UNIT-III
INTERVAL ESTIMATION

* Interval estimation: The evaluation of parameter for mean $or a population by computing an interval or range of a values within the parameter lies b/w two means. This estimation is known as interval estimation. The use of sample data to estimate an interval of possible values of a parameter of interest and it is quite opposite to point estimation.

The importance of interval estimation is it provides an estimate of range in which the true population parameter is likely to lie along with a measure of level of confidence that which we have to estimate.

* Properties:
  * The length of the interval shows the precision with which we can estimate. θ
  * The smaller interval is the higher precision with which we can estimate θ.
  * Interval data are measured using continuous intervals that show order, direction and consistdifference in values.
  * The difference b/w values on an interval scale is always evenly distributed.
  * The most prevalent forms of interval estimation are confidence intervals.

* Confidence intervals:
  * The probability that we associate with an interval is called confidence level.
  * In the interval estimation the estimate for the parameter lies b/w 2 limits. these limits are known as confidence limits

* Confidence limits are also called as ~~fiduaical~~ 'fiduocial limits'

* If 't' is a sample statistic and θ is a population parameter the interval estimation of θ is the estimation of parameter with the help of interval $(t-s, t+s)$ an interval estimate consists & numerical values defining the range of values with a specified degree of confidence most likely includes the parameter estimated in general, an interval estimate is expressed as

→ If 't' is a sample statistic used to estimate corresponding population parameter θ then $(1-\alpha)100\%$ confidence limits for θ are $t \pm S.E(t)$ where α is level of significance. $S.E(t)$ is standard error of 't' ~~tα~~ significant or critical value of 't' at significant level of α, $t \pm S.E(t)$ are the limits of confidence interval, then, the interval estimate is

$$\boxed{z = \frac{t - E(t)}{S.E(t)}}$$ is a standard normal variate with mean and variance i.e, $\mu = 0, \sigma = 1$ as $n \to \infty$

* Interval estimations for large samples:

* For large samples the underlying distribution of standardized variate corresponding to the sample distribution of statistic 't' will be asymptotically normal by a ~~cent~~ central limit theorem

$$\boxed{z = \frac{t - E(t)}{SE(t)} \sim N(0,1)}$$

* Central limit theorem: If $x_1, x_2 --- x_n$ is a random sample of size 'n' from any population then, the sample mean $(\bar{x})$ is normally distributed with mean μ and variance $\frac{\sigma^2}{n}$ provided 'n' is

Sufficiency large $\bar{x} \sim N(\mu, \sigma^2/n)$
i.e, $n \to \infty$

By using this theorem it has proved that sampling distribution of most of the statistics like sample proportion (p), difference of sample proportion (P_1-P_2), difference of sample mean $(\bar{x}_1-\bar{x}_2)$ & difference of sample standard deviation (s_1-s_2) as are asymptotically normal distribution i.e, $z=\frac{t-E(t)}{S.E(t)} \sim N(0,1)$ as $N \to \infty$ (asymptotically)

* This result the distribution any statistic in its standard form is asymptotically normal as $n \to \infty$ and also is extensively used in large sample tests used in construction of confidence limits for population parameter with samples are large.

* Confidence limits for mean (μ):

If we consider a large random sample of size 'n' from infinite population with mean (μ) & variance σ^2/n then the sample mean, is $\bar{x} \sim N(\mu, \sigma^2/n)$ from this

$$z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

for practical samples may be regarded as large if the sample is > 30, i.e, $n > 30$

* 95% Confidence:

From areas under probability curve we have 95% confidence limits as 1.96 values from this $P(-1.96 \le z \le +1.96)$

$$= 0.95$$

$$= P\left(-1.96 \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le 1.96\right)$$

$$P(\bar{x}-1.96\,\sigma/\sqrt{n} \le \mu \le \bar{x}+1.96\,\sigma/\sqrt{n})$$

Thus, 95% confidence limits for population mean are $\bar{x} \pm 1.96\,\sigma/\sqrt{n}$ & where σ is assumed as known. & the interval is $(\bar{x}-1.96\,\sigma/\sqrt{n},\ \bar{x}+1.96\,\sigma/\sqrt{n})$ is the 95%

Confidence limits for the estimating of sample mean μ.

