3.125)$$

$$= 2\,(0 < z < 3.125)$$

$$= 2\int_0^{3.125} \phi(z)\,dz$$

$$= 2\,(0.4991) = 0.9982$$

Thus we can ascertain with 99.82% confidence.

Example 22 : A population random variable has mean 100 and standard deviation 16 what are the mean and standard deviation of the sample mean for the random sample of size 4 drawn with replacement ? **[JNTU April 2004, (A) Apr. 2012 (Set No. 2)]**

Solution : Given $\mu = 100$, $\sigma = 16$, $n = 4$

Since the sampling is done with replacements, the population may be considered as infinite.

We have to find $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$.

$$\therefore \mu_{\bar{x}} = \mu = 100 \text{ and } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{16}{\sqrt{4}} = 8$$

Example 23 : A random sample of 400 items is found to have mean 82 and S.D. of 18. Find the maximum error of estimation at 95% confidence interval. Find the confidence limits for the mean if $\bar{x} = 82$. **[JNTU Nov. 2008, (A) Nov. 2011, (H) May 2011 (Set No. 2)]**

Solution : Given standard deviation $= \sigma = 18$

Sample size $= n = 400$

$Z_{\alpha/2}$ for 95% confidence $= 1.96$ (from tables)

Sample mean $= \bar{x} = 82$

$$\text{Maximum error, } E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = \frac{1.96 \times 18}{20} = 1.764$$

The limits for the confidence are

$$\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$\therefore$ Confidence limits are 80.236 and 83.764

Example 24 : Measurements of the weights of a random sample of 200 ball bearing made by a certain machine during one week showed a mean of 0.824 and a standard deviation of 0.042. Find maximum error at 95% confidence interval ? Find the confidence limits for the mean if $x = 32$. **[JNTU Nov. 2008, (A) Nov. 2011 (Set No. 4)]**

Solution : We are given

Mean of the sample = $\overline{x}$ = 0.824

$Z_{\alpha/2}$ = Z value for 95% level = 1.96 (From Tables)

σ = Standard deviation = 0.042

n = size of the sample = 200

$$\text{Maximum error, } E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = \frac{1.96 \times 0.042}{\sqrt{200}} = 0.0059$$

$$\text{Now } \overline{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 0.824 - 0.0059 = 0.8181$$

$$\text{and } \overline{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 0.824 + 0.0059 = 0.8299$$

Hence the limits for the confidence are

$$\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$\therefore$ Confidence limits are 0.8181 and 0.8299

Example 25 : A sample of size 300 was taken whose variance is 225 and mean 54. Construct 95% confidence interval for the mean. **[JNTU (H) Nov. 2010 (Set No. 1)]**

Solution : Since the sample size 300 is large (> 30), normal distribution is used as the sampling distrubution.

Here $n = 300$, $\overline{x}$ = sample mean = 54, $\sigma = \sqrt{225} = 15$

$$\therefore \text{ S.E. of } \overline{x} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{300}} = 0.866$$

95% confidence limits for the population mean are

$$\overline{x} \pm 1.96 \text{ (S.E. of } \overline{x}) = 54 \pm 1.96\,(0.866) = 54 \pm 1.697 = 55.697 \text{ and } 52.3$$

$\therefore$ The required confidence interval is (52.3, 55.7).

Example 26 : In a random sample of 100 packages shipped by air freight 13 had some damage. Construct 95% confidence interval for the true proportion of damage package.

Solution : Here p = Sample proportion of damage packages = $\frac{13}{100} = 0.13$

$\therefore q = 1 - p = 1 - 0.13 = 0.87$

S.E. of $p = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{pq}{n}}$ ($\because$ P is not known, we take p for P)

$$= \sqrt{\frac{0.13 \times 0.87}{100}} = 0.034$$

$\therefore$ 95% confidence limits for the population proportion of P of damage packages are

$p \pm 1.96$ (S.E. of p) $= 0.13 \pm 1.96\,(0.034) = 0.13 \pm 0.067$

i.e., the confidence limits are 0.063 and 0.197

Hence the 95% confidence interval for the true proportion of damage packages is $(0.063, 0.197)$.

