Significance tests (cont.…)

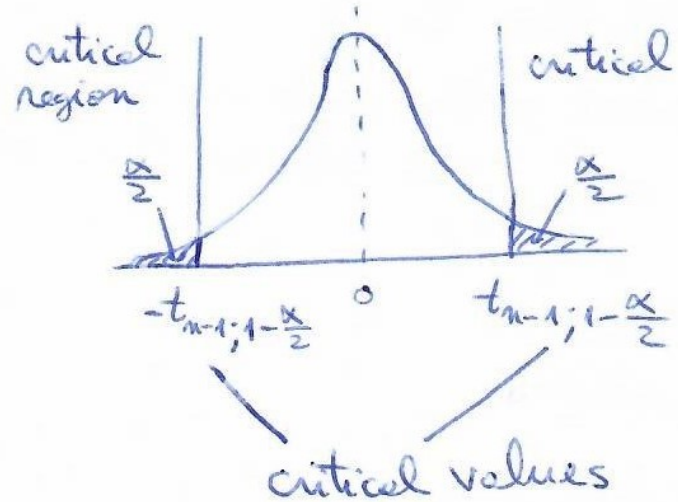
For the acceptance/critical region approach (also called the critical value approach), we need to compute the test statistic and to find the critical values (i.e. the quantiles) corresponding to a given significance level α.

For the one-sample T-tests, depending on the nature of the alternative hypothesis, we use a two-tailed or a one-tailed test:

One sample t-test

- two-tailed test:

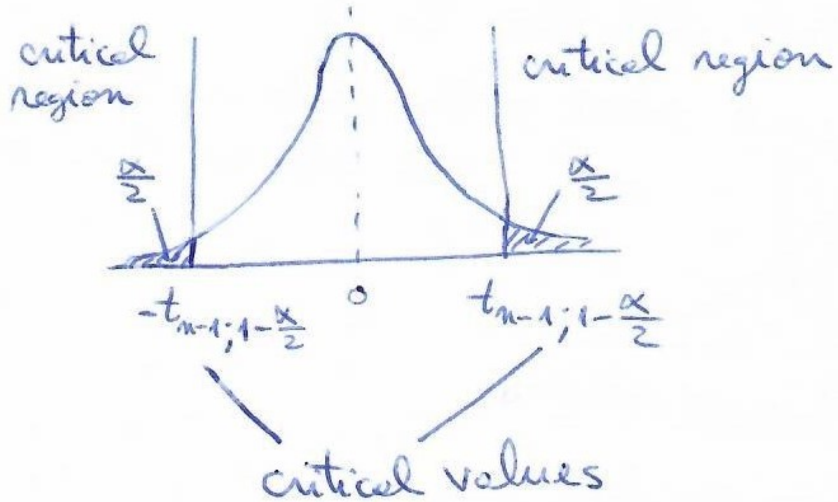
  $H_0 : \{\mu = \mu_0\}$ $\quad H_a : \{\mu \neq \mu_0\}$



One sample t-test

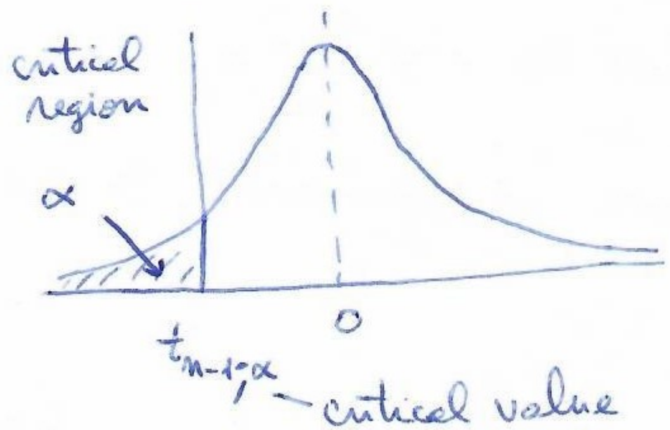
- two-tailed test:

