Confidence Intervals for Parameters of Two Normal Populations

Let $X_1, \ldots, X_m$ be a random sample from $N(\mu_1, \sigma_1^2)$ and let $Y_1 \ldots, Y_n$ be another independent random sample from $N(\mu_2, \sigma_2^2)$.

Confidence Interval for $\eta = \mu_1 - \mu_2$.

Case I: σ_1^2 and σ_2^2 are known

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

indep

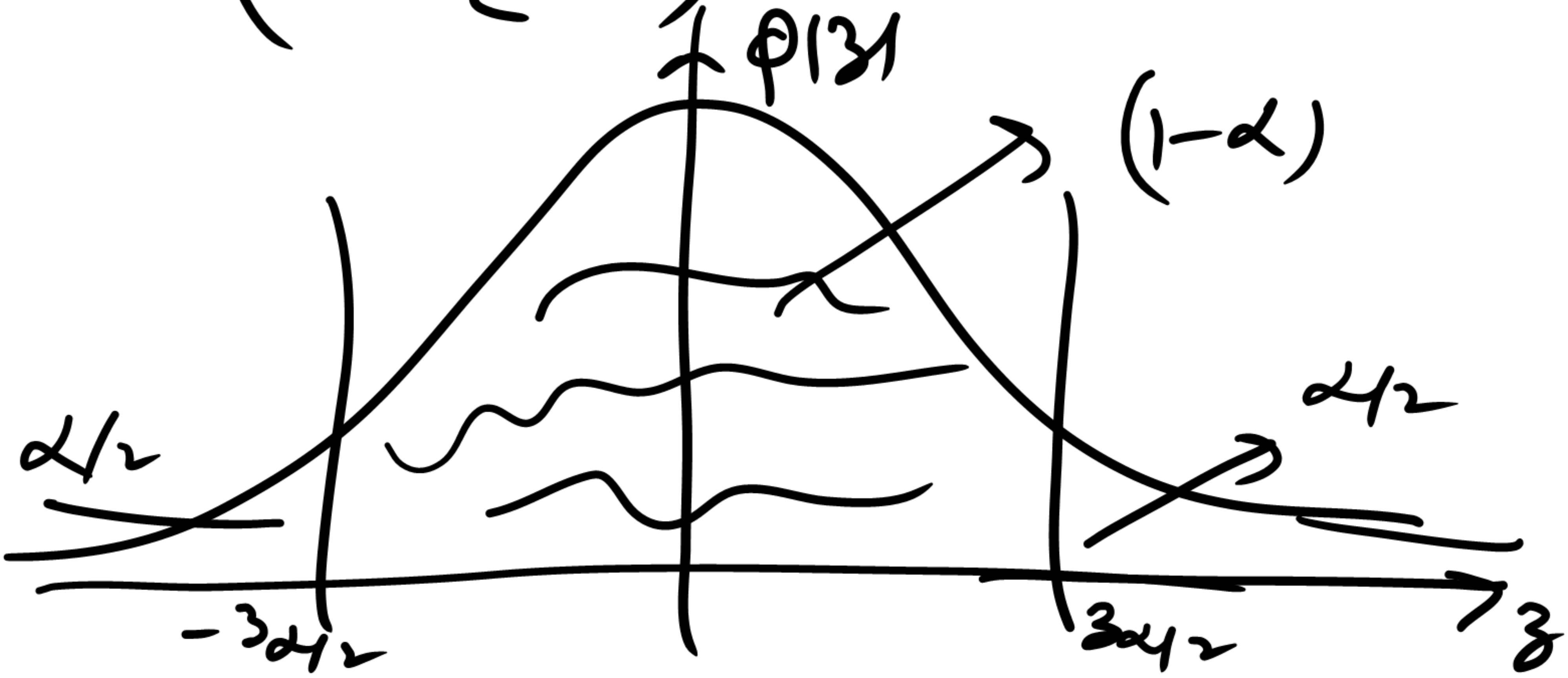
$$\bar{Y} \sim N(\mu_2, \sigma_2^2/n)$$

$$\bar{X} - \bar{Y} \sim N\Big(\underbrace{\mu_1 - \mu_2}_{\eta}, \underbrace{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}_{\tau^2}\Big)$$

