Unit - 6

Estimation

After studying this unit, the students will be able to

- Define estimation of a parameter, point estimation of a parameter, point estimator and point estimate.
- Differentiate between point estimator and point estimate.
- Describe from a random sample, the point estimators and point estimate for population mean and variance.
- Define un-biasedness, un-biased estimator, biased estimator and biased.
- Describe the methods to reduce bias in sample surveys.
- Describe and verify the un-biasedness of sample mean, sample proportion and sample variance.
- Use calculator in statistical mode to directly find the un-biased estimates of mean and variance of the population from which the sample was drawn.
- Define efficiency and explain best estimator.
- Identify the best estimator of population mean, population variance and population proportion.
- Find the best estimates of population mean, population variance and population proportion from a given random sample.
- Identify the pooled estimators, from two samples, of population mean, population variance and population proportion.
- Find the pooled estimates of population mean, population variance and population proportion from two given random samples.
- Define interval estimation of a parameter, interval estimate and confidence coefficient.
- Explain and estimate the confidence interval for the mean of a normal population (known and unknown standard deviations), the difference between means of two normal populations (known and unknown standard deviations), the population proportion (large sample) and the difference between proportions of two populations (large samples).

NOT FOR SALE

6.1 Introduction to statistical inference

Statistical inference is a process of drawing conclusions (inferences) about the population on the basis of sample information obtained from that population. There are two branches of statistical inference.

i. Estimation of parameters
ii. Hypothesis testing

Statistical inference is of immense importance because complete knowledge regarding population is seldom available. In the previous unit concepts of sampling and sampling distributions were discussed which is actually a base for statistical inference i.e. sampling allow us to make use of the information gathered for the sample to draw inferences about the entire population.

6.1.1 Estimation of parameters

Statistical estimation is a process by which the unknown value of a parameter is obtained from the sample observations. Suppose we want to know the average age of people in our country, the percentage of smokers in the Khyber Pakhtunkhwa etc. then these are the problems which come under estimation.

6.1.2 Types of estimation

There are two types of estimation (i) point estimation (ii) interval estimation. When a specific value is obtained from sample observations and is used to estimate the unknown value of the parameter, the process is called point estimation and the single estimated value is called point estimate or simply estimate. For example, the values obtained by $\bar{X}$, S^2, $\hat{p}$ are point estimates for the parameters μ, σ^2, p respectively. When a range of values is obtained from sample observations within which the unknown value of parameter is believed to lie, the process is called interval estimation and the resulting interval of two numbers is called interval estimate.

NOT FOR SALE

6.1.3 Difference between an estimator and an estimate

A rule, usually expressed as a formula that tells us how to calculate an estimate from the sample data is called an estimator and the resulting number is called an estimate.

For example; if $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = 100$ then $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ is an estimator and 100 is an estimate. Similarly, if $s^2 = \frac{\sum(x-\bar{x})^2}{n-1} = 2.63$ then $s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$ is an estimator and its numerical value 2.63 is an estimate.

6.2 Point estimation

This process provides a single value which is calculated from the sample data as an estimate for the unknown population parameter. Point estimate may or may not be close to the parameter because the random sample used is one of the many possible samples which could be selected from the population.

Example 6.1

A random sample has ten observations 4, 3, 7, 8, 4, 10, 5, 5, 4 and 9. Compute a point estimate of (i) population mean (ii) population standard deviation (iii) standard error of the mean (iv) population proportion of even numbers.

Solution:

(i) Point estimate of the population mean μ is;

$$\bar{X} = \frac{\sum x}{n} \qquad \text{(estimator)}$$

$$= \frac{59}{10} = 5.9 \qquad \text{(point estimate)}$$

NOT FOR SALE

(ii) Point estimate of the population standard deviation σ is;

$$s = \sqrt{\frac{1}{n-1}\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]} \qquad \text{(estimator)}$$

$$= \sqrt{\frac{1}{10-1}[401 - \frac{(59)^2}{10}]}$$

$$= \sqrt{\frac{1}{9}[401 - 348.1]}$$

$$= \sqrt{\frac{52.9}{9}} = 2.42 \qquad \text{(point estimate)}$$

(iii) Point estimate of the standard error $\sigma_{\bar{x}}$ is ;

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \qquad \text{(estimator)}$$

$$= \frac{2.42}{\sqrt{10}} = 0.77 \qquad \text{(point estimate)}$$

(iv) Point estimate of the population proportion p is;

$$\hat{p} = \frac{X}{n} \qquad \text{(estimator)}$$

$$= \frac{5}{10} = 0.5 \qquad \text{(point estimate)}$$

6.2.2 Properties of a good point estimator

In point estimation the unknown value of a parameter is estimated by a single number which is quite risky job. This single value may or may not be equal to the true value of the parameter. The closeness of the estimate to the parameter value depends on random sample and the estimator. An ideal estimator is one which gives exactly the correct value of the parameter but such estimator does not exist in general. Hence to search a point estimate close to the parameter, it is necessary that the choice of one appropriate estimator in a given circumstance should be made on the basis of certain properties, called criteria for a good point estimator. A good point estimator is one which is unbiased, consistent,

NOT FOR SALE

efficient and sufficient. Only two properties, unbiasedness and efficiency are discussed here in detail.

6.2.3 Unbiasedness

It is not possible for an estimator to obtain correct estimate from each and every sample, even though if samples are drawn randomly. However, if this estimator on the average (i.e. considering all possible samples), give an estimate equal to the population parameter then this property of the estimator is called unbiasedness.

Definition of unbiased estimator:

An estimator is said to be unbiased if the mean of its sampling distribution is equal to the true value of the parameter, otherwise, the estimator is said to be biased. For example, if $\hat{\theta}$ is a point estimator of the parameter θ and $E(\hat{\theta})=\theta$, then $\hat{\theta}$ is called an unbiased estimator of θ. If $E(\hat{\theta})\neq\theta$ means that $\hat{\theta}$ is a biased estimator of θ. Further, if $E(\hat{\theta})>\theta$ means $\hat{\theta}$ is positively biased. If $E(\hat{\theta})<\theta$ means that $\hat{\theta}$ is negatively biased. Remember that in previous unit we have shown in practical examples that $E(\bar{X})=\mu_{\bar{x}}=\mu$. This implies that $\bar{X}=\frac{\sum x}{n}$ is an unbiased estimator for μ. Similarly, $E(\hat{p})=p$ means that $\hat{p}=\frac{X}{n}$ is an unbiased estimator of p. The bias of an estimator $\hat{\theta}$ is given by $B(\hat{\theta})=[E(\hat{\theta})-\theta]$.