  $H_0: \{\mu = \mu_0\}$ $H_a: \{\mu \neq \mu_0\}$



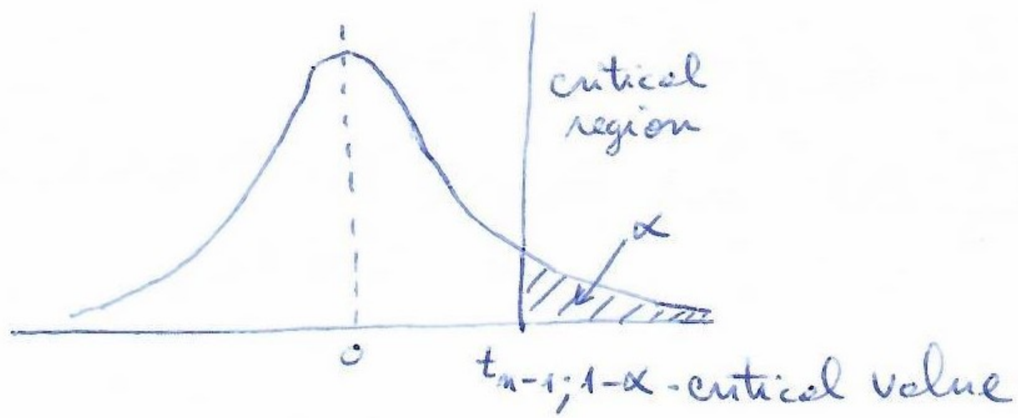
- left one-tailed test:

  $H_0: \{\mu = \mu_0\}$ $H_a: \{\mu < \mu_0\}$



- right one-tailed test:

  $H_0: \{\mu = \mu_0\}$ $H_a: \{\mu > \mu_0\}$



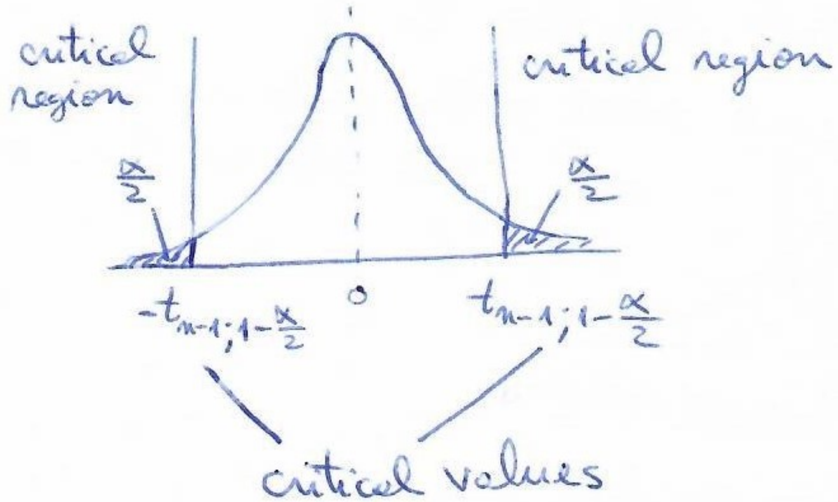
One sample t-test

in all cases, the test statistic is $t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s}$

if $t \in$ critical region then H_0 is rejected; otherwise, we fail to reject H_0

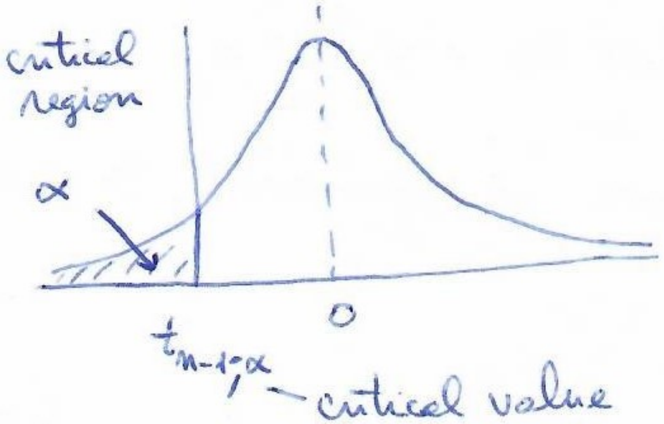
- two-tailed test:

  $H_0: \{\mu = \mu_0\}$ $H_a: \{\mu \neq \mu_0\}$



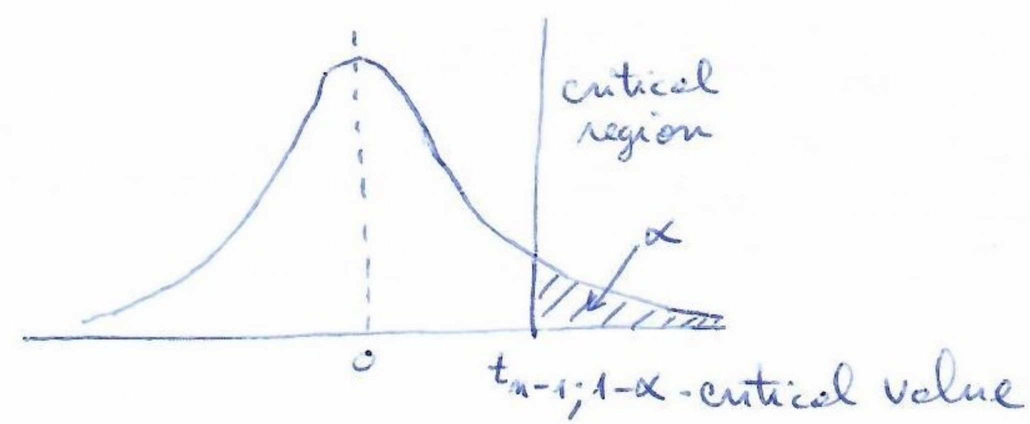
- left one-tailed test:

  $H_0: \{\mu = \mu_0\}$ $H_a: \{\mu < \mu_0\}$



- right one-tailed test:

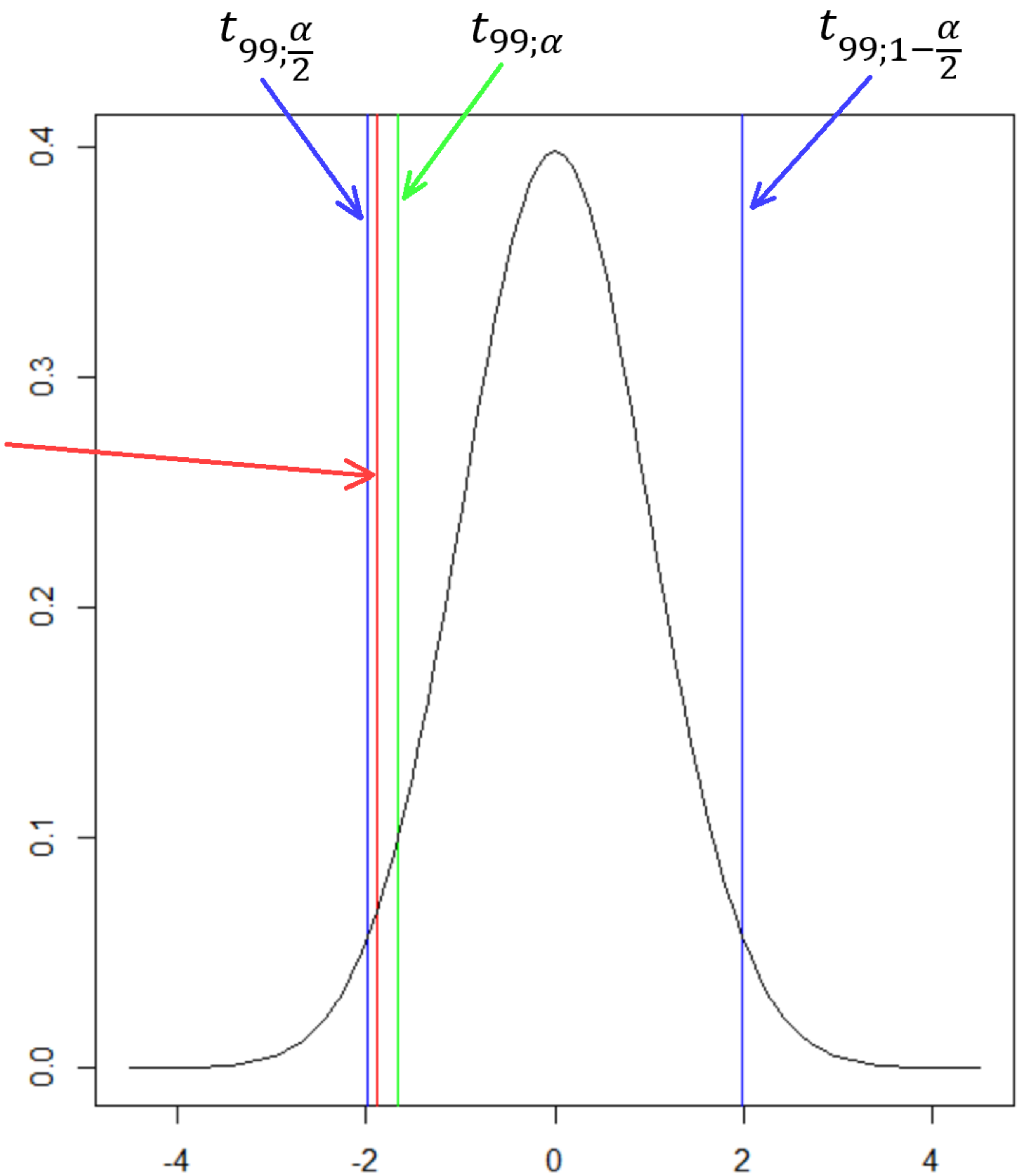
  $H_0: \{\mu = \mu_0\}$ $H_a: \{\mu > \mu_0\}$