$$\bar{x} - \bar{y} \sim N(\eta, \tau^2)$$

$\tau^2 \to$ known

So. $Z = \left(\dfrac{\bar{x} - \bar{y} - \eta}{\tau} \right) \sim N(0,1)$



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

$$\Rightarrow P\left(-z_{\alpha/2} \le \frac{\bar{X}-\bar{Y}-\eta}{\tau} \le z_{\alpha/2}\right)$$

$$= 1-\alpha.$$

$$\Rightarrow P\left(\bar{X}-\bar{Y}-\tau z_{\alpha/2} \le \eta \le \bar{X}-\bar{Y}+\tau z_{\alpha/2}\right)$$

$$= 1-\alpha$$

So $100(1-\alpha)\%$ confidence interval for $\eta = (\mu_1 - \mu_2)$ when σ_1^2 & σ_2^2

are known is

$$\left(\bar{x} - \bar{y} - \sqrt{\left(\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right)} \cdot z_{\alpha/2} \, , \; \bar{x} - \bar{y} + \sqrt{\left(\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right)} z_{\alpha/2} \right)$$

<u>Example</u> : $m = 36$ & $n = 64$

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

95% C.I for $\eta = \mu_1 - \mu_2$ is required.

$$z_{0.025} = 1.96$$

$$\left(\bar{x} - \bar{y} \pm \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\; z_{0.025}\right)$$

$$2 \pm \frac{10}{6 \times 8} \times 1.96 \rightarrow (1.592, 2.408)$$

is 95% C.I. for $\mu_1 - \mu_2$.

We can find confidence intervals

for any linear combination

$$\delta = c_1\mu_1 + c_2\mu_2 .$$

$$c_1\bar{X} + c_2\bar{Y} \sim N\Big(\delta,\ \underbrace{\frac{c_1^2\sigma_1^2}{m} + \frac{c_2^2\sigma_2^2}{n}}_{\xi^2}\Big)$$

$$\frac{c_1\bar{X} + c_2\bar{Y} - \delta}{\xi} \sim N(0,1)$$

So $100(1-\alpha)\%$ C.I for δ is then $\left(c_1\bar{X}+c_2\bar{Y}-\xi\, z_{\alpha/2},\ c_1\bar{X}+c_2\bar{Y}+\xi z_{\alpha/2}\right)$

<u>Case II</u>: $\sigma_1^2=\sigma_2^2=\sigma^2$ (unknown)

That is the variability of two populations is assumed to be unknown but equal.

Here $\bar{X}, \bar{Y}, S_1^2, S_2^2$ are indept.

$$S_1^2=\frac{1}{m-1}\sum(X_i-\bar{X})^2, \quad S_2^2=\frac{1}{n-1}\sum(Y_j-\bar{Y})^2$$

$$\bar{X} \sim N(\mu_1, \sigma^2/m), \quad \bar{Y} \sim N(\mu_2, \sigma^2/n)$$

$$\bar{X} - \bar{Y} \sim N\left(\eta, \ \sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)\right)$$

$$Z = \sqrt{\frac{mn}{m+n}} \frac{(\bar{X} - \bar{Y} - \eta)}{\sigma} \sim N(0,1).$$

$$\frac{(m-1)S_1^2}{\sigma^2} \sim \chi^2_{m-1}, \quad \frac{(n-1)S_2^2}{\sigma^2} \sim \chi^2_{n-1}$$