Example 6.2

Show that sample mean $\bar{X}$ is an unbiased estimator of the population mean μ.

Proof:

Consider a random sample $X_1,\cdots,X_n$ from a normal population having mean μ and variance σ^2, then by definition,

NOT FOR SALE

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

Taking expectation on both sides to have

$$E(\bar{X})=\frac{1}{n}E\left[\sum_{i=1}^{n}X_i\right]$$

$$=\frac{1}{n}[E(X_1)+E(X_2)+\cdots+E(X_n)]$$

$$=\frac{1}{n}(\mu+\mu+\cdots+\mu)$$ (As X_i's are drawn from a population having mean μ)

$$=\frac{n\mu}{n}=\mu$$ This implies that $\bar{X}$ is an unbiased estimator of μ.

Example 6.3

Show that sample proportion $\hat{p}$ is an unbiased estimator of population proportion p.

Proof:

By definition $\hat{p}=\frac{X}{n}$

$$E(\hat{p})=E\left(\frac{X}{n}\right)=\frac{1}{n}E(X)$$

$=\frac{1}{n}(np)$ (As X is Binomial random variable having mean np)

$=p$. It means that $\hat{p}$ is an unbiased estimator of p.

Example 6.4

Show that sample variance $s^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ is an unbiased estimator of population variance σ^2.

NOT FOR SALE

Proof:

By definition $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

$(n-1)\, s^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2$

$= \sum_{i=1}^{n}(X_i - \mu + \mu - \bar{X})^2$ i.e. μ is added and subtracted

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

$= \sum_{i=1}^{n}[(X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu)]$

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

$= \sum_{i=1}^{n}(X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2(\bar{X} - \mu)n(\bar{X} - \mu)$ As $\sum_{i=1}^{n}(X_i - \mu) = n(\bar{X} - \mu)$

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

$= \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2$

Now taking expectation on both sides

$(n-1)E(s^2) = nE(X_i - \mu)^2 - nE(\bar{X} - \mu)^2$

$= n\sigma^2 - n\sigma_{\bar{x}}^2 \quad As\ \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$

NOT FOR SALE

$= (n-1)\sigma^2$

$E(s^2) = \sigma^2$

This implies that $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an unbiased estimator of σ^2.

Example 6.5

Draw all possible samples of size 2 with replacement from a population having elements 4, 8, 12, 16 and show that;

i) $E(\bar{X}) = \mu$ *i.e.* $\bar{X}$ is an unbiased estimator of μ.

ii) $E(s^2) = \sigma^2$ i.e. $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an unbiased estimator of σ^2.

iii) $E(S^2) \neq \sigma^2$ *i.e* $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is a biased estimator of σ^2.

Solution:

Given 4, 8, 12, 16; $N = 4$; and $n = 2$

Population mean $\mu = \frac{\sum X}{N} = \frac{4+8+12+16}{4} = \frac{40}{4} = 10$

Population variance $\sigma^2 = \frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2 = \frac{480}{4} - \left(\frac{40}{4}\right)^2 = 20$

NOT FOR SALE

Since sampling is with replacement therefore possible samples are $N^n = 4^2 = 16$

S.No	Samples	$\bar{X}$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x-\bar{x})^2$	$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^2$
1	(4,4)	4	0	0
2	(4,8)	6	8	4
3	(4,12)	8	32	16
4	(4,16)	10	72	36
5	(8,4)	6	8	4
6	(8,8)	8	0	0
7	(8,12)	10	8	4
8	(8,16)	12	32	16
9	(12,4)	8	32	16
10	(12,8)	10	8	4
11	(12,12)	12	0	0
12	(12,16)	14	8	4
13	(16,4)	10	72	36
14	(16,8)	12	32	16
15	(16,12)	14	8	4
16	(16,16)	16	0	0

i) The sampling distribution of $\bar{X}$

$\bar{X}$	Tally bar	f	$f(\bar{x})$	$\bar{X}f(\bar{x})$
4	I	1	1/16	4/16
6	II	2	2/16	12/16
8	III	3	3/16	24/16
10	IIII	4	4/16	40/16
12	III	3	3/16	36/16
14	II	2	2/16	28/16
16	I	1	1/16	16/16
Total	-	16	1	160/16

NOT FOR SALE

$$E(\bar{X}) = \mu_{\bar{x}} = \sum \bar{x} f(\bar{x}) = \frac{160}{16} = 10 = \mu$$

As mean of the sampling distribution of $\bar{X}$ is equal to the population mean μ, so by definition of unbiasedness we say that $\bar{X}$ is unbiased estimator of μ.

(ii) The sampling distribution of s^2

s^2	Tally bar	f	$f(s^2)$	$s^2 f(s^2)$
0	IIII	4	4/16	0
8	IIII I	6	6/16	48/16
32	IIII	4	4/16	128/16
72	II	2	2/16	144/116
Total		16	1	320/16

$$E(s^2) = \mu_{s^2} = \sum s^2 f(s^2) = \frac{320}{16} = 20 = \sigma^2$$

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

(iii) The sampling distribution of S^2

S^2	Tally bar	f	$f(S^2)$	$S^2 f(S^2)$
0	IIII	4	4/16	0
4	IIII I	6	6/16	24/16
16	IIII	4	4/16	64/16
36	II	2	2/16	72/16
Total		16	1	160/16

NOT FOR SALE

$$E(S^2)=\sum S^2 f(S^2)=\frac{160}{16}=10$$

$$E(S^2)\neq\sigma^2$$

Hence $S^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ is a biased estimator of σ^2.

Example 6.6

Draw all possible samples of size 2 with replacement from a population consisting of units 1,2,5,6 and find the proportion of even numbers in the samples.

Construct the sampling distribution of sample proportion $\hat{p}$ and check that $E(\hat{p})=p$ i.e. $\hat{p}$ is an unbiased estimator of p.

Solution:

Given 1, 2, 5, 6 $\quad N=4 \quad n=2$

Population proportion $p=\frac{X}{N}=\frac{2}{4}=0.5$

Possible samples of size $n = 2$ in S.W.R. case are $N^n = 4^2 = 16$ which are shown in the table.