One sample t-test – one-tailed vs. two-tailed

α=0.05 df=99

$$t = \frac{\sqrt{n}\,(\bar{x}-\mu)}{s}$$



x<-rnorm(100,2,2)

a=t.test(x,alternative="less",mu=2.13)

b=t.test(x,alternative="two.sided",mu=2.13)

a$statistic is -1.877531

b$statistic is -1.877531

The p-value

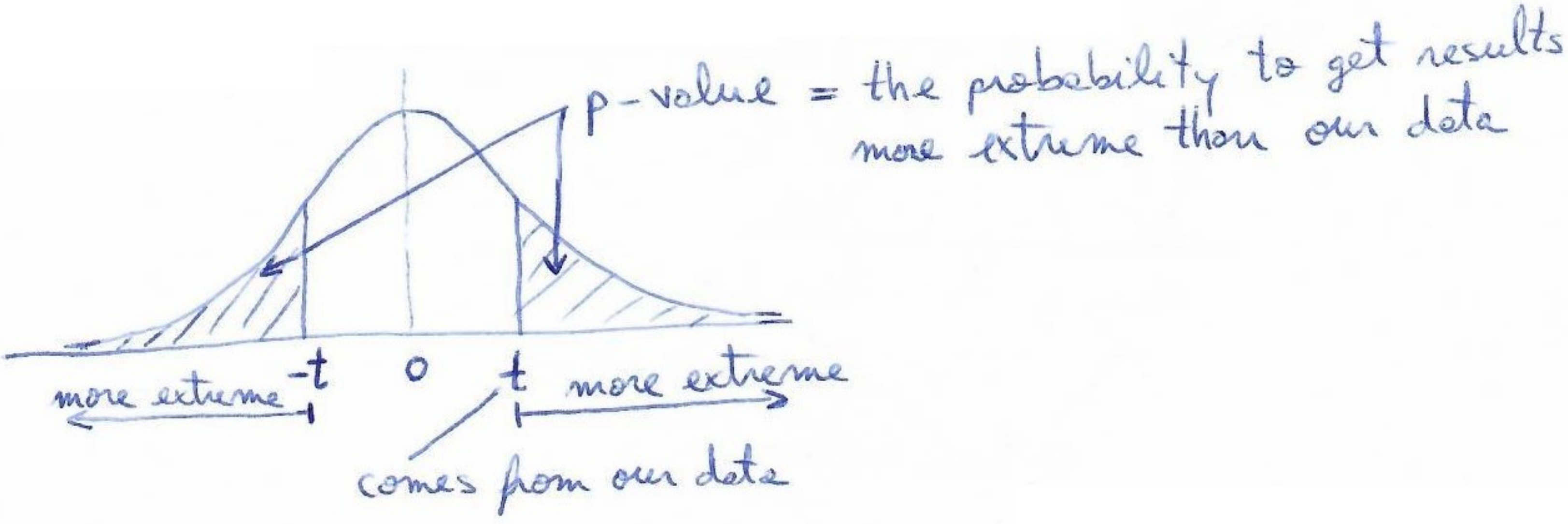
It gives the same result as the critical value approach when testing hypothesis (it is used by most statistical software).

