# BHARATHIDASAN UNIVERSITY

## Tiruchirappalli- 620024

## Tamil Nadu, India.

## Programme: M.Sc. Statistics

## Course Title: Statistical Inference-I

## Course Code: (23ST06CC)

# Unit-IV

# Interval Estimation

**Dr. T. Jai Sankar**
**Associate Professor and Head**
**Department of Statistics**

**Ms. N. Saranya**
**Guest Faculty**
**Department of Statistics**

# Unit-4

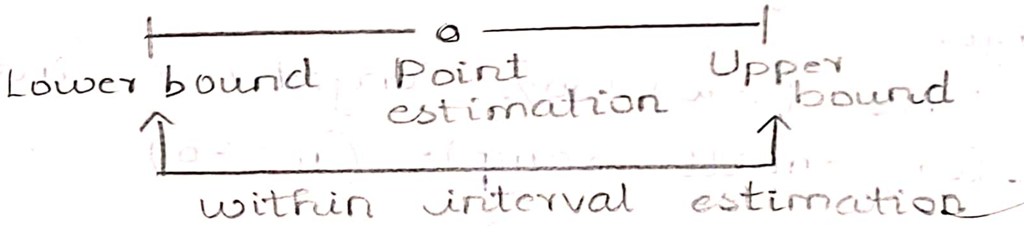
## Interval Estimation

(An interval estimate is defined by two numbers between which a population parameter is said to lies

For example:

$a < \bar{x} < b$ is an interval estimate of the population mean $\mu$. It indicates that the population mean is greater than a but less than b.)

A confidendence interval provides additional information about variability




Confidence Interval:

[A range of value constructed from the sample data so that the population parameter is likely to occur within that range at a specified probability. Specified probability is called the level of confidence.]

It states how much confidence we have that this interval contains the true population parameter. The confidence level is denoted by $(1-\alpha)100\%$.

Example:

95% level of confidence would mean that if 100% confidence interval were constructed each based on the different sample from the same population. we would expect 95% of the intervals to contain the population mean.

To compute a confidence interval, we will consider two situations:

We use sample data to estimate new with $\bar{x}$ and the population s.d ($\sigma$) is unknown. In this case, we substitute the sample standard deviation for the population standard deviation ($\sigma$).

Confidence Interval Estimates of the population mean ($\mu$):

The $(1-\alpha)100\%$ confidence interval for $\mu$ for large sample ($n > 30$)

i) $\bar{x} \pm Z_{\alpha/2}\ \sigma/\sqrt{n}$ if $\sigma$ is known and normally distributed population.

ii) $\bar{x} \pm Z_{\alpha/2}\ S/\sqrt{n}$, if $\sigma$ is not known $n$ large.

The $(1-\alpha)\ 100\%$ confidence interval $\mu$ for small sample ($n \leq 30$)

$x \pm t_{n-1,\alpha/2}\ (S/\sqrt{n})$. if $\sigma$ is not known

Example 1:

Find 95% confidence limit interval of a population mean for these values

a) $n = 36$, $\bar{x} = 13.3$ and $S^2 = 3.42$

b) $n = 64$, $\bar{x} = 2.73$ and $S^2 = 0.1047$

Soln:

a) 1st step: $(1-\alpha)100 = 95$

$1-\alpha = 0.95$

$\alpha = 0.05$

$\alpha/2 = 0.025$

find from table $Z_{0.025} = 1.96$

$$C.I = \bar{x} \pm Z_{\alpha/2}\left(S/\sqrt{n}\right)$$

$$= 13.3 \pm 1.96\left(\frac{1.8493}{\sqrt{36}}\right)$$

$$= 13.3 \pm 1.96\,(0.30821)$$

$$= 13.3 \pm 0.60409$$

$$= 13.90409 \;,\; 12.69591$$

$$C.I = 13.90 \;,\; 12.69$$

$$C.I = \bar{x} \pm Z_{\alpha/2}\left(S/\sqrt{n}\right)$$

$$= 2.73 \pm 1.96\left(\frac{0.3235}{8}\right)$$

$$= 2.73 \pm 1.96\,(0.0404)$$

$$= 2.73 \pm 0.0791$$

$$= 2.8091 \;,\; \cancel{0.0791}\; 2.6509$$

$$C.I = 2.80 \;,\; 2.65$$

Ex:2

The brightness of a television picturetube can be evaluated by measuring the amount current required to achieve a particular brightness level. A random sample of 10 tube indicates a sample mean 317.2 micro. Sample S.D is 15.7. Find the 99% C.I estimates for mean current require achieve a brightness level.

Soln:

$S. = 15.7$

$\bar{X} = 317.2$

$n = 10$

For 99% $C.I = 1-\alpha\,(100)$

$1-\alpha = 0.99$

$\alpha = 0.01$

$\alpha/2 = 0.005$

From t normal distribution table.

$t_{\alpha/2, n-1} = t_{0.005,9} = 3.250$

Hence, $99\%\ C.I = 317.2 \pm t_{0.005,9}\left(\frac{15.7}{\sqrt{10}}\right)$

$= 317.2 \pm 3.250\left(\frac{15.7}{3.1622}\right)$

$= 317.2 \pm 3.250\,(4.9648)$

$= 317.2 \pm 16.1356,$

$= 333.3356,\ 301.064$

Thus we are 99% C.I that the mean current required to achieve a particular brightness level is between 333.33 and 301.64

Confidence level for proportion:

Confidence interval estimates of the proportion $\hat{p}$ for large sample ($np \geq 5$ and $n(1-p) \geq 5$).

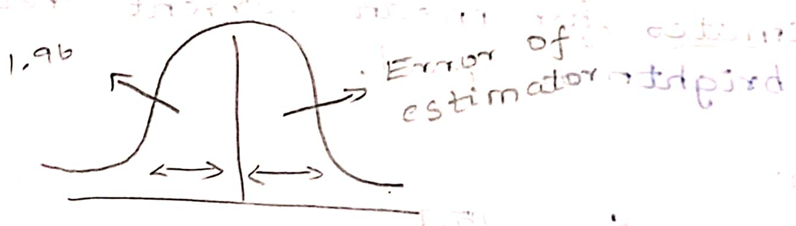
The $(1-\alpha)100\%$ C.I for $\hat{p}$ large sampl

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \text{ where } (1-\hat{p}) = \hat{q}.$$

The $(1-\alpha)100\%$ confidence interval for $(\hat{P}_1 - \hat{P}_2)$ for large sample ($n_1 \geq 30$, $n_2 \geq 30$)

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

Margin for error $1.96 \times \frac{\sigma}{\sqrt{n}}$ where $\sigma/\sqrt{n}$ is the standard error of the estimator




$C.I = \bar{x} \pm Z_c\,(S/\sqrt{n})$

$\bar{x}$ = Sample mean

$Z_c$ = Z value of confidence level

$S$ = Sample S.D

$n$ = no. of elements in a sample

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

90% confidence interval for $Z_{\alpha/2} = 1.65$

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

99% confidence interval for $Z_{\alpha/2} = 2.58$

## Confidence intervals:

1) Population mean for large sample
$$\bar{x} \pm Z_{\alpha/2}\left(s/\sqrt{n}\right)$$

2) Population mean normal data with unknown variance.
$$\bar{x} \pm t_{\alpha/2, n-1}\left(s/\sqrt{n}\right)$$

3) Difference of two means independent sample
$$\bar{x} - \bar{y} \pm Z_{\alpha/2}\sqrt{\frac{s^2x}{nx} + \frac{s^2y}{ny}}$$

