Types of errors

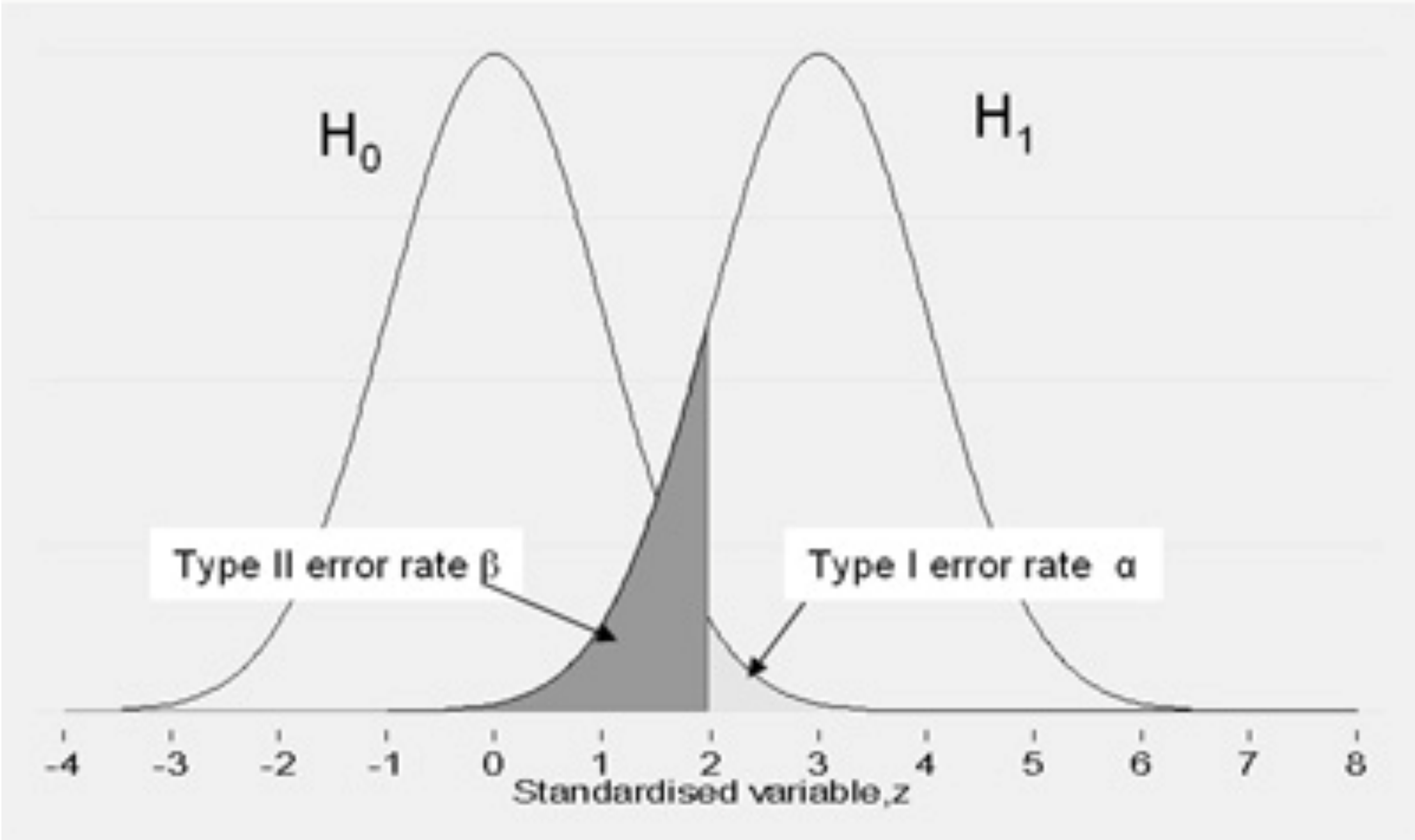
The conclusion in Step 4 could be wrong, as we are looking at a sample with a limited number n of elements

False positive

		Decision	
		Retain H_0	Reject H_0
Truth in the population	True	**Correct** $1 - \alpha$	**Type I error** α
	False	**Type II error** β	**Correct** $1 - \beta$ **(Power)**