The p-value is the probability that, under the assumption that the null hypothesis is true, we obtain a result "more extreme" than the one observed in our data.

p-values are triggers to decide when to reject the null hypothesis and when to fail to reject it.

The p-value

For a two-tailed T-test, let's assume that the sample data give a positive test statistic t



The p-value

if p-value $< \alpha$ then the test statistic is "extreme enough" to reject H_0

if p-value $\geq \alpha$ then we fail to reject H_0

Obs. The condition p-value $< \alpha$ is equivalent to $t \in$ critical region of the test.

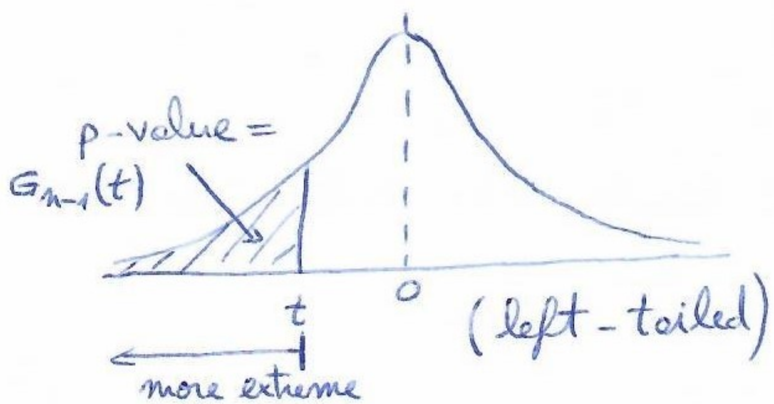
For the two-tailed T-test, the p-value is as follows:

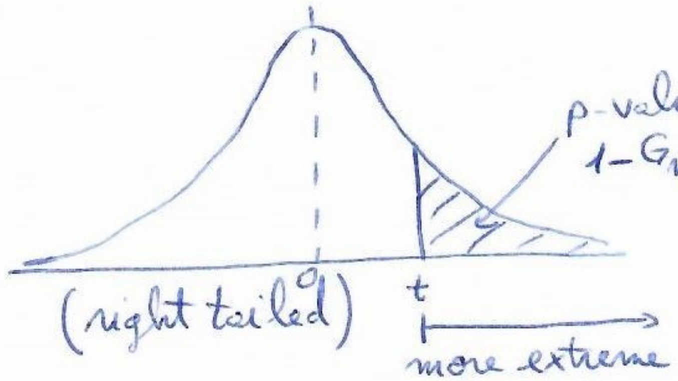
$$p = \begin{cases} 2\, G_{n-1}(t), & \text{if } t \leq 0 \\ 2\left(1 - G_{n-1}(t)\right), & \text{if } t > 0 \end{cases}$$

where $t = \dfrac{\sqrt{n}\,(\bar{x} - \mu_0)}{s}$ and G_{n-1} is the distribution function for t_{n-1}

The p-value

For the left one-tailed T-test, the p-value is $G_{n-1}(t)$ and for the right one-tailed T-test, $1 - G_{n-1}(t)$.





Obs. The p-value approach lets us choose easily the significance level value.

