Interval Estimation

Let $x_1, x_2, \ldots, x_n$ be a random sample from a popn with distn $f(x, \underline{\theta})$, $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k) \in \Omega$.

We want to estimate a function $\eta = g(\underline{\theta})$. In interval estimation we find two statistics $T_1(\underline{x})$ and $T_2(\underline{x})$ such that interval $(T_1(\underline{x}), T_2(\underline{x}))$ is likely to include $g(\underline{\theta})$.

Confidence interval:- We say that $(T_1(\underline{x}), T_2(\underline{x}))$ is a $100(1-\alpha)\%$ confidence interval for $g(\underline{\theta})$ if

$$P(T_1(\underline{x}) < g(\theta) < T_2(\underline{x})) = 1-\alpha \;\forall\; \underline{\theta} \in \Omega$$

Confidence interval for parameters of a Normal Distributions:-

Let $x_1, x_2, \ldots, x_n \sim N(\mu, \sigma^2)$.

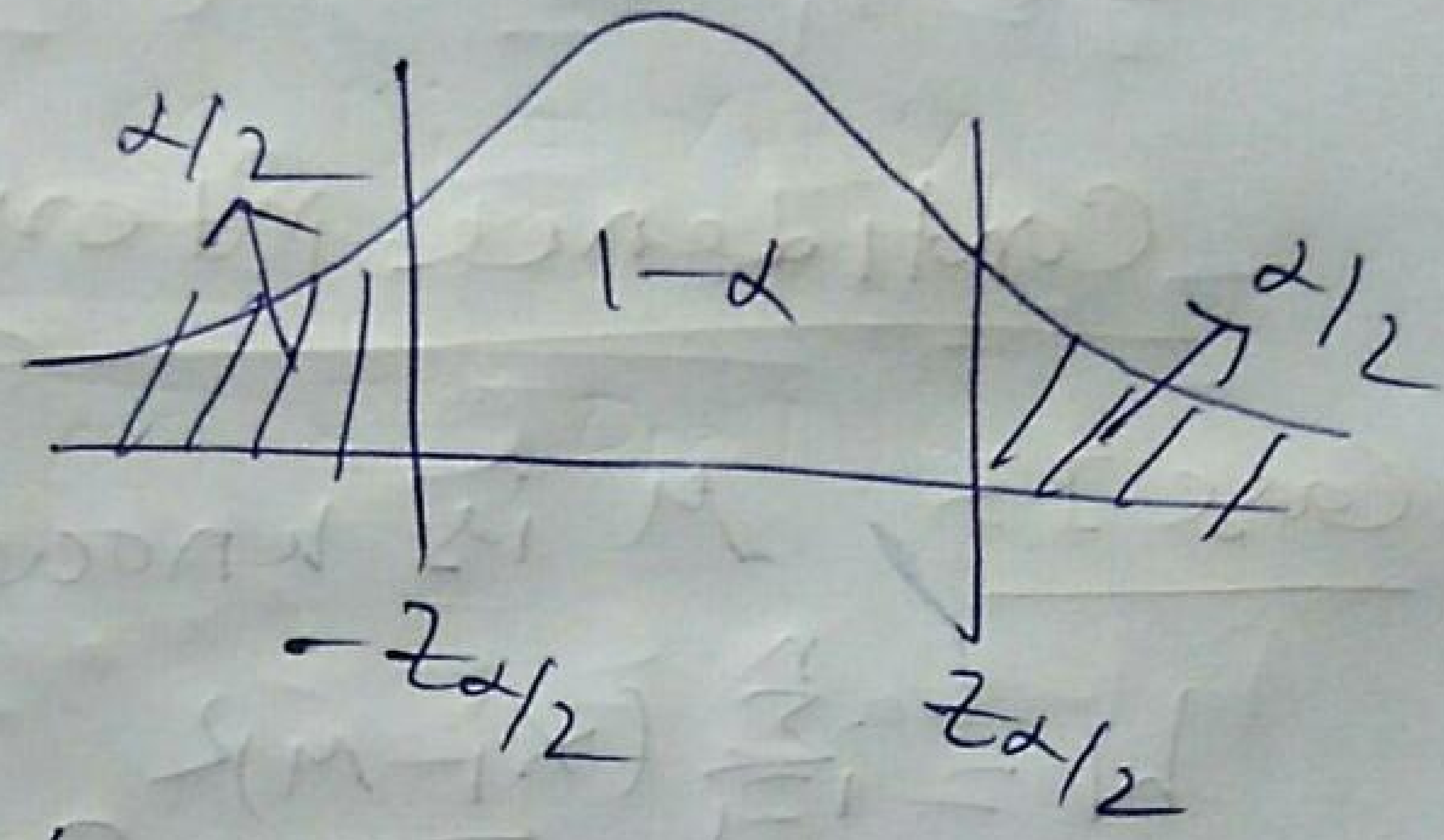
<u>Case I</u>:- σ^2 is known, we want confidence interval for μ.