* 99% Confidence limits:

The 99% confidence limits the value is given as 2.58. from $P(-2.58 \le z \le +2.58) = 0.99$

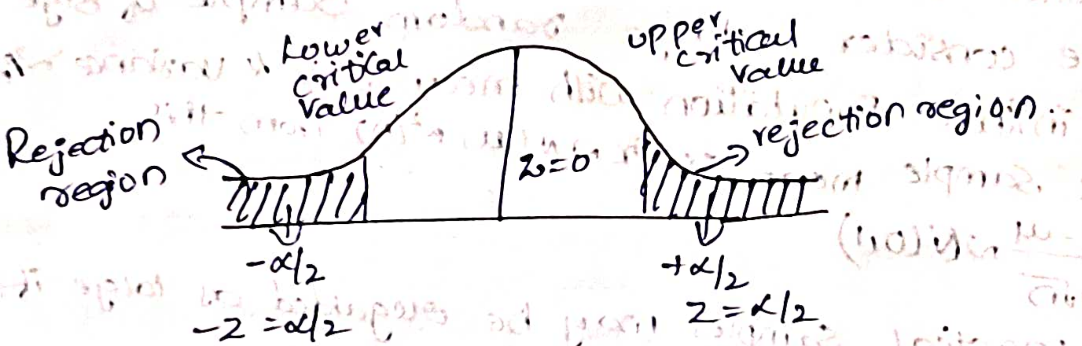
$$P\left(-2.58 \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le 2.58\right)$$

$$P(\bar{x} - 2.58\,\sigma/\sqrt{n} \le \mu \le \bar{x} + 2.58\,\sigma/\sqrt{n})$$

99% interval $(\bar{x} - 2.58\,\sigma/\sqrt{n},\ \bar{x} + 2.58\,\sigma/\sqrt{n})$. Same as above

* Two-Tailed Significant (critical) values of Z:

Confidence limits $(1-\alpha)$	50%	68.27%	90%	95%	95.45%	98%	99%	99.73%
Significant values	0.6745	1	1.645	1.96	2	2.33	2.58	3



Example 1: A random sample of 100 observations is yields a sample mean $\bar{x} = 150$ & sample variance $\sigma^2 = 400$, compute 95% and 99% confidence limits for population mean.

Sol 95% Given

$n = 100$

$\bar{x} = 150$.

$s^2 = 400,\ s = 20$

i) 95% confidence limits:

$$\bar{x} \pm 1.96\,\sigma/\sqrt{n}$$

95% interval $(\bar{x} - 1.96\,\sigma/\sqrt{n},\ \bar{x} + 1.96\,\sigma/\sqrt{n})$

$$=\left(150-1.96\left(\frac{20}{\sqrt{100}}\right), 150+1.96\left(\frac{200}{\sqrt{100}}\right)\right)$$

$$=(146.08, 153.92)$$

$\therefore 146.08 \le \mu \le 153.92$.

ii) 99% confidence limits:

$x \pm 2.58\ \sigma/\sqrt{n}$

Interval:

$(\bar{x} - 2.58\ \sigma/\sqrt{n},\ \bar{x} + 2.58\ \sigma/\sqrt{n})$

$\left(150 - 2.58\left(\frac{20}{100}\right), 150 + 2.58\left(\frac{20}{\sqrt{100}}\right)\right)$

$(144.84, 155.16)$

$\therefore 144.84 \le \mu \le 155.16$.

* Confidence limits for proportion:

* Taking $t = p$ for ~~range~~ large sample 'n' (>30) we get $z = \frac{p - E(p)}{S.E(p)} \sim N(0,1)$

$$= \frac{p - P}{\sqrt{\frac{PQ}{n}}} \sim N(0,1)$$ follows normal distribution.