The F-test

- for the variances of two normal distributions

  $X_1, \dots, X_n$ iid $N(\mu_1, \sigma_1^2)$

  $Y_1, \dots, Y_m$ iid $N(\mu_2, \sigma_2^2)$

$H_0 : \{\sigma_1^2 = \sigma_2^2\}$ against $H_a : \{\sigma_1^2 \neq \sigma_2^2\}$ (two-tailed)

(or one-tailed alternative
$H_a : \{\sigma_1^2 > \sigma_2^2\}$ or $H_a : \{\sigma_1^2 < \sigma_2^2\}$)

The F-test

Let $S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \quad S_2^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \overline{Y})^2$

It can be proven that
$$\left.\begin{array}{l} \dfrac{(n-1)\,S_1^2}{\sigma_1^2} \sim \chi^2(n-1) \\[2ex] \dfrac{(m-1)\,S_2^2}{\sigma_2^2} \sim \chi^2(m-1) \\[2ex] \text{they are independent} \end{array}\right\} \Longrightarrow$$

$$\frac{\frac{1}{n-1}\cdot\frac{(n-1)\,S_1^2}{\sigma_1^2}}{\frac{1}{m-1}\cdot\frac{(m-1)\,S_2^2}{\sigma_2^2}} = \frac{S_1^2}{S_2^2}\cdot\frac{\sigma_2^2}{\sigma_1^2} \sim F_{n-1,\,m-1} \qquad F \text{ distribution}$$

The F-test

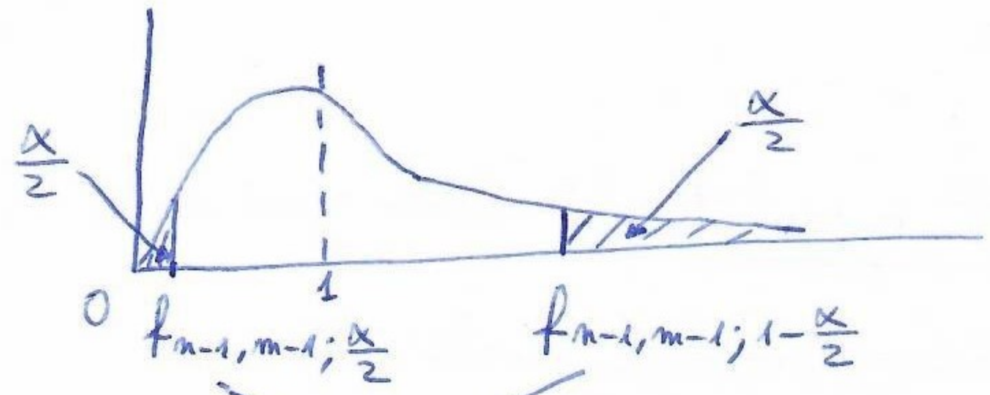
If we denote by $\gamma = \frac{\sigma_1^2}{\sigma_2^2}$, then our testing problem can be written as

$$H_0 : \{\gamma = 1\} \qquad H_a : \{\gamma \neq 1\}$$

The test statistic is $\frac{S_1^2}{S_2^2}$. Under the null hypothesis $(\gamma = 1)$, $\frac{S_1^2}{S_2^2}$ has an $F_{n-1, m-1}$ distribution.

For the significance level α, the critical values of the two tailed F-test are:

The F-test



critical values = quantiles of $F_{n-1,m-1}$ distribution

The acceptance region of H_0 is:

$$W_{n,m;1-\alpha}(\gamma=1) = \left\{(x_1,\dots,x_n,y_1,\dots,y_m) \;\middle|\; f_{n-1,m-1;\frac{\alpha}{2}} \le \frac{s_1^2}{s_2^2} \le f_{n-1,m-1;1-\frac{\alpha}{2}}\right\}$$

The probability of type I error $P(H_a|H_0) = \alpha$

The F-test

The p-value $= \begin{cases} 2\, S_{n-1,m-1}\left(\frac{s_1^2}{s_2^2}\right), & \text{if } \frac{s_1^2}{s_2^2} < 1 \\ 2\left(1 - S_{n-1,m-1}\left(\frac{s_1^2}{s_2^2}\right)\right), & \text{if } \frac{s_1^2}{s_2^2} \geq 1 \end{cases}$ $S_{n-1,m-1}$ is the $F_{n-1,m-1}$ distribution function

<u>Obs.</u> p-value $< \alpha$ iff $\frac{s_1^2}{s_2^2} \in \overline{W_{n,m;1-\alpha}}(\gamma=1)$