S.No.	Samples	$\hat{p}$
1	1,1	0
2	1,2	1/2
3	1,5	0
4	1,6	1/2
5	2,1	1/2
6	2,2	1
7	2,5	1/2
8	2,6	1
9	5,1	0
10	5,2	1/2
11	5,5	0
12	5,6	1/2
13	6,1	1/2
14	6,2	1
15	6,5	1/2
16	6,6	1

The sampling distribution of sample proportions $\hat{p}$

$\hat{p}$	Tally bar	f	$f(\hat{p})$	$\hat{p}f(\hat{p})$
0	IIII	4	4/16	0
1/2	IIII III	8	8/16	4/16
1	IIII	4	4/16	4/16
Total		16		8/16

$$E(\hat{p})=\mu_{\hat{p}}=\sum \hat{p}\, f(\hat{p})=\frac{8}{16}=0.5=p.$$

It means that $\hat{p}=\frac{X}{n}$ is an unbiased estimator of population proportion p.

6.2.4 Efficiency

If there are two estimators, both possessing the property of unbiasedness which can be used for the estimation of a parameter, it is difficult for us to choose the best one between them. The efficiency property decides to prefer the one which has minimum variance. Hence efficiency is a selection criterion for efficient estimator between unbiasead estimators.

Definition of efficient estimator:

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators of the same parameter θ and $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$, then $\hat{\theta}_1$ is efficient estimator than $\hat{\theta}_2$.

Efficiency is generally expressed in relative terms as:

$$R.E=\frac{Variance\ of\ efficient\ estimator}{Variance\ of\ other\ estimator}$$

If $0 \le R.E < 1 \Rightarrow \hat{\theta}_1$ is efficient

If $R.E > 1 \Rightarrow \hat{\theta}_2$ is efficient.

NOT FOR SALE

If R.E = 1 $\Rightarrow$ both estimators are equally efficient.

For example, in case of normal distribution both sample mean and sample median are unbiased estimators for μ. However $Var(\bar{X}) = \frac{\sigma^2}{n}$ and $Var(\text{median}) = \frac{\pi\sigma^2}{2n}$. Now comparing these estimators by relative efficiency criteria as:

$$R.E = \frac{Var(\bar{X})}{Var(median)} = \frac{\sigma^2/n}{\pi\sigma^2/2n} = \frac{\sigma^2}{n}.\frac{2n}{\pi\sigma^2} = \frac{2}{\pi}$$

$$= \frac{2}{22/7} = \frac{14}{22} = 0.64 < 1$$

This implies that $\bar{X}$ is more efficient than median for μ. Thus, an unbiased estimator which has minimum variance is called the best or the most efficient estimator.

Example 6.7

A random sample X_1, X_2, X_3 is drawn from a normal population having mean μ and variance σ^2. Let $\hat{\theta}_1 = \frac{X_1 + 2X_2 + X_3}{4}$ and $\hat{\theta}_2 = \frac{X_1 + X_2 + X_3}{3}$ are the estimators for μ.

i) Which of the estimators is unbiased?

ii) Show that $\hat{\theta}_1$ is efficient estimator for μ than $\hat{\theta}_2$.

Solution:

i) To check unbiasedness, we first consider:

$$\hat{\theta}_1 = \frac{X_1 + 2X_2 + X_3}{4}$$

Take expectation on both sides

$$E(\hat{\theta}_1) = \frac{1}{4}E[X_1 + 2X_2 + X_3]$$

$$= \frac{1}{4}[E(X_1) + 2E(X_2) + E(X_3)]$$

$$= \frac{1}{4}[\mu + 2\mu + \mu]$$, As all X_i's have same mean μ

$$= \frac{4\mu}{4} = \mu$$ Hence $\hat{\theta}_1$ is unbiased estimator of μ.

Now consider the second estimator:

$$\hat{\theta}_2 = \frac{X_1 + X_2 + X_3}{3}.$$

$$E(\hat{\theta}_2) = \frac{1}{3}[E(X_1) + E(X_2) + E(X_3)]$$

$$= \frac{1}{3}[\mu + \mu + \mu]$$

$$= \frac{3\mu}{3} = \mu$$ Thus, θ_2 is also an unbiased estimator of μ.

(ii) To check the efficiency, we have first compute variances for both estimators.

$$Var(\hat{\theta}_1) = Var\left[\frac{X_1 + 2X_2 + X_3}{4}\right]$$

$$= \frac{1}{16}[Var(X_1) + 4Var(X_2) + Var(X_3)]$$

$$= \frac{1}{16}[\sigma^2 + 4\sigma^2 + \sigma^2]$$, As X_i's are drawn from the population having variance σ^2.

$$= \frac{6\sigma^2}{16} = \frac{3\sigma^2}{8}.$$

Similarly $Var(\hat{\theta}_2) = Var\left[\frac{X_1 + X_2 + X_3}{3}\right]$

NOT FOR SALE

$$=\frac{1}{9}[Var(X_1)+Var(X_2)+Var(X_3)]$$

$$=\frac{1}{9}[\sigma^2+\sigma^2+\sigma^2] \quad =\sigma^2/3$$

Now compare the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ as;

$$R.E=\frac{V(\hat{\theta}_1)}{V(\hat{\theta}_2)}=\frac{3\sigma^2/8}{\sigma^2/3}=\frac{3\sigma^2}{8}\times\frac{3}{\sigma^2}=\frac{9}{8}=1.125>1\Rightarrow\hat{\theta}_2 \text{ is efficient than } \hat{\theta}_1$$

Thus relative efficiency criteria shows that the estimator $\hat{\theta}_2$ is efficient than $\hat{\theta}_1$. It means that $\text{var}(\hat{\theta}_2)<\text{var}(\hat{\theta}_1)$ and hence $\hat{\theta}_2$ is the best estimator for the parameter μ.

6.2.5 Pooled estimator from two samples

Parameters can also be estimated by estimators which are obtained by pooling (combining) the estimators computed from two or more random samples drawn from the same population, such an estimator is called pooled estimator. For example two random samples of sizes n_1 *and* n_2 have means $\bar{x}_1, \bar{x}_2$ and variances s_1^2, s_2^2 respectively, then combined mean given as $\bar{x}_c=\frac{n_1\bar{x}_1+n_2\bar{x}_2}{n_1+n_2}$ is pooled estimator of μ and combined variance $s_p^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ is pooled variance estimate of σ^2. Similarly, if random samples of sizes n_1, n_2 are drawn from a binomial population having sample proportions $\hat{p}_1$, $\hat{p}_2$, then $\hat{p}_c=\frac{n_1\hat{p}_1+n_2\hat{p}_2}{n_1+n_2}$ is pooled estimator for p. If needed, pooled estimators from 3 or more samples can be obtained to extend in a similar way.