False negative

Types of errors



Some errors are worse than others

The incorrect decision is to retain a false null hypothesis. This is equivalent of doing nothing. We can do more experiments and test again the null hypothesis. Not so bad…

…ld be wrong, as we are looking at …er n of elements

False positive

		Decision	
		Retain H_0	Reject H_0
Truth in the population	True	**Correct** $1 - \alpha$	**Type I error** α
	False	**Type II error** β	**Correct** $1 - \beta$ **(Power)**

False negative

Some errors are worse than others

The … s we are looking at
a sa… …ts

The incorrect decision is to reject a true null hypothesis. This means rejecting a previous notions of truth that are in fact true (this is equivalent to finding an innocent person guilty).

False positive

		…ion	
		Retain H_0	Reject H_0
Truth in the population	True	**Correct** $1 - \alpha$	**Type I error** α
	False	**Type II error** β	**Correct** $1 - \beta$ **(Power)**

False negative

Note that you can directly control the probability of a Type I error by stating an alpha level

The decision we are looking for

Strong conclusion
This is the decision we are looking for when we test the hypothesis. If we test it, it means we have doubts about such hypothesis.

g, as we are looking at
ments

False positive

		ecision	
		Retai	Reject H_0
Truth in the population	True	**Correct** $1 - \alpha$	**Type I error** α
	False	**Type II error** β	**Correct** $1 - \beta$ **(Power)**

False negative

Decision we want

Measuring the size of an effect

- A decision to reject the null hypothesis means that **an effect is significant**. Hypothesis testing does not inform on how big the effect is.
- **Effect size** is a statistical measure of the size of an effect in a population. It particularly makes sense when the null hypothesis is rejected.
- **Cohen's *d*** measures the number of standard deviations an effect shifted above or below the population mean stated by the null hypothesis

Cohen's d measure formula

Mean of the sample

Mean of the population

Standard deviation of the population

$$\text{Cohen's } d = \frac{M - \mu}{\sigma}$$

Cohen's effect size conventions are often used to interpret the effect size

If values of *d* are negative, the effect shifted below the population mean

Description of Effect	Effect Size (d)
Small	$\lvert d \rvert < 0.2$
Medium	$0.2 < \lvert d \rvert < 0.8$
Large	$\lvert d \rvert > 0.8$

The example again: *Cohen's d*

Assume you are the database administrator of a big information system and you are unhappy with the execution time of a given SQL package.

From historical data (thousands of previous package executions), you know that the average execution time of the package is **83.54** seconds with a standard deviation of **16.36**.

You change the tuning of the database and run the package several times to check the effect.

$$\textit{Cohen's d} = \frac{M - \mu}{\sigma} = \frac{78.15 - 83.54}{16.36} = -0.33$$