The operating characteristic function

$$
\begin{aligned}
OC(\gamma) &= P_\gamma\left((X_1, \ldots, X_n, Y_1, \ldots, Y_m) \in W_{n,m;1-\alpha}(\gamma=1)\right) \\
&= P_\gamma\left(f_{n-1,m-1;\frac{\alpha}{2}} \leq \frac{S_1^2}{S_2^2} \leq f_{n-1,m-1;1-\frac{\alpha}{2}}\right) \\
&= P_\gamma\left(\frac{1}{\gamma} f_{n-1,m-1;\frac{\alpha}{2}} \leq \underbrace{\frac{S_1^2}{S_2^2} \cdot \frac{1}{\gamma}}_{\sim F_{n-1,m-1}} \leq \frac{1}{\gamma} \cdot f_{n-1,m-1;1-\frac{\alpha}{2}}\right) \\
&= S_{n-1,m-1}\left(\frac{1}{\gamma} \cdot f_{n-1,m-1;1-\frac{\alpha}{2}}\right) - S_{n-1,m-1}\left(\frac{1}{\gamma} f_{n-1,m-1;\frac{\alpha}{2}}\right)
\end{aligned}
$$

Two sample t-test

- for the means of two normal distributions

$X_1, \dots, X_n$ iid $N(\mu_1, \sigma_1^2)$

$Y_1, \dots, Y_m$ iid $N(\mu_2, \sigma_2^2)$

1. <u>The case of equal variances</u> $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$H_0 : \{\mu_1 = \mu_2\}$ $H_a : \{\mu_1 \neq \mu_2\}$ (or the one-sided alternatives $\{\mu_1 > \mu_2\}$ or $\{\mu_1 < \mu_2\}$)

Two sample t-test

it follows that $\overline{X} - \overline{Y} \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right) \Longrightarrow$

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0,1)$$

As $\frac{1}{\sigma^2}\left((n-1)S_1^2 + (m-1)S_2^2\right) \sim \chi^2(n+m-2)$, it follows that

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)\left((n-1)S_1^2 + (m-1)S_2^2\right)}} \sim t_{n+m-2}$$

if we introduce $\delta = \mu_1 - \mu_2$ then the hypotheses become:

$H_0: \{\delta = 0\}$ against $H_a: \{\delta \neq 0\}$ (for the two-tailed case)

Two sample t-test

The test statistic is

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{1}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)\left((n-1)S_1^2 + (m-1)S_2^2\right)}}$$

Under the null hypothesis assumption, $t \sim t(n+m-2)$

The test is built now in a similar way to the one-sample T-test

Thus, the acceptance region of H_0 is

$$W_{n,m;1-\alpha}(\delta = 0) = \{(x_1, \dots, x_n, y_1, \dots, y_m) \mid -t_{n+m-2;1-\frac{\alpha}{2}} \le t \le t_{n+m-2;1-\frac{\alpha}{2}}\}$$

Two sample t-test

$$P(H_a \mid H_0) = \alpha$$

The p-value $= \begin{cases} 2 \cdot G_{n+m-2}(t), & \text{if } t \leq 0 \\ 2(1 - G_{n+m-2}(t)), & \text{if } t > 0 \end{cases}$

The operating characteristic function is:

$$OC(\delta) = P_\delta\left(-t_{n+m-2;1-\frac{\alpha}{2}} \leq t \leq t_{n+m-2;1-\frac{\alpha}{2}}\right)$$

$$= P_\delta\Big(-t_{n+m-2;1-\frac{\alpha}{2}} - \Delta \leq \underbrace{t - \Delta}_{\sim t_{n+m-2}} \leq t_{n+m-2;1-\frac{\alpha}{2}} - \Delta\Big)$$

$$= G_{n+m-2}\left(t_{n+m-2;1-\frac{\alpha}{2}} - \Delta\right) - G_{n+m-2}\left(-t_{n+m-2;1-\frac{\alpha}{2}} - \Delta\right),$$

where $\Delta = \dfrac{\delta}{\sqrt{\frac{1}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)\left((n-1)S_1^2 + (m-1)S_2^2\right)}}$

Two sample t-test

2. <u>The case of unequal variances</u> $\sigma_1^2 \neq \sigma_2^2$

The test statistic is $t' = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}}$

If H_0 is true, then $t' \sim t_{df}$, where

$$df = \frac{\left(\frac{S_1^2}{n} + \frac{S_2^2}{m}\right)^2}{\frac{1}{n-1}\left(\frac{S_1^2}{n}\right)^2 + \frac{1}{m-1}\cdot\left(\frac{S_2^2}{m}\right)^2}$$

(this is the Welch approximation for the degrees of freedom)

Using now the quantile $t_{df;1-\frac{\alpha}{2}}$, the test is similar with the case of equal variances.