Since S_1^2 and S_2^2 are independently distributed, we have

$$\frac{(m-1)S_1^2 + (n-1)S_2^2}{\sigma^2} \sim \chi^2_{m+n-2}$$

Let
$$S_p^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{(m+n-2)}$$

$\rightarrow$ Pooled sample variance

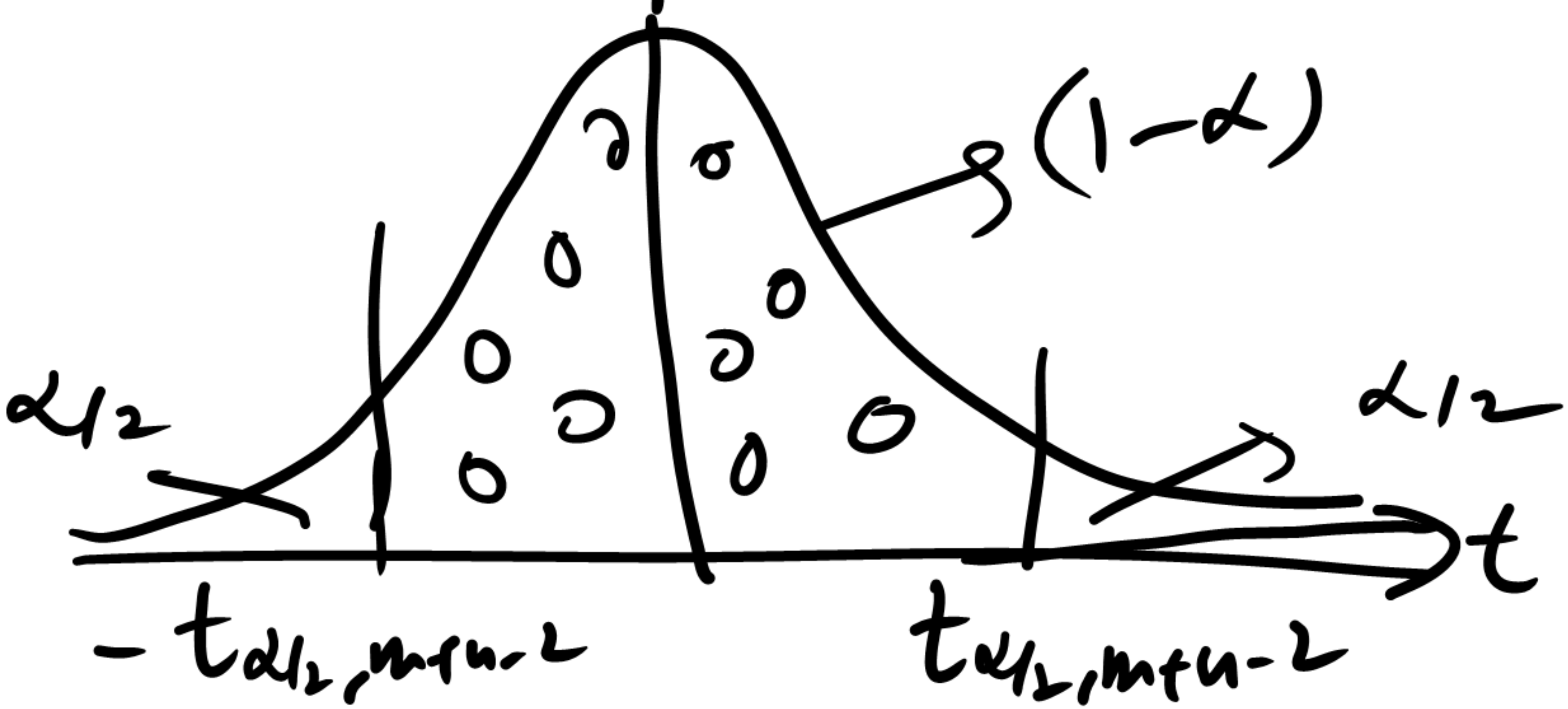
ie $W = \frac{(m+n-2) S_p^2}{\sigma^2} \sim \chi^2_{m+n-2}$

Z and W are independently dist^d.

$$T = \frac{Z}{\sqrt{W/(m+n-2)}} \sim t_{m+n-2}$$

$$= \frac{\sqrt{\left(\frac{mn}{m+n}\right)}\left(\bar{X}-\bar{Y}-\eta\right)/\cancel{\sigma}}{S_p/\cancel{\sigma}}$$

$$= \sqrt{\frac{mn}{m+n}} \frac{(\bar{X} - \bar{Y} - \eta)}{S_p} \sim t_{m+n-2}$$



$$P\left(-t_{\alpha/2, m+n-2} \le \sqrt{\frac{mn}{m+n}} \frac{(\bar{X} - \bar{Y} - \eta)}{S_p} \le\right.$$

$$t_{\alpha/2,\, m+n-2}) = 1-\alpha$$

$$P\left(\bar{X}-\bar{Y} - \sqrt{\frac{m+n}{mn}}\, S_p\, t_{\alpha/2,\, m+n-2} \le \eta \le \bar{X}-\bar{Y} + \sqrt{\frac{m+n}{mn}}\, S_p\, t_{\alpha/2,\, m+n-2}\right)$$

$$= 1-\alpha$$

Thus $100(1-\alpha)\%$. C.I for $\mu_1-\mu_2$ in this case is

$$\bar{X} - \bar{Y} \pm \sqrt{\frac{m+n}{mn}}\, S_p \cdot t_{\alpha/2,\, m+n-2}$$

<u>Case III</u> : σ_1^2 and σ_2^2 are completely unknown.