~~M~~ethod of Pivoting :-

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

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

$P(-z_{\alpha/2} \le Z \le z_{\alpha/2})$
$= 1-\alpha$



$\Rightarrow P\left(-z_{\alpha/2} \le \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \le z_{\alpha/2}\right) = 1-\alpha$.

$\Rightarrow P\left(-\frac{\sigma}{\sqrt{n}} z_{\alpha/2} \le \bar{X}-\mu \le \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right) = 1-\alpha$

$\Rightarrow P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \le \mu \le \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right) = 1-\alpha$

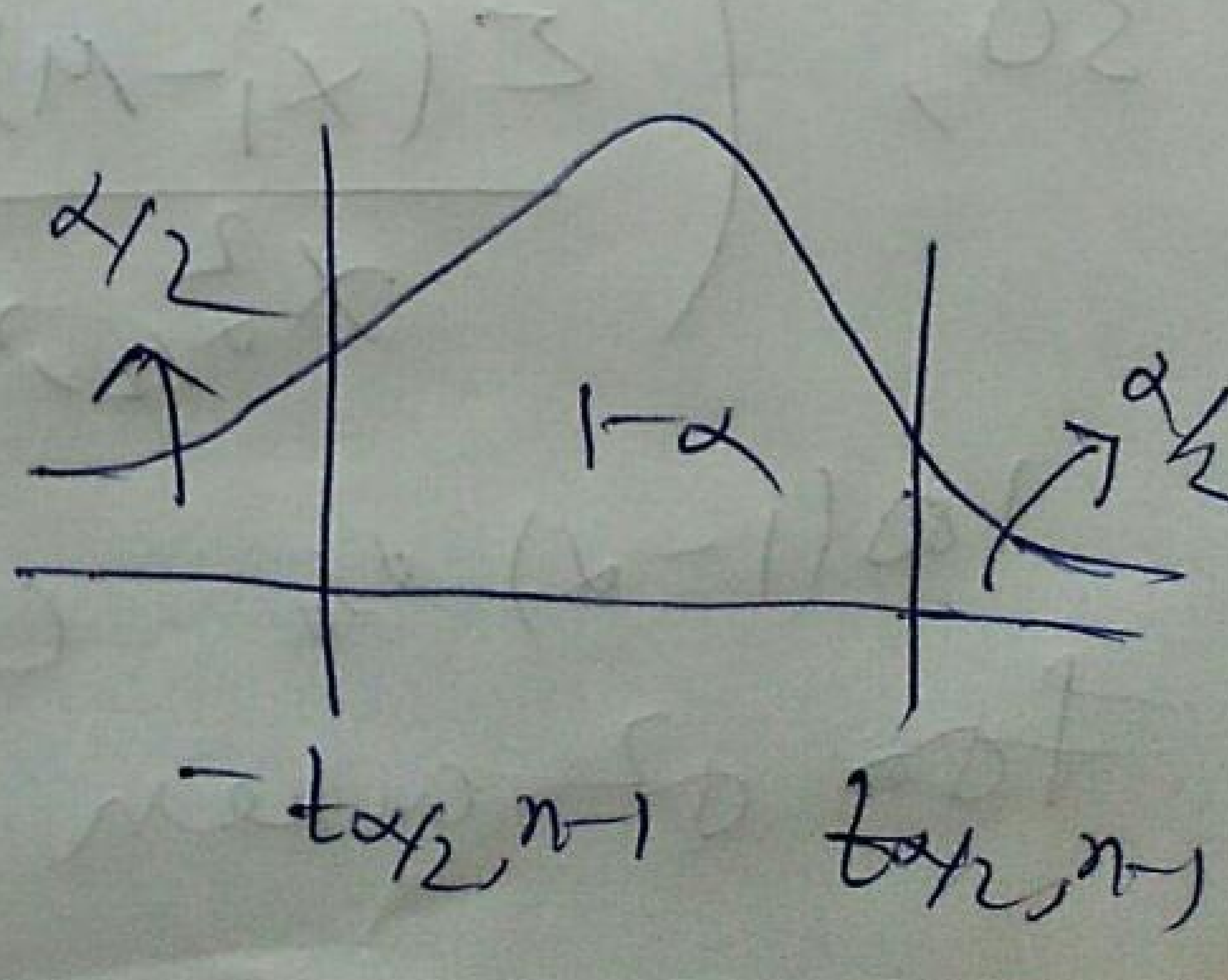
So, $\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)$ is

$100(1-\alpha)\%$ confidence interval for μ.