* In sampling from an infinite population then, ~~100α%~~ $100(1-\alpha)\%$ confidence limits for P are given as $p \pm z_{\alpha/2}\ S.E(p) = p \pm z_{\alpha/2}\sqrt{pq/n}$ where, $\sqrt{pq/n}$ is estimate of standard error (p) (or) $S.E(p) = \sqrt{\frac{PQ}{n}}$ because P is unknown, hence, 95% and 99% confidence limits for P, for large samples, are given as follows

$$95\% = p \pm 1.96\sqrt{\frac{pq}{n}}$$

$$99\% = p \pm 2.56\sqrt{\frac{pq}{n}}$$

* Confidence limits for difference of means:

* If $\bar{x}_1$ and $\bar{x}_2$ are the sample means based on

large and independent random Samples of size n_1 and n_2 from two infinite population with mean μ_1 and μ_2 and standard deviations σ_1 and σ_2 resp-ectively then.

$$E[\bar{x}_1 - \bar{x}_2] = E[\bar{x}_1] - E[\bar{x}_2]$$
$$= \mu_1 - \mu_2$$

and
$$var[\bar{x}_1 - \bar{x}_2] = var[\bar{x}_1] - var[\bar{x}_2]$$
$$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\therefore S.E[\bar{x}_1 - \bar{x}_2] = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Thus, for large samples by central theorem. we get,
$$z = \frac{(\bar{x}_1 - \bar{x}_2) - E[\bar{x}_1 - \bar{x}_2]}{S.E[\bar{x}_1 - \bar{x}_2]}$$
$$= \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Hence, the confidence limits for the difference of means $\mu_1 - \mu_2$ of confidence co-efficient $(1-\alpha)$ are $(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} S.E(\bar{x}_1 - \bar{x}_2)$

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

If σ_1^2 and σ_2^2 are unknown then they estimates provides by corresponding sample ~~abs~~ variations s_1^2 and s_2^2 are used for large samples. i.e.

$\hat{\sigma}_1^2 = s_1^2$ and $\hat{\sigma}_2^2 = s_2^2$ and $100(1-\alpha)\%$ confid-ence limits for μ_1 and μ_2 are given by

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Problems :-

1) A sample of 100 gave a mean of 7.4 kg and a S.d of 1.2kg find 95% confidence limits for the Population mean.

Sol Given $n = 100$

$\text{mean } \bar{x} = 7.4 \text{ kg}$

$SD\ \sigma = 1.2 \text{ kg}$

95% confidence limits

$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$

$7.4 \pm 1.96 \frac{1.2}{\sqrt{100}}$

$(7.6352, 7.1648)$

2) A random sample of 700 units from a large consinment that 200 were damaged find 95% and 99% confidence limits for the proportion of damaged units in the sample

Sol $n = 700$

P = Damaged values

$P = \frac{200}{700} = 0.285$

$q = 1 - 0.28 = 0.715$

i) 95% Confidence limits

$= P \pm 1.96 \sqrt{\frac{Pq}{n}}$

$= 0.285 \pm 1.96 \sqrt{\frac{0.285 \times 0.715}{700}}$

$= (0.318, -0.048)$

ii) 99% Confidence limits

$= P \pm 2.56 \sqrt{\frac{Pq}{n}} \Rightarrow 0.285 \pm 2.56 \sqrt{\frac{0.285 \times 0.715}{700}}$

$= (0.328, 0.241)$

(3) A random sample of 100 items taken from a large batch of articles contains 5 defective items. find i) 95% confidence limits for the proportion of defective limits in the batch.

ii) If the batch contain 2669 items are defective then find 95% confidence limits for the proportion of defective items

Sol

i)

Given $n = 100$

$p =$ damaged values

$$p = \frac{5}{100} = 0.05$$

$$q = 1 - p$$

$$q = 1 - 0.05$$

$$q = 0.95$$

Standard error

$$S.E = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.05)(0.95)}{100}}$$

$$= 0.0217$$

i) 95% confidence limits :-

$$p \pm 1.96\, S.E$$

$$0.05 \pm 1.96(0.0217)$$

$$(0.091, \cancel{0.06}\ -0.006)$$

ii) 9 2669 = N

$$p \pm 1.96 \sqrt{\frac{(N-n)pq}{\cancel{N(n-1)}\ N(n-1)}}$$

2.0100

$$= 0.05 \pm 1.96 \sqrt{\frac{(2669-100) \times 0.05 \times 0.95}{2669(100-1)}}$$

$$= (0.050, 0.049)$$

(4) A random sample of $n=50$ males showed a mean avg. daily intake of diary Products = 756 gms with a S.D of 35 gms find 95% and 99% confidence interval for the Population avg μ.