In this case we do not have exact distribution of the pivot quantity. However it is noted that

$$T^* = \frac{\bar{X} - \bar{Y} - \eta}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}}$$

has

approximately t_ν dist^n where

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

we round off ν to the nearest integer.

Welch
Smith - Satterwaite procedure.

Based on T^*, we can find

$100(1-\alpha)\%$ confidence interval for $\eta = \mu_1 - \mu_2$ as

$$\bar{x} - \bar{y} \pm t_{\alpha/2, \nu} \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}$$

Sometimes we do not have independent random samples. We have situation where the observations $(X_1, Y_1), (X_2, Y_2), \cdots$

(X_n, Y_n) are on the same set of subjects.

We assume here that

$$(X_i, Y_i) \sim BVN(\underbrace{\mu_1, \mu_2}, \sigma_1^2, \sigma_2^2, \underline{\underline{\rho}})$$

We want C.I. for $\eta = \mu_1 - \mu_2$

Define $d_i = X_i - Y_i$

Then $d_i \overset{i.i.d.}{\sim} N(\eta, \sigma_D^2)$

where $\sigma_D^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$

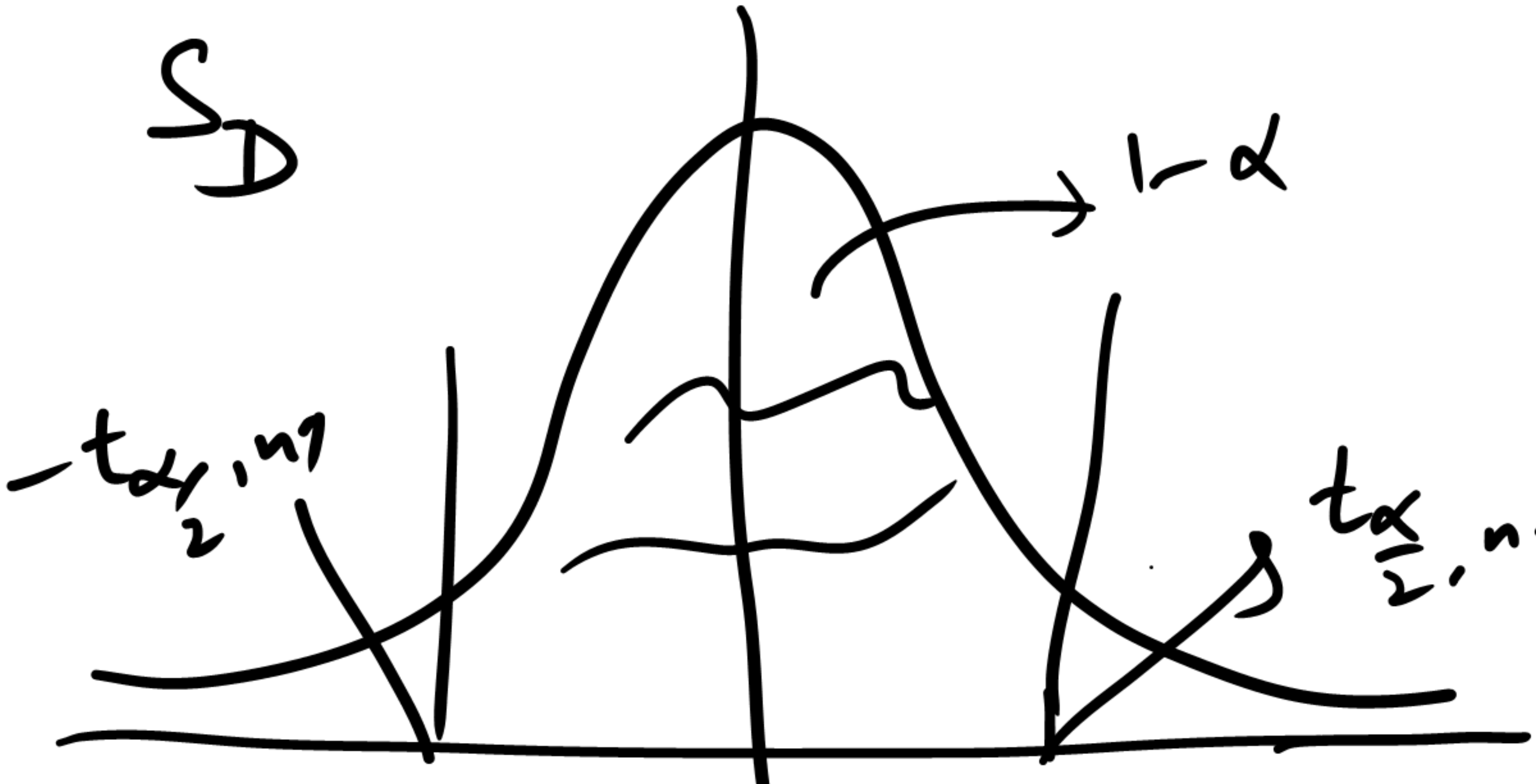
$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$, $S_D^2 = \frac{1}{n-1}\sum(d_i - \bar{d})^2$