NOT FOR SALE

Example 6.8

Consider the following two samples

Sample A:	1	4	8	5
Sample B:	3	0	9	4

Compute pooled estimate for (i) population mean μ (ii) population variance σ^2 and (iii) proportion of odd numbers in the population.

Solution:

(i) $\bar{x}_1=\frac{1+4+8+5}{4}=\frac{18}{4}=4.5$ $\quad \bar{x}_2=\frac{3+0+9+4}{4}=\frac{16}{4}=4$

$\therefore \bar{x}_c=\frac{n_1\bar{x}_1+n_2\bar{x}_2}{n_1+n_2}=\frac{4(4.5)+4(4)}{4+4}=\frac{18+16}{8}=\frac{34}{8}=4.25$ is the pooled estimate of μ.

(ii) $s_1^2=\frac{1}{n_1-1}\left[\sum x_1^2-\frac{(\sum x_1)^2}{n_1}\right]=\frac{1}{4-1}\left[106-\frac{(18)^2}{4}\right]=\frac{1}{3}[25]=8.333$

$s_2^2=\frac{1}{n_{2-1}}\left[\sum x_2^2-\frac{(\sum x_2)^2}{n_2}\right]=\frac{1}{4-1}\left[106-\frac{(16)^2}{4}\right]=\frac{1}{3}[42]=14$

$\therefore s_p^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}=\frac{(4-1)(8.333)+(4-1)(14)}{4+4-2}$

$=\frac{24.999+42}{6}=\frac{66.999}{6}=11.17$ is the pooled estimate of σ^2.

(iii) Proportion of odd numbers in sample A: $\hat{p}_1=\frac{2}{4}=0.5$

Proportion of odd numbers in sample B: $\hat{p}_2=\frac{2}{4}=0.5$

$\therefore \hat{p}_c=\frac{n_1\hat{p}_1+n_2\hat{p}_2}{n_1+n_2}=\frac{4(0.5)+4(0.5)}{4+4}=\frac{2+2}{8}=0.5$ is the pooled estimate of p.

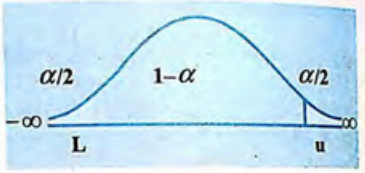
NOT FOR SALE

6.3 Interval estimation

This process provides an interval of values calculated from sample data, as an estimate for the unknown population parameter. Interval estimate has a high probability of containing the parameter of interest.

6.3.1 Confidence interval

An interval which is constructed from sample observations in such a way that it has a high probability of containing the unknown value of parameter is called confidence interval. For example if L and U are two statistics, then the confidence interval for the parameter θ is given as:



$$P[L<\theta<U]=1-\alpha$$

▪ Main points about confidence interval

i. The end points that bound a confidence interval i.e. L and U are called critical values or confidence limits where L is the lower confidence limit and U is the upper confidence limit.

ii. The region between L and U is called confidence interval or confidence region or acceptance region for the unknown parameter θ.

iii. The region beyond the acceptance region i.e. from $-\infty$ to L and U to ∞ is known as critical region or rejection region.

iv. The probability that the interval contains the parameter is called confidence coefficient or confidence level and is denoted by $(1-\alpha)$. It is also written as $100(1-\alpha)\%$.

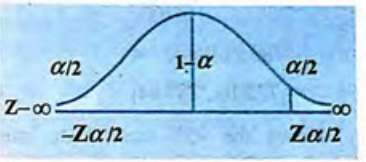
v. The probability that parameter lies in the rejection region is called significance level and is denoted by α.

vi. $(U-L)$ is called width or length of confidence interval which is a measure of precision for confidence interval.

vii. Precision means accuracy of the confidence interval. It can be increased either by increasing the sample size or decreasing the confidence coefficient.

6.3.2 Large sample confidence interval for population mean μ (when σ is known):

To find a $100(1-\alpha)\%$ large sample $(n \geq 30)$ confidence interval for population mean μ when σ is known, we begin with the statistic $Z=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ which has a standard normal distribution and make a probability statement as:



$$P[-Z_{\alpha/2} < Z < Z_{\alpha/2}] = 1-\alpha$$

Where $Z_{\alpha/2}$ means that probability or area to the right of $Z_{\alpha/2}$ is equal to $\alpha/2$. Similarly probability to the left of $-Z_{\alpha/2}$ is equal to $\alpha/2$ such that $\alpha/2+\alpha/2=\alpha$.

A $100(1-\alpha)\%$ confidence interval for μ when σ is known is given by;

$$\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Note for particular sample use $\bar{x}$ instead of $\bar{X}$.

Example 6.9

An electrical firm manufactures light bulbs that have a length of life with mean μ and a standard deviation of 40 hours. If a random sample of 100 bulbs has an average life of 780 hours, find a 95% confidence interval for the population mean of all bulbs produced by this firm.

Solution:

A $100(1-\alpha)\%$ confidence interval for μ is $\bar{X} \pm Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$

NOT FOR SALE

We are given $\sigma = 40$, $n = 100$, $\bar{x} = 780$, and confidence level $1-\alpha = 95\% = 0.95$ so $\alpha = 1-0.95 = 0.05$, $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

Now make inverse use of area table of standard normal distribution and search $0.5-0.025 = 0.475$ in the body of the table which correspond to $Z_{\alpha/2} = Z_{0.025} = 1.96$.

Putting values in the above interval estimator, we get

$$780 \pm 1.96 \frac{40}{\sqrt{100}}$$

$$780 \pm 1.96(4)$$

$$[772.16, 787.84]$$

Hence the 95% confidence interval for μ is from 772.16 to 787.84 hours.

Example 6.10

A random sample of size n = 400 selected without replacement from a population of size N = 2000 with $\sigma = 4$ gives $\bar{x} = 80$. Use this sample information to construct a 90% confidence interval for the mean of the population.

Solution:

A $100(1-\alpha)\%$ confidence interval for μ is $\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

Given that $N = 2000$, $n = 400$, $\sigma = 4$, $\bar{x} = 80$,

$1-\alpha = 0.90$, $\alpha = 0.10$, $\frac{\alpha}{2} = 0.05$.