4) Difference of two means matched pairs
$$\bar{x} - \bar{y} + Z_{\alpha/2}\left(\frac{sd}{\sqrt{n}}\right)$$
$$\bar{x} - \bar{y} \pm Z_{n-1, \alpha/2}\left(\frac{sd}{\sqrt{n}}\right)$$

5) Population proportion large sample
$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

6) difference of two population, proportion independent large samples.
$$\hat{p}_x - \hat{p}_y \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$$

## ~~Confidence~~ interval for variance:

* A single population variance $\sigma^2$
* The ratio of two population variance $\frac{\sigma_x^2}{\sigma_y^2}$ (or) $\sigma_y^2/\sigma_x^2$

## One Variance:

Let $x_1, \ldots x_n$ are normally distributed and

$a = \chi^2_{1-\alpha/2, n-1}$ and $b = \chi^2_{\alpha/2, n-1}$

Then a $(1-\alpha)$% C.I for population variance $\sigma^2$ is
$$\left\{\frac{(n-1)s^2}{b} \le \sigma^2 \le \frac{(n-1)s^2}{a}\right\}$$

Example 1::

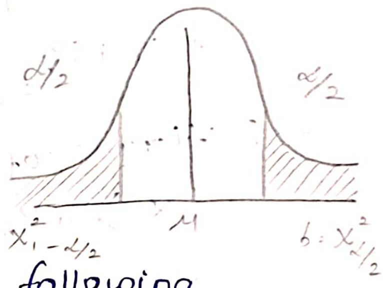
$(1-\alpha)\%$ C.I for the population S.D $\sigma$ is

$$\left(\left(\frac{\sqrt{n-1}}{\sqrt{b}}\right) s \le \sigma \le \left(\frac{\sqrt{(n-1)}}{\sqrt{a}}\, s\right)\right)$$

Proof:

We learned previously that if $X_1, X_2 \dots X_n$ are normally distributed with mean $\mu$ and population variance $\sigma^2$ then,

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$




with

$a = \chi^2_{1-\alpha/2}$

$b = \chi^2_{\alpha/2}$, we can write the following probability statement

$$P\left[a \le \frac{(n-1)s^2}{\sigma^2} \le b\right] = 1-\alpha$$

now

$$a \le \frac{(n-1)s^2}{\sigma^2} \le b$$

Taking the reciprocal of all the term and there by changing the directio of the inequalities we get

$$1/a \ge \sigma^2/(n-1)s^2 \ge 1/b$$

Now multiplying $(n-1)s^2$ we get confidenc intervals for $\sigma^2$

$$\frac{(n-1)s^2}{b} \le \sigma^2 \le \frac{(n-1)s^2}{a},$$

$$\frac{\sqrt{(n-1)s^2}}{\sqrt{b}} \le \sigma \le \frac{\sqrt{(n-1)s^2}}{\sqrt{a}}.$$

Hence the proof.

A ratio of two population variance:

If $X_1, X_2\, X_n \sim N(\mu_x, \sigma_x^2)$ and $Y_1, Y_2 \dots Y_n$ $\sim N(\mu_y, \sigma_y^2)$ are independent random samples.

1) $c = F_{1-\alpha/2}(m-1, n-1) = \dfrac{1}{F_{\alpha/2}(n-1, m-1)}$ and

2) $d = F_{\alpha/2}(m-1)(n-1)$, then a $(1-\alpha)100\%$ confidence

interval for $\sigma_x^2 / \sigma_y^2$ is

$$\frac{1}{F_{\alpha/2}(n-1, m-1)} \, S_x^2/S_y^2 \leq \frac{\sigma_x^2}{\sigma_y^2} \leq F_{\alpha/2}(m-1, n-1)\frac{S_x^2}{S_y^2}$$

Proof:

Because $X_1, X_2, \ldots X_n \sim N(\mu_x, \sigma_x^2)$

$$\frac{(n-1)S_x^2}{\sigma_x^2} \sim \chi^2_{n-1} \text{ and } \frac{(m-1)S_y^2}{\sigma_y^2} \sim \chi^2_{m-1}$$

$$F = \frac{[(m-1)S_y^2/\sigma_y^2]/m-1}{\frac{[(n-1)S_x^2/\sigma_x^2]}{n-1}} = \frac{\sigma_x^2}{\sigma_y^2} \cdot \frac{S_y^2}{S_x^2} \sim F(m-1)(n-1)$$

$$P\left[F_{1-\alpha/2}(m-1, n-1) \leq \frac{\sigma_x^2}{\sigma_y^2} \cdot \frac{S_y^2}{S_x^2} \leq F_{\alpha/2}(m-1)(n-1)\right]$$

Finding the $(1-\alpha)100\%$ confidence interval for the ratio $\frac{S_x^2}{S_y^2}$ and recalling the fact that

$$F_{1-\alpha/2}(m-1)(n-1) = \frac{1}{F_{\alpha/2}(n-1)(m-1)}$$

$$\frac{1}{F_{\alpha/2}(n-1, m-1)} \cdot \frac{S_x^2}{S_y^2} \leq \frac{\sigma_x^2}{\sigma_y^2} \leq F_{\alpha/2}(m-1, n-1)\frac{S_x^2}{S_y^2}$$

∴ Hence the proof.]

Confidence intervals for mean; known variance:

Suppose $X_1, X_2 \ldots X_n$ are independent and identically distributed (i.i.d) random variables and we want to make inference about mean $\mu$ of the population. i.e) $\mu = E(X_i)$

Since $\mu$ determines the population distribution it is called a parameter.

A point estimator such that the sample mean $\bar{X}$, provides a single guess for the true value of parameter $\mu$.

An interval estimator consists of a range of values designed to contain $\mu$ with prespecified probability.

The interval estimator automatically provides a margin of error to account for the sampling variability of $\bar{X}$.

An interval with random end points which, contains the parameter of interest with a pre-specified probability denoted by $1-\alpha$ (the confidence level).

$\alpha = 0.05$ or $\alpha = 0.01$.

$\sigma^2 = Var[x_i]$, here we assume that $\sigma^2$ is known in practical situations we will real know the value of $\sigma^2$, but this assumption is convenient for now.

There are two assumptions.

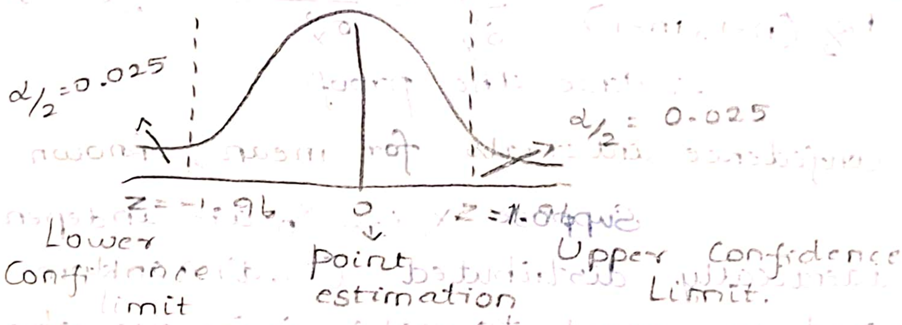
Population Variance ($\sigma^2$) is known.

Population is normally distributed.

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

where $Z_{\alpha/2}$ is the normal distribution value for a probability of $\alpha/2$ in each tail. This value is the Z score with $\alpha/2$ probability upper tail. $\sigma/\sqrt{n}$ is standard error.

$\bar{x}$ is sample mean




Constructing a confidence Interval: for the varia

* We know that if $x_1, x_2 \dots x_n$ is a random taken a normal population with mean $\mu$ and variance $\sigma^2$ and if the sample variance is denoted by $s^2$ the random variable.