Case II :- σ^2 is unknown.

$\frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{n-1}$

So, $P\left(-t_{\alpha/2, n-1} \le \frac{\sqrt{n}(\bar{X}-\mu)}{S} \le t_{\alpha/2, n-1}\right) = 1-\alpha$



Scanned by TapScanner

$$\Rightarrow P\left(\bar{x} - \frac{s}{\sqrt{n}}\, t_{\alpha/2,\,n-1} \le \mu \le \bar{x} + \frac{s}{\sqrt{n}}\, t_{\alpha/2,\,n-1}\right) = 1-\alpha.$$

So, $\left(\bar{x} - \frac{s}{\sqrt{n}}\, t_{\alpha/2,\,n-1},\ \bar{x} + \frac{s}{\sqrt{n}}\, t_{\alpha/2,\,n-1}\right)$

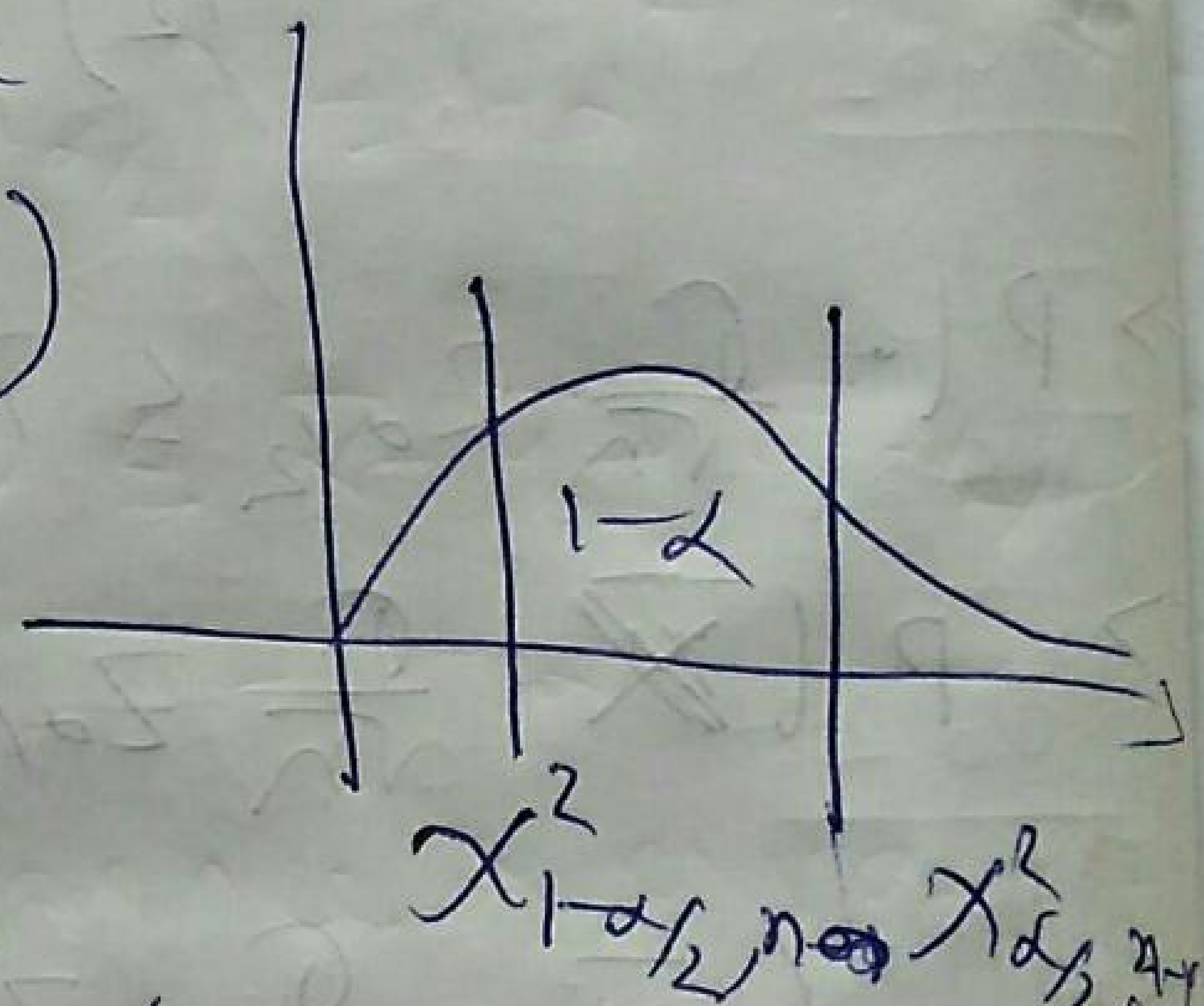
is $100(1-\alpha)\%$ confidence interval for μ.