From the area table we have $Z_{\alpha/2} = Z_{0.05} = 1.645$ corresponding to the probability $0.50 - 0.05 = 0.45$

Hence the 90% confidence interval for μ is

NOT FOR SALE

$$80 \pm 1.645 \frac{4}{\sqrt{400}} \sqrt{\frac{2000-400}{2000-1}}$$

or 80 ± 0.294

or $80 - 0.294$, $80 + 0.294$

or $[79.706, 80.294]$

6.3.3 Large sample confidence interval for population mean μ when σ is unknown

Practically, σ is usually not known but if $n \geq 30$ then, central limit theorem allows us to consider the sampling distribution of $\bar{X}$ as approximately normal with mean μ and S.E $= \frac{s}{\sqrt{n}}$ that is, σ is replaced by sample standard deviation s. An approximate $100(1-\alpha)\%$ confidence interval for μ in this case is given as;

$$\bar{X} \pm Z_{\alpha/2} \frac{s}{\sqrt{n}}.$$

Example 6.11

A scientist, interested in monitoring chemical contaminants in food, selected a random sample of $n = 50$ male adults. It was found that the average daily intake of dairy products was $\bar{x} = 756$ grams per day with a standard deviation of $s = 35$ grams per day. Construct a 95% confidence interval for the mean daily intake of dairy products for males.

Solution:

Here σ is unknown but n = 50 is large, therefore, $100(1-\alpha)\%$ confidence interval for μ is:

$$\bar{X} \pm Z_{\alpha/2} \frac{s}{\sqrt{n}}$$

We have $1-\alpha = 0.95$, $\alpha = 0.05$, $Z_{\alpha/2} = Z_{0.025} = 1.96$

Hence the approximate 95% confidence interval for μ is

NOT FOR SALE

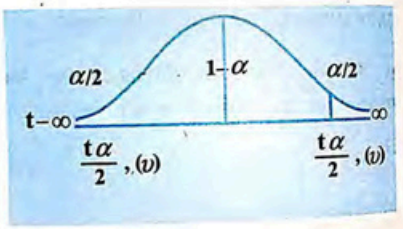
$756 \pm 1.96 \frac{35}{\sqrt{50}}$ or 756 ± 9.70

or $756-9.70$, $756+9.70$ or [746.30 , 765.70]

This means that the mean daily intake of dairy products for males varies from 746.30 to 765.70 grams per day.

6.3.4 Confidence interval for population mean μ when σ is unknown (small sample)

When σ is known and n is small, then the interval estimator $\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ will still be used for μ but if σ is unknown and $n < 30$, then the statistic $t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$ is used which has t-distribution with $v = n-1$ degrees of freedom and $s = \sqrt{\frac{1}{n-1}\sum(x-\bar{x})^2}$.



Probability statement can be made as:

$$P\left[-t_{\alpha/2,(v)} < t < t_{\alpha/2,(v)}\right] = 1-\alpha$$

Putting value of t to have a $100(1-\alpha)\%$ confidence interval for μ when σ is unknown and n is small as;

$$\bar{X} \pm t_{\alpha/2,(v)} \frac{s}{\sqrt{n}}$$

- **Degrees of freedom**

The number of values in the sample that are independent and free to vary is called degrees of freedom e.g. if $x_1 + x_2 = 10$. we can give value only to x_1 or x_2 the second one will be automatically computed. Hence $d.f = 2-1 = 1$. Generally if we have $x_1 + x_2 + \ldots\ldots x_n = 10$ we can gave values to first n-1 scores freely and one score is restricted, so we have degrees freedom $d.f = n-1$

TABLE 6.1 [Critical values of student t-distribution]

df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	df
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.204	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.146	2.650	3.012	13
14	1.345	14.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
∞	1.282	1.645	1.960	2.326	2.576	∞

NOT FOR SALE

Example 6.12

Find a 99% confidence interval for population mean when a random sample selected from the population gives the values 1.03, 1.01, 0.097, 1.04, 0.99, 0.98, 1.03, 1.01, 0.99.

Solution:

A $100(1-\alpha)\%$ confidence interval for μ is $\bar{X} \pm t_{\frac{\alpha}{2},(n-1)} \frac{s}{\sqrt{n}}$

We have values 1.03, 1.01, 0.97, 1.04, 0.99, 0.98, 1.03, 1.01, 0.99.

Here $n=9$, $\Sigma x=9.05$, $\Sigma x^2=9.1051$, therefore,

$$\bar{x}=\frac{\Sigma x}{n}=\frac{9.05}{9}=1.0056,$$

$$s=\sqrt{\frac{1}{n-1}\left[\Sigma x^2-\left(\frac{\Sigma x}{n}\right)^2\right]}=\sqrt{\frac{1}{9-1}\left[9.1051-\frac{(9.05)^2}{9}\right]}=0.02245$$

$1-\alpha=0.99$ or $\alpha=0.01$

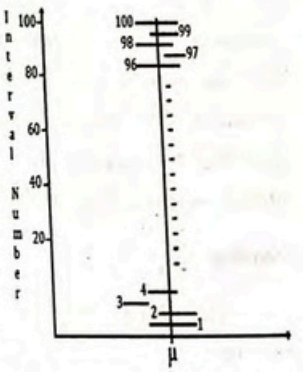
Now form t-table 6.1, we have $t_{\alpha/2,(n-1)}=t_{0.01/2,(9-1)}=t_{0.005,(8)}=3.355$

Putting values in expression for confidence interval, we get 99% confidence interval for μ as

$$1.0056 \pm 3.355\left(\frac{0.0245}{\sqrt{9}}\right)$$ or [1.0056 ± 0.0274], [0.9782 , 1.033]

6.3.5 Interpreting the confidence interval

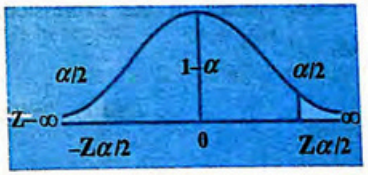
What does it mean to say we are 98% confidant that the true value of the population mean μ is within a given interval? It means that if we construct 100 such intervals for different samples, almost 98 out of 100 will contain the parameter and only 2 out of 100 will not contain the true value of the parameter. It can diagrammatically be shown as given in the Figure. Two of the intervals at serial number 3 and 97 do not contain the parameter μ while remaining 98 intervals contain the parameter. Remember that we cannot be absolutely sure that any one particular interval contains the mean μ.