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

has the chi-square distribution with $n-1$ degrees of freedom. This knowledge enable us to Construct a confidence interval as follo

Firstly we decide distribution with $n-1$ d.f and 95% of confidence level.

* Then we have n-1 df so that the value of 95% $\chi^2$ will lie between the left tail value of $\chi^2_{0.975, n-1}$ and the right tail value of $\chi^2_{0.025, n-1}$ if we known. The confidence interval is developed is shown below,

we have

$$\chi^2_{0.025,n-1} \leq \chi^2 \leq \chi^2_{0.975,n-1}$$

So that,

$$\chi^2_{0.025,n-1} \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi^2_{0.975,n-1}$$

Hence,

$$\frac{1}{\chi_{0.975,n-1}} \leq \frac{\sigma^2}{(n-1)s^2} \leq \frac{1}{\chi^2_{0.025,n-1}}$$

So that

$$\left[\frac{(n-1)s^2}{\chi^2_{0.975,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{0.025,n-1}}\right] = 0.95$$

Another, we say the using probability directly in say that,

$$P\left(\frac{(n-1)s^2}{\chi^2_{0.975,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{0.025,n-1}}\right)$$

Nothing that 0.95 = 100 (1-0.5) and the working with right-hand tail value of $\chi^2$ distribution it is used to generalize that above result as follows:

Taking the confidence level as 100(1-α)%. (a 95% interval gives α=0.05) our confidence interval becomes

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}$$

The confidence interval for the standard deviation σ is obtained by taking the appropriate square roots.

The following key summaries the development of this confidence interval,

A one sided 100(1-α)% Upper confidence limit,

$$\bar{X} + Z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$$

One sided 100(1-α)% lower confidence limit, $\bar{X} - Z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$

Upper Limit

$$\bar{X} + \frac{t_{1-\alpha, n-1}\, \hat{\sigma}}{\sqrt{n}}$$

Lower Limit

$$\bar{X} - \frac{t_{1-\alpha, n-1}\, \hat{\sigma}}{\sqrt{n}}$$

Standard Deviation is known that

$$D = \frac{t_{1-\alpha/2, n-1}\, \hat{\sigma}}{\sqrt{n}}$$

Confidence interval for mean, unknown variance:

Let us $x_1, x_2 \dots x_n$ are iid with unknown mean $\mu$ and unknown variance $\sigma^2$. clearly we will now have to estimate $\sigma^2$ from the available data. The most commonly used estimator of $\sigma^2$ is the sample variance

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

The reason for using n-1 in the denominator that is makes $s^2$ an unbiased estimator of $\sigma^2$ other word $E(s_X^2) = \sigma^2$. The interval $\bar{X} \pm Z_{\alpha/2} \frac{S_X}{\sqrt{n}}$ is an asymptotic level 1-α C.I for $\mu$.

In other words the sample size is large we can use $S_X$ in place of the unknown $\sigma$ and the confidence intervals.

It can be shown that $S_X^2$ Converge in Probability to $\sigma^2$.

$$\lim_{n \to \infty} P_r(|S_X^2 - \sigma^2| > \epsilon) \to 0 \text{ for any } \epsilon > 0$$

the distribution,

$$\frac{\bar{X} - \mu}{S_X / \sqrt{n}}$$

we get

$$P_r(\text{C.I contains } \mu) = P_r\left(-z_{\alpha/2} < \frac{\bar{x}-\mu}{S_x/\sqrt{n}} < z_{\alpha/2}\right)$$

The sample size is small we get C.I is $\bar{x} \pm t_{\alpha/2} \frac{S_x}{\sqrt{n}}$ [$t_{\alpha/2}$ is defined below when n is small]

$$t = \frac{\bar{x}-\mu}{S_x/\sqrt{n}}$$

Here doesn't whether we use the C.I

$$\bar{x} - t_{\alpha/2} \frac{S_x}{\sqrt{n}} \text{ or } \bar{x} \pm z_{\alpha/2} \frac{S_x}{\sqrt{n}}$$