Confidence interval for σ^2

Case I: μ is known

$$W = \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{\sigma^2} \sim \chi^2_n$$

$$P\left(\chi^2_{1-\alpha/2,\,n} \le W \le \chi^2_{\alpha/2,\,n}\right) = 1-\alpha.$$



$$\Rightarrow P\left(\chi^2_{1-\alpha/2,\,n} \le \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{\sigma^2} \le \chi^2_{\alpha/2,\,n}\right) = 1-\alpha$$

$$\Rightarrow P\left(\frac{\sum (x_i-\mu)^2}{\chi^2_{\alpha/2,\,n}} \le \sigma^2 \le \frac{\sum (x_i-\mu)^2}{\chi^2_{1-\alpha/2,\,n}}\right) = 1-\alpha$$

So, $\left(\frac{\sum (x_i-\mu)^2}{\chi^2_{\alpha/2,\,n}},\ \frac{\sum (x_i-\mu)^2}{\chi^2_{1-\alpha/2,\,n}}\right)$ is

$100(1-\alpha)\%$ confidence interval for σ^2 when μ is known.

Case II :- When μ is unknown.

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

$$P\left(\chi^2_{1-\alpha/2, n-1} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\alpha/2, n-1} \right) = 1-\alpha$$

$$\Rightarrow P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}} \right) = 1-\alpha$$

So, $\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}} \right)$ is $100(1-\alpha)\%$ confidence interval for σ^2.

Confidence intervals for parameters of Two normal Populations :—

Let $X_1, X_2, \ldots, X_m$ be a random sample from $N(\mu_1, \sigma_1^2)$

$Y_1, \ldots, Y_n$ be another independent sample from $N(\mu_2, \sigma_2^2)$ pop^n.

We want find confidence interval for $\theta = \mu_1 - \mu_2$.