6.3.6 Confidence interval estimate for the difference between two populations means when σ_1 and σ_2 are known

In this case $\bar{X}_1-\bar{X}_2$ is the point estimator of $(\mu_1-\mu_2)$ which in standard form can be written as;

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



Probability statement for the random variable Z can be made as $P[-Z_{\alpha/2}<Z<Z_{\alpha/2}]=1-\alpha$.

The $100(1-\alpha)\%$ confidence interval for $(\mu_1-\mu_2)$ in this case is given by

$$(\bar{X}_1-\bar{X}_2)\pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

NOT FOR SALE

Example 6.13

The wearing qualities of two types of automobile tires were compared by road – testing samples of $n_1 = n_2 = 100$ tires of each type. The test results are $\bar{x}_1 = 26400$ miles and $\bar{x}_2 = 25100$ miles. The standard deviations for the two types of populations are $\sigma_1 = 37.9473$ and $\sigma_2 = 44.2719$ respectively. Estimate $(\mu_1 - \mu_2)$, the difference in mean miles to wear out, using a 99% confidence interval.

Solution:

The $100(1-\alpha)\%$ confidence interval for $(\mu_1 - \mu_2)$ in this case is given by

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Given $n_1 = 100$ $\bar{x}_1 = 26400$ $\sigma_1 = 37.9473$ $\sigma_1^2 = 1440$

$n_2 = 100$ $\bar{x}_2 = 25100$ $\sigma_2 = 44.2719$ $\sigma_2^2 = 1960$

$1-\alpha = 0.99$ $\alpha = 0.01$ $\frac{\alpha}{2} = 0.005$

Hence $z_{\alpha/2} = 2.57$

Putting values in the expression, we have 99% confidence interval for $(\mu_1 - \mu_2)$ as;

$$(26400 - 25100) \pm 2.57\sqrt{\frac{1440}{100} + \frac{1960}{100}}$$

$$1300 \pm 2.57(5.831)$$

or $[1285.014 < (\mu_1 - \mu_2) < 1314.986]$

NOT FOR SALE

6.3.7 Confidence interval for $(\mu_1 - \mu_2)$ when σ_1 and σ_2 are unknown (large sample case)

When population variances are unknown but sample sizes n_1 and n_2 are large (≥ 30) then σ_1^2, σ_2^2 are estimated by sample variances S_1^2, S_2^2 respectively. The statistic Z in this case will be of the form

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

The $100\,(1-\alpha)\%$ confidence interval for $(\mu_1 - \mu_2)$ is then

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Example 6.14

Intelligence test of two groups of boys and girls gave the following results:

Girls: $n_1 = 60$ mean $= 75$ $S.D = 8$

Boys: $n_2 = 100$ mean $= 73$ $S.D = 10$

Compute a 95% confidence interval for $(\mu_1 - \mu_2)$ where μ_1 is the mean score of girls and μ_2 is the mean score of boys in respective populations.

Solution:

Here σ_1, σ_2 are not given but n_1, n_2 are large, so $100(1-\alpha)\%$ confidence interval for $(\mu_1 - \mu_2)$ in this case is

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

We have $n_1 = 60$, $\bar{x}_1 = 75$, $S_1 = 8$

$n_2 = 100$, $\bar{x}_2 = 73$, $S_2 = 10$

NOT FOR SALE

$1-\alpha = 0.95$, $\alpha = 0.05$, $\frac{\alpha}{2} = 0.025$

From area table of normal distribution, we have $Z_{\alpha/2} = Z_{0.025} = 1.96$

Hence the 95% confidence interval for $(\mu_1 - \mu_2)$ is:

$$(75-73) \pm 1.96 \sqrt{\frac{64}{60} + \frac{100}{100}}$$

$$2 \pm 1.96\ (1.44)$$

$$2 \pm 2.8224$$

$$(-0.8224\ ,\ 4.8224)$$

6.3.8 Confidence interval estimate for the difference between two normal populations means $(\mu_1 - \mu_2)$ when σ_1^2 , σ_2^2 are unknown (small sample case)

When population variances are not given and n_1 , n_2 are small (< 30) then we have to make another assumption i.e. $\sigma_1 = \sigma_2 = \sigma$ and this σ is estimated by unbiased pooled estimator. In this case the statistic $\bar{X}_1 - \bar{X}_2$ has the t-distribution with $(n_1 + n_2 - 2)$ degrees of freedom given below;

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$

A $100(1-\alpha)\%$ confidence interval for $(\mu_1 - \mu_2)$ in this case is given by

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2,(n_1+n_2-2)}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

NOT FOR SALE

Example 6.15

The following two samples were randomly selected from two normal populations for which $\sigma_1^2 = \sigma_2^2$ but unknown.

Sample-I	103	94	110	87	98		
Sample-II	97	82	123	92	175	88	118

Compute 90% confidence interval for the difference between the two population means.

Solution:

$100(1-\alpha)\%$ confidence Interval for $(\mu_1 - \mu_2)$ in this case is

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,(n_1+n_2-2)}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Given that

$n_1 = 5$, $\Sigma x_1 = 492$, $\Sigma x_1^2 = 48718$

$n_2 = 7$, $\Sigma x_2 = 775$, $\Sigma x_2^2 = 92019$

Now $\bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{492}{5} = 98.4$

$$\bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{775}{7} = 110.71$$

$$(n_1-1)s_1^2 = \Sigma(x_1 - \bar{x}_1)^2 = \Sigma x_1^2 - \frac{(\Sigma x_1)^2}{n_1} = 48718 - \frac{(492)^2}{5} = 305.2$$

$$(n_2-1)s_2^2 = \Sigma(x_2 - \bar{x}_2)^2 = \Sigma x_2^2 - \frac{(\Sigma x_2)^2}{n_2} = 92019 - \frac{(775)^2}{7} = 6215.4286$$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$$

NOT FOR SALE

$$= \sqrt{\frac{305.2+6215.4286}{5+7-2}} \quad = \sqrt{\frac{6520.6286}{10}} = 25.54$$

Here $1-\alpha = 0.90$, $\alpha = 0.10$

From $t-$table 6.1 we have $t_{\alpha/2,(n_1+n_2-2)} = t_{0.10/2,(5+7-2)} = t_{0.05,(10)} = 1.812$

Hence 90% confidence Interval for $(\mu_1 - \mu_2)$ is

$$(98.4-110.71) \pm 1.812(25.54)\sqrt{\frac{1}{5}+\frac{1}{7}}$$

$$-12.31 \pm 27.10$$

$$[-39.41 < \mu_1 - \mu_2 < 14.79]$$

6.3.9 Confidence interval estimate for population proportion p (large sample case)

When the sample size is large the sample proportion $\hat{p}$ has the sampling distribution which is approximately normal with mean p and $S.E = \sqrt{\frac{pq}{n}}$ i.e.

$$Z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} \sim N(0,1) \text{ as } n \to \infty$$

A $100(1-\alpha)\%$ confidence interval for p is written as;

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

If p is unknown then the confidence interval for p is computed by

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Example 6.16

A random sample of 200 persons from a city was interviewed and 50 of them were found to be literate. Calculate a 90% confidence interval for the proportion of literate persons in the city.