$$\bar{d} \sim N\left(\eta, \frac{\sigma_D^2}{n}\right)$$

$\frac{(n-1)S_D^2}{\sigma_D^2} \sim \chi^2_{n-1}$

$\bar{d}$ & S_D^2 are independently distributed.

$$\frac{\sqrt{n}(\bar{d}-\eta)}{S_D} \sim t_{n-1}$$



$$P\left(-t_{\alpha/2,n-1} \leq \frac{\sqrt{n}(\bar{d}-\eta)}{S_D} \leq t_{\alpha/2,n-1}\right)$$

$$= 1-\alpha$$

Then

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

$$= 1-\alpha$$

So in this case $100(1-\alpha)\%$ C.I.

for $\eta = \mu_1 - \mu_2$ is

$$\bar{d} + \frac{S_D}{\sqrt{n}}\, t_{\alpha/2,\, n-1}$$

<u>Example</u>: To compare the gripping strength of left hand and right hand of left-handed persons, the measurements are made on 10 randomly selected persons.

Person	LHGS	RHGS
1	140	138
2	90	87
3	125	110
4	130	132
5	95	96
6	121	120
7	85	86

8	97	90
9	131	129
10	110	100
i	x_i	y_i

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

$\bar{d} = 3.6$, $s_d^2 = 31.26$.

For 90% C.I, $t_{0.05,9} = 1.833$

So 90% C.I. for $\eta = \mu_1 - \mu_2$ is

$$\left(3.6 \pm \sqrt{\frac{31.26}{10}} \times 1.833\right)$$

3.24

$$\simeq (0.36,\ 6.84)$$

<u>Example</u> : Equal but unknown variances

$\bar{x} = 18.5$, $s_1 = 5.8$, $\bar{y} = 20.7$

$s_2 = 6.3$, $m = n = 100$

$$S_p^2 = \frac{99 S_1^2 + 99 S_2^2}{198} = \frac{(5.8)^2 + (6.3)^2}{2}$$

$$= 36.665$$

<u>90% CI</u>

$$t_{0.05,198} = 1.645$$

<u>for η:</u>

$$18.5 - 20.7 \pm \sqrt{\frac{200}{100 \times 100}} \sqrt{36.665} \times 1.645$$

<u>Example</u>: Known σ_1^2 & σ_2^2.

$\bar{x} = 30.87$, $\bar{x}_2 = 30.68$

$\sigma_1 = 0.15, \quad \sigma_2 = 0.12, \quad m = 12,$
$n = 10$

90% C.I for $\mu_1 - \mu_2$: $\quad z_{0.05} = 1.645$

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

$\Rightarrow \quad (0.095, \; 0.285)$
is 90% C.I. for η.