Case I :- σ_1^2 & σ_2^2 are known

$$\bar{X} \sim N(\mu_1, \sigma_1^2/m)$$

independent

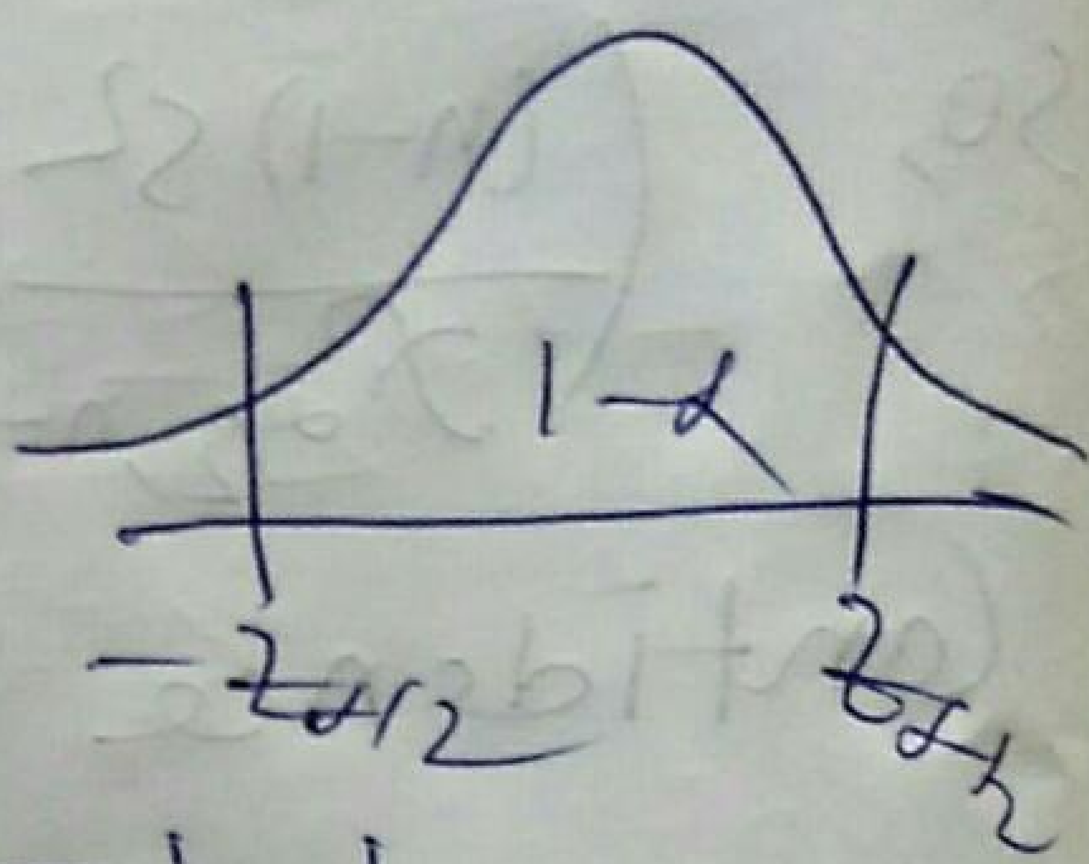
$$\bar{Y} \sim N(\mu_2, \sigma_2^2/n)$$

$$\Rightarrow \bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right)$$

so, $$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0, 1)$$

Let $\theta = \mu_1 - \mu_2$, $\tau^2 = \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}$

$$Z = \frac{\bar{X} - \bar{Y} - \theta}{\tau} \sim N(0, 1)$$



$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$

$$\Rightarrow P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \bar{Y} - \theta}{\tau} \leq z_{\alpha/2}\right) = 1 - \alpha$$

$$\Rightarrow P(\bar{X} - \bar{Y} - \tau z_{\alpha/2} \leq \theta \leq \bar{X} - \bar{Y} + \tau z_{\alpha/2}) = 1 - \alpha$$

so, $100(1-\alpha)\%$ confidence interval for $\theta = \mu_1 - \mu_2$ in this case is

$$\left(\bar{X} - \bar{Y} - \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\, z_{\alpha/2}, \bar{X} - \bar{Y} + \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\, z_{\alpha/2}\right)$$

Case II :- σ_1^2 and σ_2^2 are unknown but equal. say $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

$$U = \frac{\bar{x} - \bar{y} - \theta}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim N(0,1). \qquad c^2 = \frac{1}{m} + \frac{1}{n}.$$

$$\Rightarrow U = \frac{\bar{x} - \bar{y} - \theta}{c\sigma} \sim N(0,1).$$

$$W_1 = \frac{(m-1)S_x^2}{\sigma^2} \sim \chi^2_{m-1}, \qquad S_x^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})^2$$

independent

$$W_2 = \frac{(n-1)S_y^2}{\sigma^2} \sim \chi^2_{n-1}, \qquad S_y^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_j - \bar{y})^2$$

$$W_1 + W_2 = \frac{(m-1)S_x^2 + (n-1)S_y^2}{\sigma^2} \sim \chi^2_{m+n-2}$$

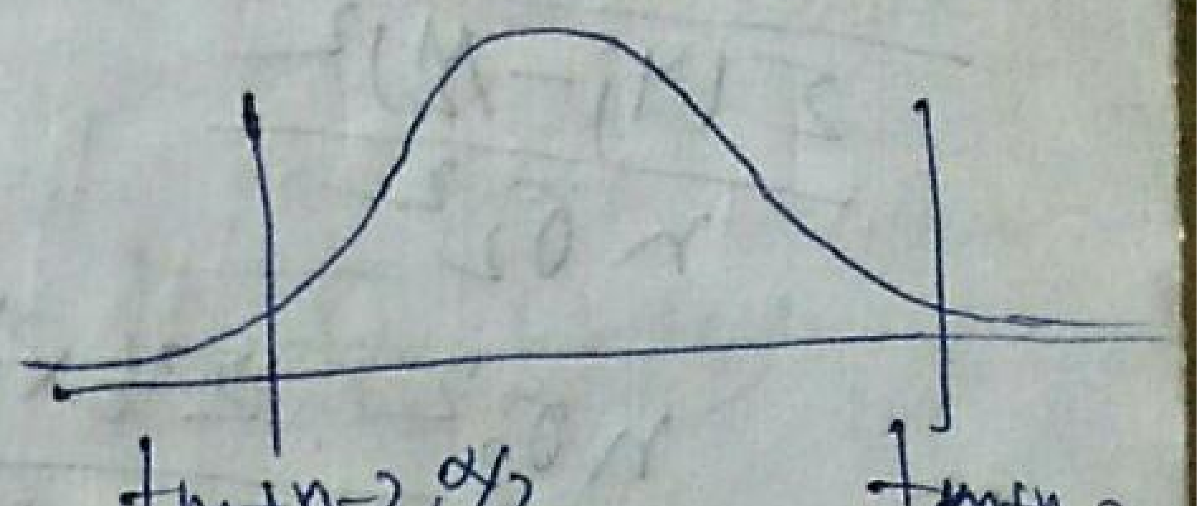
$$S_p^2 = \frac{1}{(m+n-2)}\left\{(m-1)S_x^2 + (n-1)S_y^2\right\} \rightarrow \text{Pooled sample variance}$$