Solution:

A $100(1-\alpha)\%$ confidence interval for p (literate persons) is

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Given

$n = 200$, $X = 50$ (Number of literate persons), therefore

$$\hat{p} = \frac{X}{n} = \frac{50}{200} = 0.25, \quad \hat{q} = 1-\hat{p} = 1-0.25 = 0.75,$$

$1-\alpha = 0.90$, $\alpha = 0.10$. From area table of standard normal distribution, we have $Z_{\alpha/2} = Z_{\frac{0.10}{2}} = Z_{0.05} = 1.645$

Hence the 90% confidence interval for p is $0.25 \pm 1.645\sqrt{\frac{(0.25)(0.75)}{200}}$

or 0.25 ± 0.05 or $(0.2\ ,\ 0.3)$ or $(0.2 < p < 0.3)$

6.3.10 Large sample confidence interval estimate for the difference between two population proportions ($p_1 - p_2$)

A simple extension of the estimation of a binomial proportion p is the estimation of the difference between two binomial proportions. For sufficiently large sample sizes the sampling distribution of $\hat{p}_1 - \hat{p}_2$ can be approximated by a normal distribution i.e.

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} \sim N(0,1)$$

The approximate $100(1-\alpha)\%$ confidence interval for $(p_1 - p_2)$ is given by:

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

Example 6.17

Two pairs relieving drugs were compared each on independent samples of $n_1 = 1000$, $n_2 = 1000$ individuals. Out of these individuals 750 receiving drug-I and 800 receiving drug-II reported some pain relief. Construct a 90% confidence interval for the difference between population proportions.

Solution:

As p_1, p_2 (population proportions) are unknown, therefore $100(1-\alpha)\%$ confidence interval for $(p_1 - p_2)$ is $(\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

Given information are;

$$n_1 = 1000 \ , \quad X_1 = 750 \ , \ \hat{p}_1 = \frac{X_1}{n_1} = \frac{750}{1000} = 0.75$$

$$\hat{q}_1 = 1 - \hat{p}_1 = 1 - 0.75 = 0.25,$$

$$n_2 = 1000 \ , \quad X_2 = 800 \ , \ \hat{p}_2 = \frac{X_2}{n_2} = \frac{800}{100} = 0.80$$

$$\hat{q}_2 = 1 - \hat{p}_2 = 1 - 0.80 = 0.20,$$

$$1 - \alpha = 0.90 \ , \quad \alpha = 0.10 \ , \quad \frac{\alpha}{2} = 0.05.$$

From the area table of standard normal distribution, we have $Z_{\alpha/2} = Z_{0.05} = 1.645$.

Hence, the 90% confidence interval for $(p_1 - p_2)$ is

$$(0.75 - 0.80) \pm 1.645\sqrt{\frac{(0.75)(0.25)}{1000} + \frac{(0.80)(0.20)}{1000}}$$

$$-0.05 \pm 0.03$$

$$(-0.08 \ , \ -0.02)$$

NOT FOR SALE

Key points

- Statistical inference is a process of drawing conclusions (inferences) about the population on the basis of sample information from that population.
- Statistical estimation is a process by which the unknown value of a parameter is obtained from the sample observations.
- When a specific value is obtained from sample observations and is used to estimate the unknown value of the parameter, the process is called point estimation.
- When a range of values is obtained from sample observations within which the unknown value of parameter is believed to lie, the process is called interval estimation.
- A rule, usually expressed as a formula that tells us how to calculate an estimate from the sample data is called an estimator and the resulting number is called an estimate.
- An estimator is said to be unbiased if the mean of its sampling distribution is equal to the true value of the parameter, otherwise the estimator is said to be biased
- If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators of the same parameter θ and $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$ then $\hat{\theta}_1$ is called efficient estimator than $\hat{\theta}_2$.
- An estimator which is linear, unbiased and has minimum variances among a group of linear unbiased estimators is called best linear unbiased estimator or BLUE.
- A rule for calculating two numbers to create an interval which has a high probability of containing the parameter of interest is the confidence interval.
- The region between L and U is called confidence interval or confidence region or acceptance region for the unknown parameter θ.
- The region beyond the acceptance region i.e. from $-\infty$ to L and U to ∞ is known as critical region or rejection region.
- The probability that parameter lies in the rejection region is called significant level and is denoted by α.

NOT FOR SALE

Exercise

6.1 Read the following statements carefully and indicate which statement is true or false.

i. Statistical inference means making confidence interval for the parameter.

ii. Parameters are constant and statistics are random variables.

iii. Inference regarding population parameters can be done in three ways.

iv. A statistic will always be an unbiased estimator if the sample itself is chosen without bias.

v. $s^2 = \frac{1}{n-1}\Sigma(X-\bar{X})^2$ is an unbiased estimator of σ^2.

vi. In estimation procedure we estimate the value of a statistic.

vii. A 99% confidence interval will be wider than a 95% confidence interval constructed from the same data.

viii. The precision of the confidence interval will increase by increasing the sample size.

ix. Confidence limits will vary from sample to sample.

x. t-distribution is used for interval estimation of μ when σ is known and n is large.

6.2 Fill in the blanks.

i. An estimator is itself a _____.

ii. A value of an estimator is called an _____.

iii. If $x_1, x_2 ... x_n$ be a random sample, the expression $\bar{x} = \frac{\Sigma x}{n}$ is an _____.

iv. A single value of an estimator for a population parameter is called its _____estimate.

v. If expected value of an estimator $\hat{\theta}$ is equal to the value of the parameter θ, then $\hat{\theta}$ is said to be an _____ estimator of θ.

vi. If two estimators T_1 and T_2 such that $Var(T_1) < Var(T_2)$, then T_1 is called _____estimator than T_2.

vii. An interval estimate with _____ interval is best.