Example 27 : It is desired to estimate the mean number of hours of continuous use until a centain computer will first require repairs. If it can be assumed that $\sigma = 48$ hours. How large the sample will be needed so that one will be able to assert with 90% confidence that the sample mean is off by at most 10 hours. **[JNTU(H) Nov. 2010 (Set No. 3)]**

Solution : Sampling error, $E = 10$ hours, $\sigma = 48$ hours, $z_{\alpha/2} = 1.645$

We have $E = \frac{z\sigma}{\sqrt{n}}$ or $\sqrt{n} = \frac{z\sigma}{E}$

$$\therefore n = \left(\frac{z_{\alpha/2}\,\sigma}{E}\right)^2 = \left(\frac{1.645 \times 48}{10}\right)^2 = (7.896)^2 = 62.35$$

Hence, sample size, $n = 62$

Example 28 : Among 100 fish caught in a large lake, 18 were inedible due to the pollution of the environment. With what confidence can we assert that the error of this estimate is at most .065? **[JNTU(H) Dec. 2011 (Set No. 3)]**

Solution : We are given

$n =$ Sample size $= 100$

and Max. error of estimate, $E = 0.065$

Here $p =$ sample proportion of inedible fish $= \frac{18}{100} = 0.18$

$\therefore \quad q = 1 - p = 1 - 0.18 = 0.82$

Maximum error of estimate for true proportion $= E$

$$= z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}} \quad [\because \text{P is not known, we take } p \text{ for P}]$$

$$\Rightarrow 0.065 = z_{\alpha/2} \cdot \sqrt{\frac{0.18 \times 0.82}{100}} = z_{\alpha/2}(0.038)$$

$$\therefore z_{\alpha/2} = \frac{0.065}{0.038} = 1.71$$

When $z = 1.71$, the probability $= 0.4564$ (from normal tables)

$P(z > z_{\alpha}) = 0.4564 \Rightarrow 2 \times P(z > z_{\alpha}) = 2(0.4564) = 0.9128 = 91.28\% = 91\%$

Hence with 91% confidence we can assert that the error of this estimate is at most 0.065.

Example 29 : A random sample of 500 points on a heated plate resulted in an average temperature of 73.54 degrees Fahrenheit with a standard deviation of 2.79 degree Fahrenheit. Find a 99% confidence interval for the average temperature of the plate.

[JNTU(H) Dec. 2011 (Set No. 4)]

Solution : We are given

$n = 500, \bar{x} = 73.54$ and $\sigma = 2.79$

We have $z_{\alpha/2} = 2.58$ (for 99%)

We know that 99% confidence interval is $\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$

Now $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 2.58 \times \frac{2.79}{\sqrt{500}} = \frac{7.1982}{\sqrt{500}} = 0.32$

$\therefore$ 99% confidence interval $= (73.54 - 0.32, 73.54 + 0.32)$

$= (73.22, 73.86)$

Example 30 : In a study of an automobile insurance a random sample of 80 body repair costs had a mean of Rs.472.36 and a standard deviation of Rs.62.35. If x is used as point estimate to the true average repair costs, with what confidence we can assert that the maximum error does not exeed Rs.10?

[JNTU(H) Apr. 2012 (Set No. 4)]

Solution : We are given

$n =$ sample size $= 80$, $\sigma =$ S. D. $= 62.35$

Maximum error, $E = 10$

We have $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow 10 = z_{\alpha/2} \cdot \frac{62.35}{\sqrt{80}}$

$$\therefore z_{\alpha/2} = \frac{10\sqrt{80}}{62.35} = 1.43$$

From normal table, the area corresponding to $z_{\alpha/2} = 1.43$ is 0.4236.

Then the area between $z_{-\alpha/2}$ to $z_{\alpha/2}$ is $2(0.4236) = 0.8472$

Thus we can ascertain with 84.7% confidence.

Example 31 : The mean mark in mathematics in common entrance that will vary from year to year. If this variation of the mean mark is expressed subjectively by a normal distribution with mean $\mu_0 = 72$ and variance $\sigma_0^2 = 5.76$.

(*i*) What probability can we assign to the actual mean mark being somewhere between 71.8 and 73.4 for the next years test?

(*ii*) Construct a 95% Bayesian interval for μ if the test is conducted for a random sample of 100 students from the next incoming class yielding a mean mark of 70 with s.d. of 8.