$$V = \frac{(m+n-2)S_p^2}{\sigma^2} \sim \chi^2_{m+n-2}.$$

Since sample mean & sample variance are independent so, U & V are independent.

$$\text{so, } \frac{U}{\sqrt{\frac{V}{m+n-2}}} \sim t_{m+n-2}.$$

$$\Rightarrow \frac{\frac{\bar{X}-\bar{Y}-\theta}{c\sigma}}{\sqrt{\frac{\cancel{(m+n-2)}S_p^2}{\sigma^2\cancel{(m+n-2)}}}} = \frac{\bar{X}-\bar{Y}-\theta}{cS_p} \sim t_{m+n-2}.$$



$$P\left(-t_{\alpha/2,m+n-2} \le \frac{\bar{X}-\bar{Y}-\theta}{cS_p} \le t_{\alpha/2,m+n-2}\right) = 1-\alpha.$$

$$\Rightarrow P\left(\bar{X}-\bar{Y}-cS_p t_{\alpha/2,m+n-2} \le \theta \le \bar{X}-\bar{Y}+cS_p t_{\alpha/2,m+n-2}\right) = 1-\alpha.$$

So, $\bar{x}-\bar{y} \pm cS_p t_{\alpha/2,m+n-2}$ gives $100(1-\alpha)\%$ confidence interval for $\theta = \mu_1-\mu_2$ when $\sigma_1^2=\sigma_2^2=\sigma^2$ (unknown).

Case III :- When σ_1^2 and σ_2^2 are unknown & unequal.

Welch / Sathemitle formula

$$T = \frac{\bar{X}-\bar{Y}-\theta}{\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}}}$$ has an approximate t_ν dist^n.

where ν can be taken as integer part of

$$\frac{\left(\frac{S_1^2}{m}+\frac{S_2^2}{n}\right)^2}{\left[\frac{S_1^4}{m^2(m-1)}+\frac{S_2^4}{n^2(n-1)}\right]}$$

Theory of two stage / multistage

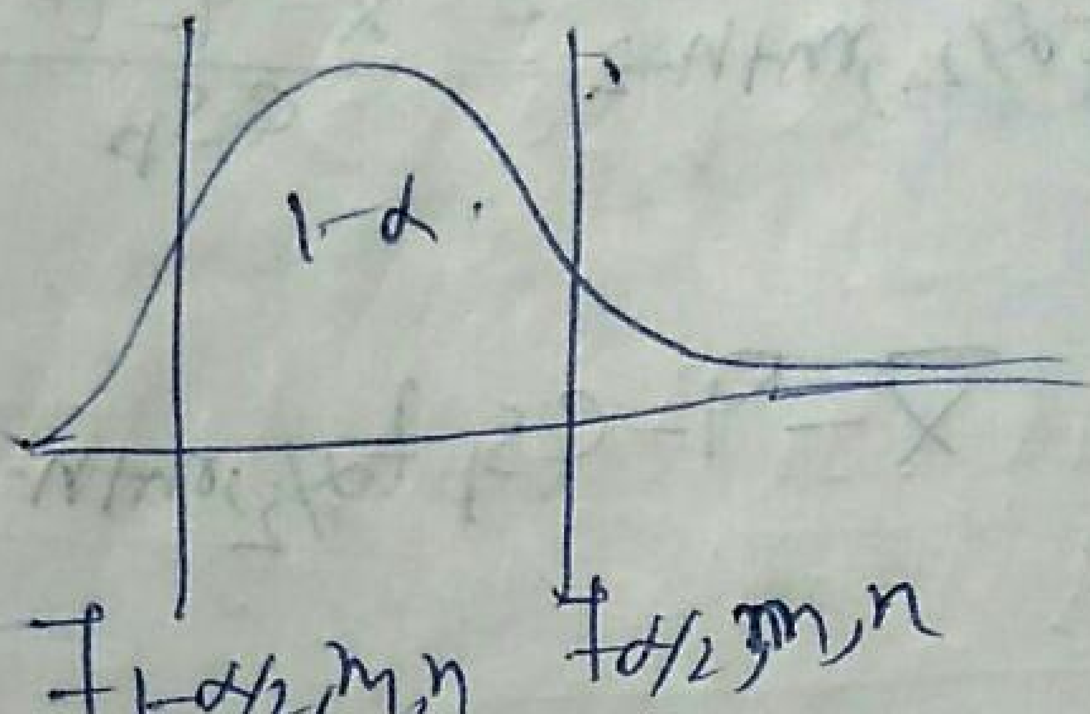
Confidence interval for $\xi = \frac{\sigma_1^2}{\sigma_2^2}$