Sol Given $n=50$

$\bar{x} = 756$ gms

S.D $\sigma = 35$ gms.

i) 95% Confidence limits

$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$

$756 \pm 1.96 \left(\frac{35}{\sqrt{50}}\right)$

$= (765.70, 746.298)$

ii) 99% Confidence limits

$x \pm 2.56 \left(\frac{\sigma}{\sqrt{n}}\right)$

$756 \pm 2.56 \left(\frac{35}{\sqrt{50}}\right)$

$= (768.67, 743.328)$.

* Confidence limits for difference for proportions if P_1 and P_2 are sample proportions in large and independent random samples of sizes n_1 and n_2 respectively from infinite Populations with proportions P_1 and P_2 ~~exepeta~~ expectation $E(P_1 - P_2) - E(P_1) - E(P_2) = P_1 - P_2$. Since the samples are independent then variance $var(P_1 - P_2) = var(P_1) - var(P_2)$ $= \frac{P_1 Q_1}{n_1} - \frac{P_2 Q_2}{n_2}$, Hence for large samples ~~$Z = (P_1 - P_2) - E$~~

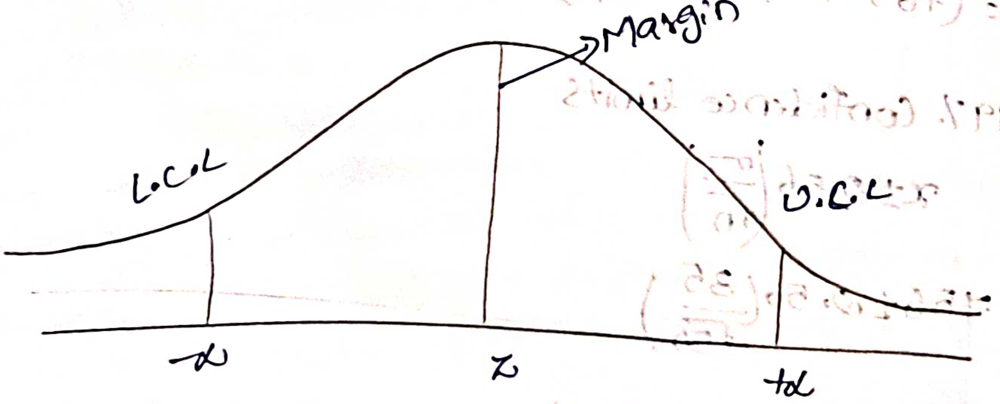
$Z = \frac{(P_1 - P_2) - E(P_1 - P_2)}{S.E(P_1 - P_2)} \sim N(0,1)$, $Z = \frac{(P_1 - P_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 Q_1}{n_1} - \frac{P_2 Q_2}{n_2}}} \sim N(0,1)$

Hence, the confidence limits for the difference of Population proportions P_1-P_2 at level of significance (P_1-P_2) $\pm Z_{\alpha/2} S.E(P_1-P_2)$ since P_1 and P_2 are unknown their un-baised estimates P_1 and P_2 are used instead of P_1Q_1 is p_1q_1 thus, $(1-\alpha)\%$ confidence limits for P_1-P_2 are

$$(P_1-P_2) \pm Z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1}-\frac{p_2q_2}{n_2}}$$

* Confidence interval:

* A confidence interval is the mean of your estimate +& - of variation in that estimate i.e., a range with an upper and lower number calculated from a sample. it is denoted as 'Z' It can be written as (L.C.L ± Z ± U.C.L)



Point interval
$\bar{x} \pm z\left(\frac{\sigma}{\sqrt{n}}\right)$

(Pb) A random sample of 500 pineapples was taken from a large consinment and 65 of them were found to be bad. Show that 95% and 99% confidence interval with S.D that the % of bad pineapples in that consinment

Sol Given,

$n=500$

$p=\frac{65}{500}=0.130$, $q=1-p=1-0.130$
$=0.870$

i) 95% confidence interval

$x \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$

$P \pm 1.96\sqrt{\frac{pq}{n}}$

$0.130 \pm 1.96\sqrt{\frac{0.130 \times 0.870}{500}}$

$= (0.159, 0.100)$

ii) 99% confidence interval.