⇒ Confidence interval for proportion:

Let z be N(0,1) and P be a number between 0 and 1 Critical value $-z$, $z_p$ are

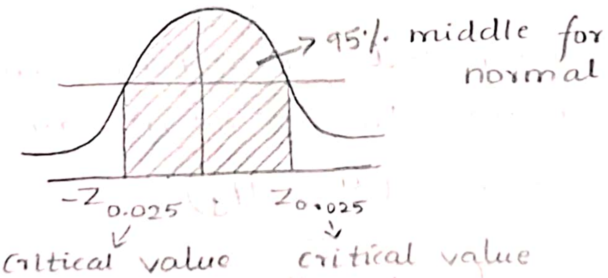
$$P(z > z_p) = 1 - \Phi(z_p) = P$$

Let $0 < \alpha < 1$ and $x$ be a number of success in n observed trials of Bernoulli experiment with unknown probability of success p for $\hat{p} = \frac{x}{n}$ the $100(1-\alpha)\%$ Confidence interval for proportion

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \text{ and } \sqrt{\hat{p}(1-\hat{p}/n)}$$

$$\left[\hat{p} - z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, 1\right] \cdot \left[0, \hat{p} + z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$




and $z_{0.05}$ Critical value for the success to the sample size is $\hat{p} = x/n$ $\therefore \alpha = 5\% = 0.05$

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \text{ and}$$

$$S.E(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Confidence interval for $\mu$ ($\sigma$ unknown)

Assumption:

* Population standard deviation is unknown.
* Population is normally distributed.
* In population is not normal use large sample
* Confidence interval estimate

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

where $S/\sqrt{n} \rightarrow$ standard error

$t_{\alpha/2, n-1} \, S/\sqrt{n} \rightarrow$ margin of error

Confidence intervals for the difference of two normal population means:

Let $X_1, X_2 \ldots X_n$ and $Y_1, Y_2 \ldots Y_n$ be a random sample for two independent normal distribution with means $M_x, M_y$ and standard deviation $\sigma_x$ and $\sigma_y$ respectively.

The sample means and variance for $X$ and $Y$ are also denoted as $\bar{X}, \bar{Y}, S_x^2$ and $S_y^2$ respectively.

We are interested in $100(1-\alpha)\%$ C.I for $\mu = \mu_x = \mu_y$ when we known the ratio of variance say,

$$\frac{\sigma_y^2}{\sigma_x^2} = c \quad \text{where } c \geq 1$$

Confidence interval for $\mu$ with a known ratio of variance:

The proposed confidence interval is constructed using the pivotal quantity

$$T_3 = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\bar{S}_p \sqrt{1/n + c/m}}$$

$$\bar{S}_p^2 = \frac{(n-1) S_{\bar{x}}^2 + (m-1) S_{\bar{y}}^2}{n+m-2}$$

$$S_{\bar{y}}^2 = (m-1)^{-1} \sum_{i=1}^{m} (\bar{y}_i - \bar{y}^*)^2$$

$$\bar{y}_i = Y_i/\sqrt{c} \; ; \; i = 1, 2 \ldots m), \; c \geq 1$$

and

$\bar{y}^*$ is the sample mean of $\bar{Y}_i$ $(i=1,2\ldots m)$

we choose the $t_{1-\alpha/2}, n+m-2$ which is the $(1-\alpha/2)^*$ percentive of the t-distribution with $n+m-2$ d.f such that

$$1-\alpha = P_r\left[-t_{1-\alpha/2, n+m-2} < T_3 < t_{1-\alpha/2, n+m-2}\right]$$