The paired two sample t-test

$(X_1, Y_1), \ldots, (X_n, Y_n) \sim \text{iid } N(\mu, \Sigma)$, where

$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}$$

Denote $\delta = \mu_x - \mu_y$.

$H_0 : \{\delta = 0\}$ against $H_a : \{\delta \neq 0\}$

The paired two sample t-test

Let $d_i = X_i - Y_i$, $i = \overline{1,n}$

$$\left.\begin{aligned}
\sigma_{xy} &= E\left((X-\mu_x)(Y-\mu_y)\right) = E(XY) - \mu_x\mu_y \\
\operatorname{var}(X-Y) &= E\left((X-Y)^2\right) - \left[E(X-Y)\right]^2 \\
&= E\left(X^2+Y^2-2XY\right) - (\mu_x-\mu_y)^2 \\
&= \underbrace{E(X^2)-\mu_x^2}_{\sigma_x^2} + \underbrace{E(Y^2)-\mu_y^2}_{\sigma_y^2} - 2E(XY) + 2\mu_x\mu_y
\end{aligned}\right\} \Rightarrow$$

$$\Rightarrow \operatorname{Var}(X-Y) = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}$$

The paired two sample t-test

Thus, $d_1, \ldots, d_n \sim iid\ N(\delta, \sigma_d^2)$, where $\delta = \mu_x - \mu_y$,

$$\sigma_d^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}$$

The estimators for δ and σ_d^2 are:

$$\hat{\delta} = \bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$$

$$\hat{\sigma}_d^2 = s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$$

The test is built in a similar way to the one-sample T-test.

We have that $\dfrac{\sqrt{n}(\bar{d} - \delta)}{s_d} \sim t_{n-1}$

The test statistic is $t = \dfrac{\sqrt{n}\,\bar{d}}{s_d}$.

The paired two sample t-test

The acceptance region for H_0 at the significance level α is:

$$W_{n;1-\alpha} = \left\{ (x_1, y_1), \dots, (x_n, y_n) \,\middle|\, -t_{n-1;1-\frac{\alpha}{2}} \le \frac{\sqrt{n}\,\bar{d}}{s_d} \le t_{n-1;1-\frac{\alpha}{2}} \right\}$$

$$P(H_a \mid H_0) = \alpha$$

$$\text{The } p\text{-value} = \begin{cases} 2 \cdot G_{n-1}\left(\frac{\sqrt{n}\,\bar{d}}{s_d}\right), & \text{if } \frac{\sqrt{n}\,\bar{d}}{s_d} < 0 \\ 2\left(1 - G_{n-1}\left(\frac{\sqrt{n}\,\bar{d}}{s_d}\right)\right), & \text{if } \frac{\sqrt{n}\cdot\bar{d}}{s_d} \ge 0 \end{cases}$$

The paired two sample t-test

The operating characteristic function is:

$$OC(\delta) = P_\delta\left(-t_{n-1;1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}\,\bar{d}}{s_d} \leq t_{n-1;1-\frac{\alpha}{2}}\right)$$

$$= P_\delta\left(-t_{n-1;1-\frac{\alpha}{2}} - \frac{\sqrt{n}\,\delta}{s_d} \leq \underbrace{\frac{\sqrt{n}(\bar{d}-\delta)}{s_d}}_{\sim t_{n-1}} \leq t_{n-1;1-\frac{\alpha}{2}} - \frac{\sqrt{n}\,\delta}{s_d}\right)$$

$$= G_{n-1}\left(t_{n-1;1-\frac{\alpha}{2}} - \frac{\sqrt{n}\,\delta}{s_d}\right) - G_{n-1}\left(-t_{n-1;1-\frac{\alpha}{2}} - \frac{\sqrt{n}\,\delta}{s_d}\right)$$