NOT FOR SALE

viii. The precision of a confidence interval increases by _____ the sample size.

ix. Estimation has _____ types.

x. $(1-\alpha)$ is known as _____.

6.3 Choose the Correct answer:

i. Estimate and estimator are:

(a) synonyms (b) different

(c) related to population (d) formulae

ii. Bias of an estimator can be:

(a) positive (b) negative

(c) either positive or negative (d) always zero

iii. If $\hat{p} = 0.5$, then 0.5 is called ;

(a) estimator (b) estimate

(c) interval (d) all of above

iv. Estimation has _____ types.

(a) 3 (b) 4

(c) 2 (d) 5

v. α is called ;

(a) level of significance (b) confidence level

(c) confidence coefficient (d) all of above

vi. The probability statement $P[-Z_{\alpha/2} < Z < Z_{\alpha/2}] = ?$

(a) α (b) β

(c) $1-\beta$ (d) $1-\alpha$

vii. If the average value of an estimator is equal to the true value of the parameter, the property is called;

(a) efficiency (b) sufficiency

(c) consistency (d) unbiasedness

viii. Statistical inference makes inferences about

(a) Sample (b) population

(c) both population and sample (d) estimator

NOT FOR SALE

ix. $\left[\hat{p}-Z_{\alpha/2}\sqrt{\frac{pq}{n}} < p < \hat{p}+Z_{\alpha/2}\sqrt{\frac{pq}{n}}\right]$ is an interval estimate for

(a) mean (b) variance

(c) proportion (d) S.D

x. A 90% confidence interval for the population mean is of the from

(a) $\bar{x} \pm 1.645\frac{\sigma}{\sqrt{n}}$ (b) $\bar{x} \pm 1.28\frac{\sigma}{\sqrt{n}}$

(c) $\bar{x} \pm 2.58\frac{\sigma}{\sqrt{n}}$ (d) $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$

6.4 What do you understand by estimation? Differentiate between an estimator and an estimate.

6.5 Describe the following:

i) Statistical inference ii) Types of estimates.

6.6 Distinguish between

i) Point estimate and interval estimate

ii) Estimator and estimate.

6.7 Define and discuss with examples the following properties of a point estimator;

i) unbiasedness. ii) efficiency.

6.8 What do you mean by bias? What are the factors which introduce bias?

6.9 Given a random sample 4, 3, 7, 8, 4, 10, 5, 5, 4, 9 Compute point estimate for (i) population mean (ii) population variance (iii) S.E of the mean.

6.10 If random samples of size n = 2 are drawn with replacement from population having values 10, 11, 13, 15, 16, 19.

Show that;

i) Sample mean is an unbiased estimator of population mean

NOT FOR SALE

ii) Sample variance $s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$ is unbiased estimator of the population variance.

6.11 A random sample of size 3 i.e. x_1, x_2, x_3 is drawn from a normal population having mean μ and variance σ^2. The following two statistics $T_1 = \frac{x_1+2x_2+x_3}{4}$, $T_2 = \frac{x_1+x_2+x_3}{3}$ are taken for μ.

i) Which of the above is unbiased?

ii) Which of the above is most efficient?

6.12 Describe the concept of confidence interval estimation.

6.13 Define the following terms:

i) confidence limits, ii) confidence coefficient iii) level of significance, iv) precision of a confidence interval.

6.14 A random sample 12, 9, 14, 10, 12, 07, 13, 11 is drawn from a normal population whose $\sigma = 2$. Compute a 90% confidence interval for the mean of this normal population.

6.15 A sample of 100 chocolate bars is taken at random from a large shipment have an average $\bar{x} = 0.8$ pound with a standard deviation of $S = 0.1$ pound. Find a 99% confidence interval for the mean weight (μ) of chocolate bars for the entire shipment.

6.16 An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed with a standard deviation of 42 hours. If a random sample of 49 bulbs has an average life of 800 hours. Find a 95% confidence interval for the population mean of all bulbs produced by this firm.

6.17 Given the sample 2.3, −0.2, −0.4, −0.9. Compute a 90% confidence interval for the mean of a normal population with σ=3.

6.18 The heights of a random sample of 100 college students showed a mean height of 64 inches. If standard deviation of the height distribution of the population is 3 inches, find a 95% confidence interval for the mean height of the population.

NOT FOR SALE

6.19 If $n=50$, $\Sigma x=2163$, $\Sigma x^2=144949.6$. Compute a 99% confidence interval for μ.

6.20 A random sample of size $n_1=36$ taken from a normal population with a variance $\sigma_1^2=9$ has mean $\overline{x}_1=75$. A second random sample of size $n_2=25$ taken from a different population with a variance $\sigma_2^2=25$ has a mean $\overline{x}_2=70$. Find a 98% confidence interval for $\mu_1-\mu_2$.

6.21 The number of accidents per day in two cities was observed and the following information was obtained:

Description	city-A	city-B
Numbers of days	144	100
Mean number of accidents	4.5	5.4
Standard deviation	1.2	1.5

Estimate 95% confidence interval for the difference between the mean accidents of the two cities.

6.22 When it is appropriate to use t-distribution instead of the Z-distribution to construct confidence interval for the population mean or difference of population means?

6.23 A sample of 10 measurements of the diameter of a spare gave a mean $\overline{x}=4.38$ inches and a standard deviation s = 0.06 inches. Find a 95% confidence interval for the actual diameter.

6.24 A random sample of 12 ball bearings has weights in grams as 31.4, 33.1, 35.9, 34.7, 33.4, 34.5, 35, 32.5, 36.9, 36.4, 35.8, 33.2. Find a 90% confidence interval for the mean weight of the population from which these weights were drawn.

6.25 The following summary statistics were recorded from independent random samples drawn from two populations:

	n	$\overline{x}$	s^2
Sample A	10	74	60
Sample B	13	81	40

Construct a 99% confidence interval for $\mu_A-\mu_B$

NOT FOR SALE

6.26 Find a 95% confidence interval for p if 24 heads are obtained in 40 tosses of a coin.

6.27 In a survey carried out in a large city 170 housewives out of a random sample of 250 preferred Lipton brand of tea. Find a 95% confidence interval for the percentage of all housewives in the city preferring Lipton brand of tea.

6.28 Independent random samples of $n_1=800$ and $n_2=640$ observations were selected from binomial populations 1 and 2, and $X_1=337$, $X_2=374$ successes were observed. Find a 90% confidence interval for the difference (p_1-p_2) in the two population proportions. Interpret the interval.

NOT FOR SALE