It is easy to see that $100(1-\alpha)\%$ C.I for $M$ is

$$CI_3 = \left[(\bar{x}-\bar{y}) - t_{1-\alpha/2, n+m-2}\, \bar{S}_p\sqrt{1/n + c/m}\right.$$
$$\left.\bar{x}-\bar{y} + t_{1-\alpha/2, n+m-2}\, \bar{S}_p\sqrt{1/n + c/m}\right]$$

Pivotal Quantity method:

Definition:

Let X be a random variable with probability $P_\theta$, $\theta \in \Theta$ where $\theta \in R$. Let $(X_1, X_2 \ldots X_n)$ be a random variable of X

Let $T_1 = T_1(X_1, X_2 \ldots X_n)$ and

$T_2 = T_2(X_1, X_2 \ldots X_n)$ be two dimensional statistics

$$P_\theta(T_1 \le \theta \le T_2) \geq 1-\alpha \quad \forall\, \theta \in \Theta$$

$$T_1(X_1, X_2 \ldots X_n)\ T_2(X_1, X_2 \ldots X_n)\ \text{C.I}\ 1-\alpha$$

~~Confi.~~ Pivotal Method to find Confidence Interval:

Let $x_1, x_2 \ldots x_n$ be a random sample of n observations selected from a population having p.d.f $f(x, \theta)$. Let $Q = q(x_1, x_2 \ldots x_n; \theta)$ be the function of sample observations and parameter. Now, if the distribution of Q is free of $\theta$, Q is called "pivotal" Quantity. For ex, if the sample observation are selected from $N(\mu, \sigma^2)$ then

$$Q = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ if } \sigma^2 \text{ is known. Also}$$

$$Q = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ if } n \text{ is large, when } s^2 \text{ is an}$$

unbiased estimator of $\sigma^2$. Here the distribution of Q is free of parameter $\mu$. $\therefore$ both the Q is Pivotal quantity.

Example:

Let $x_1, x_2, \dots x_n$ be a random sample of $n$ observations from $N(\mu, \sigma^2)$. Find $100(1-\alpha)\%$ C.I for $\mu$ and $\sigma^2$.

Soln:

C.I for $\mu$:

Let us assume that $\sigma^2$ is known. Then

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

is a pivotal quantity. Therefore, the $100(1-\alpha)\%$ C.I for $\mu$ is

$$\bar{x} - q_1 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + q_2 \sigma/\sqrt{n}$$

Here $q_1$ & $q_2$ are to be replaced by $z_1 = z_{\alpha/2}$ and $z_2 = \frac{z_1 - a}{2}$. Therefore,

$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

is the req. C.I for $\mu$. Here, if $\alpha = 0.05$, then for 95% C.I $z_{\alpha/2} = 1.96$ and $z_{\alpha/2} = z_1 - \frac{a}{2}$.

When $\sigma^2$ is not known: If $\sigma^2$ is not known it is replaced by

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

In such a case the pivotal quantity $Q = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ follows students 't' distribution with $(n-1)$ d.f. Therefore, the $100(1-\alpha)\%$ C.I for $\mu$ is

$$\bar{x} - q_1 \frac{s}{\sqrt{n}} < \mu < \bar{x} + q_2 \frac{s}{\sqrt{n}}$$

where $q_1$ & $q_2$ are to be found out in such a way that

$$\int_{q_1}^{q_2} f(t)\, dt = 1-\alpha.$$

C.I for $\sigma^2$:

The estimate of $\sigma^2$ is

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum^{n} (x_i - \bar{x})^2$$

Therefore, the pivotal quantity is

$$Q = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}.$$

The distribution of $Q$ is free of $\sigma^2$. Hence the necessary $100(1-\alpha)\%$ C.I is obtained from the eqn,

$$P[q_1 < Q < q_2] = 1-\alpha$$

Here

$$q_1 \leq \frac{(n-1)s^2}{\sigma^2} \leq q_2 \quad \text{or} \quad \frac{(n-1)s^2}{q_2} \leq \sigma^2 \leq \frac{(n-1)s^2}{q_1}$$