Comparing Variances

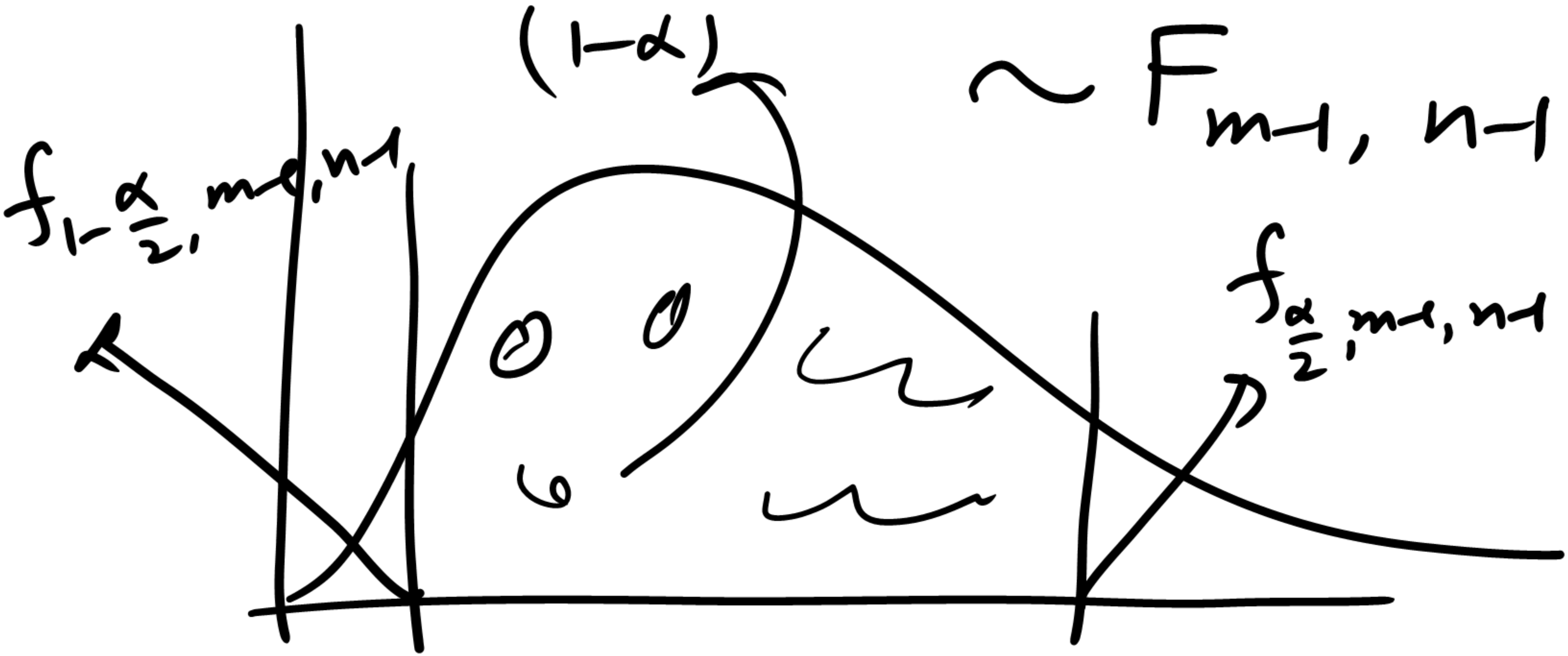
$$x = \frac{\sigma_1^2}{\sigma_2^2}$$

$$\frac{(m-1)\,S_1^2}{\sigma_1^2} \sim \chi^2_{m-1}$$

indept

$$\frac{(n-1)\,S_2^2}{\sigma_2^2} \sim \chi^2_{n-1}$$

$$\frac{(m-1)S_1^2}{\sigma_1^2(m-1)} \Big/ \frac{(n-1)S_2^2}{\sigma_2^2(n-1)} = \frac{\sigma_2^2}{\sigma_1^2}\,\frac{S_1^2}{S_2^2}$$



$$P\left(f_{1-\frac{\alpha}{2}, m-1, n-1} \le \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \le f_{\alpha/2, m-1, n-1} \right) = 1-\alpha$$

$$P\left(\frac{S_1^2}{S_2^2} f_{1-\frac{\alpha}{2}, m-1, n-1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2} f_{\frac{\alpha}{2}, m-1, n-1} \right)$$
$$= 1-\alpha$$

For $100(1-\alpha)\%$ C.I for $\gamma = \frac{\sigma_1^2}{\sigma_2^2}$ is

$$\left(\frac{S_1^2}{S_2^2} f_{1-\frac{\alpha}{2}, m-1, n-1} \;,\; \frac{S_1^2}{S_2^2} f_{\frac{\alpha}{2}, m-1, n-1} \right)$$

<u>Example</u>: Viscosity of two brands

of oil used in cars :

Brand 1 : 10.62, 10.58, 10.33, 10.72, 10.44

Brand 2 : 10.50, 10.52, 10.62, 10.53

90% C.I for $\psi = \sigma_1^2/\sigma_2^2$

$S_1^2 = 0.02362$, $S_2^2 = 0.002825$

$S_1^2/S_2^2 = 8.36$

$$f_{0.05,4,3} = 9.1172$$

$$f_{0.95,4,3} = 0.1517$$

90% C.I. for τ is $(1.268, 76.22)$