$P \pm 2.56\sqrt{\frac{pq}{n}}$

$0.130 \pm 2.56\sqrt{\frac{0.130 \times 0.870}{500}}$

$= (0.168, 0.091)$

(Pb) Builiding statement for 1000 Patients discald ~~for~~ from a Particular hospital were randomly selected out of ~~thousand~~ thousand 1000 building statement ~~one of 2~~ 102 were found an errors construct 95 & 99% confidence intervals

Sol Given, $n = 1000$

$p = \frac{102}{1000} = 0.102, q = 1 - 0.102 = 0.898$

i) 95% confidence limits

$P \pm 1.96\sqrt{\frac{pq}{n}}$

$\Rightarrow 0.102 \pm 1.96\sqrt{\frac{(0.102)(0.898)}{1000}}$

$\Rightarrow (0.120, 0.083$

ii) 99% confidence limits

$P \pm 2.56\sqrt{\frac{pq}{n}}$

$\Rightarrow 0.102 \pm 2.56\sqrt{\frac{(0.102)(0.898)}{1000}}$

$(0.126, 0.077)$

(P5) An electrical manufactures light ~~bulb~~ bulbs that have a length of life with mean μ and a S.D of 40hrs if a sample of 100 bulbs has an avg life 780hr find 95% confidence interval for the population mean of all bulbs produced by electrician.

Sol

$n = 100$

$\bar{x} = 780$

S.D $= 40$hrs

95% confidence interval:

$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$

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

$(787.84, 772.16)$

* Construction of confidence intervals: Construction of a confidence interval about mean (μ) and S.D(σ) is known Z_α is critical value. E = margin of error $x \pm E$ is point estimate or interval. There are 4 steps in constructing of confidence intervals

Step 1: Calculate the sample proportion 'p' to calculate the sample proportion P divide the no. of successful items x by the sample size N.

Step 2: Choose the confidence level i.e 95% or 99% etc

Step 3: Calculate the margin of error by estimating formula which was given as interval

Step 4: Estimate the confidence interval at a particular level of significance

By this we construct the confidence interval by the given level of α%.

Imp ~~Dupe~~ ~~Duality~~
Duality b/w acceptance region of a test and a confidence interval:

The duality b/w acceptance region of testing and confidence region is about their validity i.e., statisfing the significance level of given problem and the confidence level of the sum to be calculated as statistical inference.

* For confidence interval it can mean the length of the interval is minimum.

* For accepting region of testing it can mean the power of testing has uniformly maximum power or the length of the acceptance region is minimum.

* There is an important duality b/w confidence intervals and testing i.e., we can obtain a confidence interval by inverting a hypothesis test and viseversa a confidence interval is also known as acceptance region. for test statistic for which the null hypothesis is acceptable and this confidence interval is used to estimate a population parameter from a given sample.

* Differences b/w mean ~~and~~ ratio and ~~of~~ proportion.

Mea Ratio	Proportion
* Defines the quantitative relationship b/w 2 or more quantities	* Indicates the equality of 2 ratios
* It can be expressed as a fraction using the colon form ":" or in decimal form	* using the It is usually represented by using equals to sign "=".
* It focus on comparision of quantities	* It emphasizes the relationship b/w parts and whole.
* A ratio can exist	* The proportion requires comparision
* Doesn't guarantee an equal distribution.	* It ensures a balanced distribution.

Ex: The ratio of boys to girls is 3:2

* It can have infinite no. of Possible values

* Ratios can be simplified or expanded by multiplying or dividing both sides by the same number.

Ex: In a group of 50 Stud-ents the ratio of boys to girls is 3:2.

* Generally, a proportion has only one solution.

* Proportions can be solved using cross multiplication metho-d.