Since $Q$ is distributed as chi-square, we can write

$$P[Q \geq \chi^2_\alpha] = \int_{\chi^2_\alpha}^{\infty} f(Q)\, dQ = \alpha$$

Then,

$$P\left[\chi^2_{1-\alpha/2} < Q < \chi^2_{\alpha/2}\right] = 1-\alpha$$

$$P\left[\chi^2_{1-\alpha/2} < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{\alpha/2}\right] = 1-\alpha$$

$$P\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right] = 1-\alpha$$

Hence $\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right]$ is the $100(1-\alpha)\%$ C.I of $\sigma^2$. Hence $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ are the tabulated value of $\chi^2$ with $(n-1)$ d.f.

Confidence Interval using large Sample:

Let $x_1, x_2 \ldots x_n$ be a random sample of $n$ observations from a population having p.d.f $f(x, \theta)$, $\theta \in S$. It is assumed that $n$ is large. The likelihood function of the sample observations is

$$L = f(x_1, \theta)\, f(x_2, \theta) \cdots f(x_n, \theta).$$

It is assumed that the regularity conditions are valid for the distribution and hence the ML estimator of $\theta$ is obtained solving the equation

$$\frac{\partial \log L}{\partial \theta} = 0.$$

Again, it is also known that the ML estimator is asymptotically distributed as normal. and

$$E\left[\frac{\partial \log L}{\partial \theta}\right] = 0, \quad V\left[\frac{\partial \log L}{\partial \theta}\right] = -E\left[\frac{\partial^2 \log L}{\partial \theta^2}\right]$$

Therefore, if the sample size is large,

$$Z = \frac{\partial \log L}{\sqrt{V\left(\frac{\partial \log L}{\partial \theta}\right)}} \sim N(0,1)$$

Therefore, the $100(1-\alpha)\%$ C.I for $\theta$ is obtained from the following eqn:

$$P\left[|z| \leq Z_{\alpha/2}\right] = 1-\alpha \text{ or,}$$

$$\frac{1}{\sqrt{2\pi}} \int_{-Z_{\alpha/2}}^{Z_{\alpha/2}} e^{-\frac{1}{2}z^2} dz = 1-\alpha$$

Here

$$Q = \frac{\partial \log L}{\partial \theta} \text{ and } P\left[q_1 < Q < q_2\right] = 1-\alpha$$

The values of $q_1$ & $q_2$ are obtained from Normal Probability Table.

<u>Example:</u>

Let $x_1, x_2 \ldots x_n$ be a random Sample of n observations from a population with p.d.f

$$f(x) = \theta x^{\theta-1}, \quad 0 < x < 1$$

Find $100(1-\alpha)\%$ C.I for $\theta$.

<u>Soln:</u>

We have,

$$F(x, \theta) = x^{\theta}, \quad 0 < x < 1$$

$$\therefore P\left[q_1 < \prod_{i=1}^{n} F(x_i, \theta) < q_2\right] = 1-\alpha$$

or $P\left[q_1 < \prod_{i=1}^{n} x_i^{\theta} < q_2\right] = 1-\alpha$

or $P\left[\log q_1 < \theta \prod_{i=1}^{n} \log x_i < \log q_2\right] = 1-\alpha$

or $P\left[-\log q_2 < -\theta \prod_{i=1}^{n} \log x_i < -\log q_1\right] = 1-\alpha$

or, $P\left[\dfrac{\log q_2}{\log \prod_{i=1}^{n} x_i} < \theta < \dfrac{\log q_1}{\log \prod_{i=1}^{n} x_i}\right] = 1-\alpha$

Hence $100(1-\alpha)\%$ C.I for $\theta$ is

$$\left[\frac{\log q_2}{\log \prod_{i=1}^{n} x_i}, \frac{\log q_1}{\log \prod_{i=1}^{n} x_i}\right].$$