Case I :- μ_1 & μ_2 are known

$$\underbrace{\frac{\sum(x_i-\mu_1)^2}{\sigma^2} \sim \chi_m^2, \quad \sum_j \frac{(y_j-\mu_2)^2}{\sigma_2^2} \sim \chi_n^2.}_{\text{independent}}$$

$$\frac{\dfrac{\Sigma(x_i-\mu_1)^2}{m\sigma_1^2}}{\dfrac{\Sigma(y_j-\mu_2)^2}{n\sigma_2^2}} \sim F_{m,n}$$

$$\Rightarrow \frac{n\sigma_2^2}{m\sigma_1^2}\,\frac{\Sigma(x_i-\mu_1)^2}{\Sigma(y_j-\mu_2)^2} \sim F_{m,n} \quad (\because \text{very skewed})$$



$$P\left(f_{1-\alpha/2,m,n} \le \frac{n\sigma_2^2}{m\sigma_1^2}\,\frac{\Sigma(x_i-\mu_1)^2}{\Sigma(y_j-\mu_2)^2} \le f_{\alpha/2,m,n}\right) = 1-\alpha$$

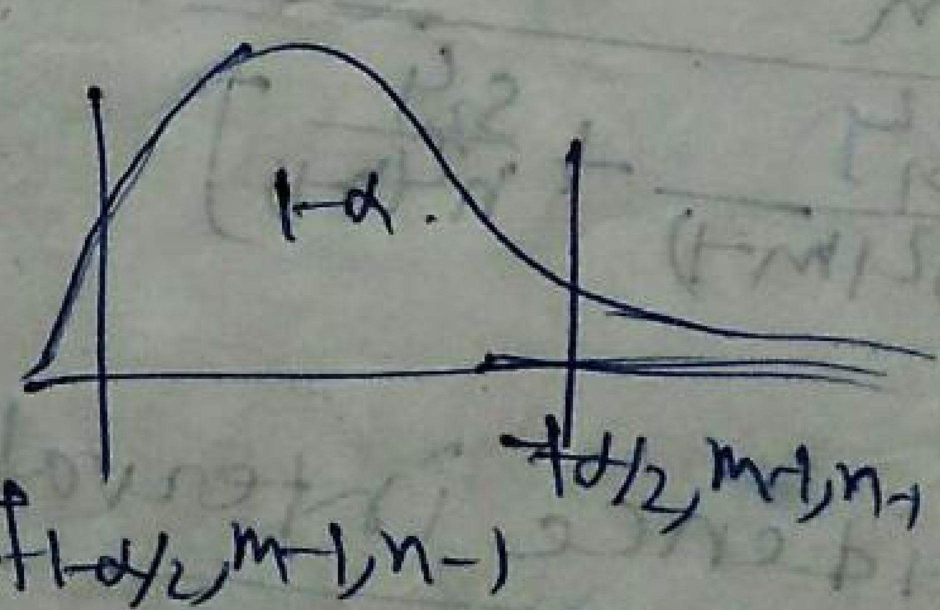
$$\Rightarrow P\left(\frac{m\,\Sigma(y_j-\mu_2)^2}{n\,\Sigma(x_i-\mu_1)^2} f_{1-\alpha/2,m,n} \le \frac{\sigma_2^2}{\sigma_1^2} \le \frac{m\,\Sigma(y_j-\mu_2)^2}{n\,\Sigma(x_i-\mu_1)^2} f_{\alpha/2,m,n}\right) = 1-\alpha$$