(*iii*) What posterior probability should we assign to the event of part (*i*)

[JNTU(K) May 2010 (Set No. 1)]

Solution : (*a*) Here μ_0 = mean $= 72$ and $\sigma_0 = \text{S.D.} = \sqrt{5.76} = 2.4, n = 100$

We have $z = \dfrac{x - 72}{2.4}$

When $x = 71.8$, $z_1 = \dfrac{71.8 - 72}{2.4} = -0.0833$

When $x = 73.4$, $z_2 = \dfrac{73.4 - 72}{2.4} = 0.5833$

Let X be the mean mark obtained in the common entrance test. Then

$$\text{Prior probability} = P(71.8 < X < 73.4)$$
$$= P(-0.083 < z < 0.583)$$
$$= 0.0319 + 0.2190 = 0.2509$$

(*b*) Here $n = 100, \bar{x} = 70, \sigma_0 = 2.4, \mu_0 = 72, \sigma = 8$

$$\therefore \text{ Posterior mean, } \mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$$

$$= \frac{100\,(70)\,(2.4)^2 + 72\,(8)^2}{100\,(2.4)^2 + (8)^2} = \frac{44928}{640} = 70.2$$

$$\text{Posterior standard deviation } \sigma_1 = \sqrt{\frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2}}$$

$$= \sqrt{\frac{(8)^2(2.4)^2}{100(2.4)^2 + (8)^2}} = \sqrt{\frac{368.64}{640}}$$

$$= \sqrt{0.576} = 0.7589 \approx 0.76$$

$\therefore$ 95% Bayesian interval for μ is given by

$$\mu_1 \pm z_{\alpha/2}(\sigma_1) = 70.2 \pm 1.96\,(0.7589) = 70.2 \pm 1.4874$$

i.e. $\quad 68.71 < \mu < 71.69$

(c) Here $\mu_1 = 70.2, \sigma_1 = 0.76$

When $x = 71.8, z_1 = \dfrac{71.8 - 70.2}{0.76} = 2.105$

When $x = 73.4, z_2 = \dfrac{73.4 - 70.2}{0.76} = 4.2105 \approx 4.21$

$\therefore$ Posterior probability $= P(71.8 < X < 73.4) = P(2.105 < z < 4.21)$

$= 0.5 - 0.4821 = 0.0179$

REVIEW QUESTIONS

1. Defiine estimate, estimator and estimation. **[JNTU (A) Nov. 2011 (Set No. 1), May 2012]**
2. Define unbiased estimator. What is the more efficiet unbiased estimator, explain briefly? **[JNTU (A) Nov. 2011 (Set No. 2)]**
3. In how many ways the estimation can be done and what are they ? Explain in detail **[JNTU (A) Nov. 2011 (Set No. 1), May 2012]**
4. Explain the ways in which an estimation can be done. Also explain the Bayesian estimation **[JNTU (A) Nov. 2010 (Set No. 1)]**
5. Give the difference between the interval estimation and Bayesian estimation **[JNTU (A) Nov. 2011 (Set No. 2)]**
6. What is the "interval estimation" ? Give the relations used to find the confidence interval of large and small samples. **[JNTU (A) Nov. 2011 (Set No. 3)]**
7. Write a short note on interval estimation and Bayesian estimation. **[JNTU (A) Nov. 2011 (Set No. 4)]**
8. Show that $\overline{x}$ is an unbiased estimator of the population mean μ **[JNTU (A) Nov. 2011(Set No. 2)]**
9. Explain briefly Point and Interval estimations. **[JNTU (K) Nov. 2011 (Set No. 2)]**
10. Define the problem of estimation. Show that if $\hat{\theta}$ is an unbiased estimator of θ then show that $\hat{\theta}^2$ is a biased estimator of θ^2. **[JNTU (K) Dec. 2013 (Set No. 2)]**
11. List the properties of estimators with the examples. **[JNTU (K) Dec. 2013 (Set No. 3)]**

EXERCISE

1. To estimate the average amount of time visitors take to move from one building to another in an office complex, the mean of a random sample of size n is used. Given $\sigma = 1.40$ minutes, determine how large should be the sample size if it is ascertained with 99% confidence that the error E is atmost 0.25.
2. Find the degree of confidence to assert that the average salary of school teachers is between Rs. 272 and Rs. 302 if a random sample of 100 such teachers revealed a mean salary of Rs. 287 with S.D of Rs. 48.
3. Using the mean of a random sample of size 150 to estimate the mean mechanical aptitude of mechanics of a large workshop and assuming $\sigma = 6.2$, what can we assert with 0.99 probability about the maximum size of the error.