$$f_{1-\alpha/2,m,n} = \frac{1}{f_{\alpha/2,n,m}}$$

<u>case II</u> :- μ_1 and μ_2 are unknown

$$\frac{\dfrac{\cancel{(m-1)}\,s_1^2}{\sigma_1^2\,\cancel{(m-1)}}}{\dfrac{\cancel{(n-1)}\,s_2^2}{\sigma_2^2\,\cancel{(n-1)}}} \sim F_{m-1,n-1}$$

$$\Rightarrow \frac{\sigma_2^2}{\sigma_1^2}\,\frac{s_1^2}{s_2^2} \sim F_{m-1,n-1}$$



$$P\left(f_{1-\alpha/2,m-1,n-1} \le \frac{\sigma_2^2}{\sigma_1^2}\,\frac{s_1^2}{s_2^2} \le f_{\alpha/2,m-1,n-1}\right) = 1-\alpha$$

$$\Rightarrow P\left(\frac{s_2^2}{s_1^2} f_{1-\alpha/2,m-1,n-1} \le \frac{\sigma_2^2}{\sigma_1^2} \le \frac{s_2^2}{s_1^2} f_{\alpha/2,m-1,n-1}\right) = 1-\alpha$$

Scanned by TapScanner

Example :- $m = 36$, $n = 64$

$\bar{x} = 10$, $\bar{y} = 8$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$.

$\alpha = 0.05$, $z_{0.025} = 1.96$

$$\left(\bar{x} - \bar{y} \pm \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\, z_{\alpha/2}\right) = 2 \pm \sqrt{\frac{1}{36} + \frac{1}{64}}\,(1.96)$$

$$= 2 \pm \frac{10}{6 \times 8} \times 1.96$$

$$= 2 \pm \frac{5}{24} \times 1.96$$

$$= (1.592, 2.408)$$

95% confidence interval for $\mu_1 - \mu_2$

Ex:- To compare age at marriage of women in two ethnic groups, a random sample of 100 women is taken from each group.

$\bar{x} = 18.5$, $\bar{y} = 20.7$

$S_p^2 = ?$ $\quad s_1 = 5.8$, $s_2 = 6.3$

$$\bar{x} - \bar{y} \pm \sqrt{\frac{m+n}{mn}}\, S_p \cdot t_{0.05, 198}$$

$$\left(18.5 - 20.7 \pm \frac{\sqrt{2}}{10} \times 1.645\right)$$

Ex:- Two machines are used to fill plastic bottles with dishwashing detergent. The s.d.s of fill volume are known to be $\sigma_1 = 0.15$ fluid and $\sigma_2 = 0.12$ fluid for the two machines. Two random samples of $n_1 = 12$ bottles from machine 1 and $n_2 = 10$ bottles from machine 2 are selected and the observations are $\bar{x}_1 = 30.87$, $\bar{x}_2 = 30.68$. ~~90~~ 90% C.I. for $\mu_1 - \mu_2$.

$$\bar{x} - \bar{y} \pm \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\, z_{0.05} \rightarrow 30.87 -$$

Paired Observations.

$(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n) \sim$ ~~BVN~~

$BVN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

$d_i = x_i - y_i \sim N(\theta, \tau^2)$, $\tau^2 =$

$\tau^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$.

So, $100(1-\alpha)\%$ confidence interval for $\theta = \mu_1 - \mu_2$ is

$$\left(\bar{d} - \frac{s_d}{\sqrt{n}} t_{\alpha/2, n-1}, \bar{d} + \frac{s_d}{\sqrt{n}} t_{\alpha/2, n-1}\right)$$

where $\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$, $s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$

Ex:- To compare the gripping strength of left hand and right hand of left handed persons the measurements are made on 10 persons

person	1	2	3	4	5	6	7	8	9	10	
left hand	140	90	125	130	95	121	85	97	131	110	x_i
Right hand	138	87	110	132	96	120	86	90	129	100	y_i

<u>Confidence interval for $\mu_1 - \mu_2$</u>.

$d_i \longrightarrow$ 2, 3, 15, −2, −1, 1, −1, 7, 2, 10

$\bar{d} = 3.6$, $S_d^2 = \frac{1}{9}\Sigma d_i^2 - \bar{d}^2$

$t_{0.05, 9} = 1.833$.

$\left(3.6 \pm \frac{S_d}{\sqrt{10}} \times 1.833\right)$