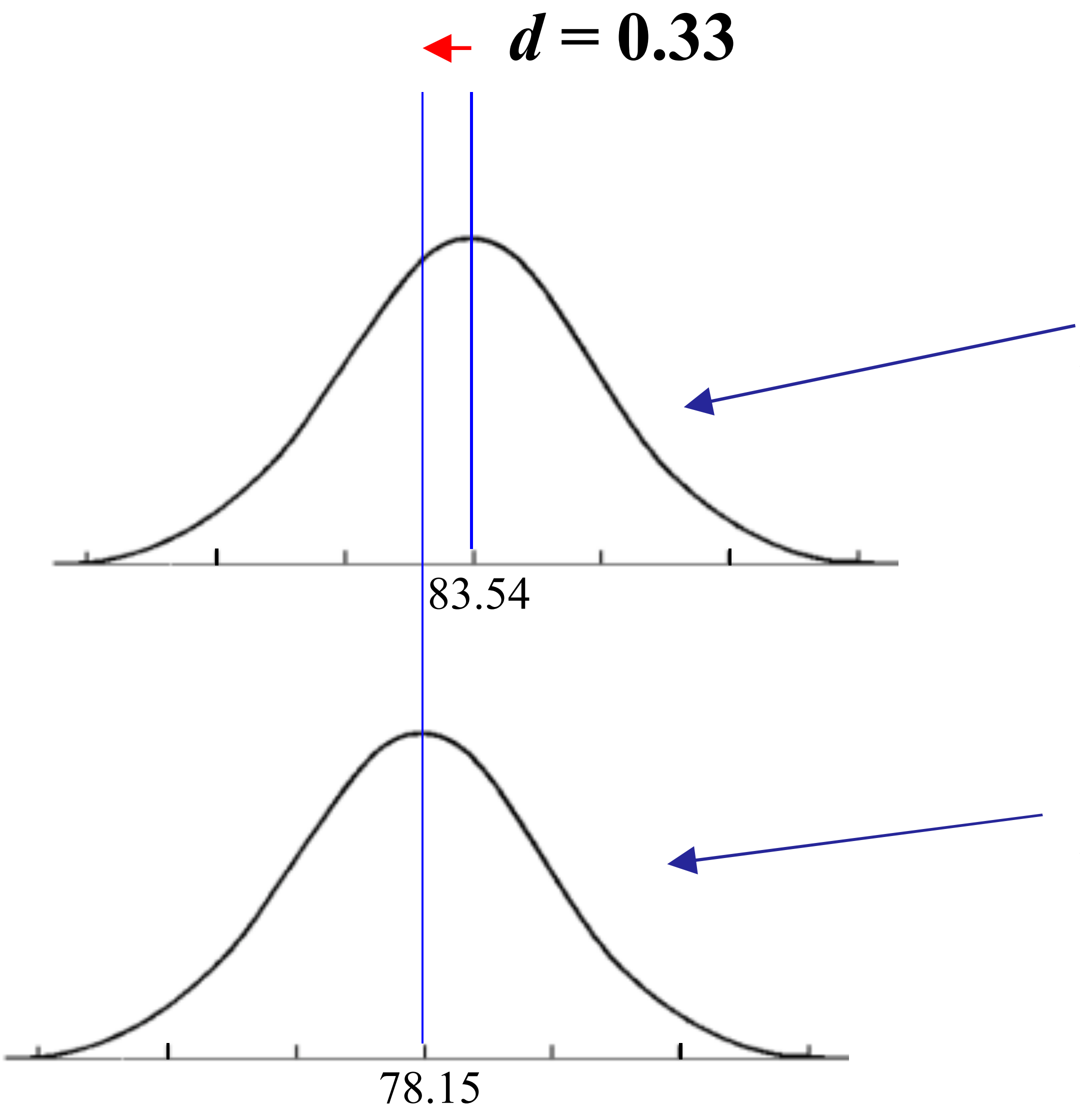
The observed effect shifted 0.33 standard deviations below the mean

Package exec. time
74
66
88
68
70
⋮
79
78
72
86
85
86

32 times

Avg = 78.15

The example again: *Cohen's d*



Population distribution assuming the null hypothesis is true

Population distribution assuming the null hypothesis is false with a 5.39 points effect

T-test

- The test statistic of t test follows Student's T distribution (under true null hypothesis)

- Two types:
  - **One-sample T-tests** → Compare population mean with a value
  - **Two-sample T-tests** → Compare two population means

Independent samples:
unrelated separate groups

- T-test should be applied when:
  - The **sample size is small** ($n < 30$)
  - The standard deviation of the population(s) is (are) unknown
  - The population(s) follows a normal distribution

(when the number of samples is large, t test and z test give similar results)

Hypothesis testing using T-test (two independent samples)

- Under the assumption of **equal variances**

Means of the two samples

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

The degree of freedom is

$$n_1 + n_2 - 2$$

Common standard deviation of the two samples

$$S_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

where n_1 and n_2 is the size of the sample of both samples and s_1 and s_2 is the (sample) standard deviation of both groups

Hypothesis testing using T-test (two independent samples)

- Under the assumption of **unequal variances**

Means of the two samples

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

The degree of freedom is

$$\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Also known as Welsh test

Example 4 - Hypothesis testing using T-test (two independent samples)

Assume you are the database administrator of a big information system. The database has just been installed and you are trying two tuning configurations: Conf. **A** and Conf. **B**. You believe that Conf. **B** should be the fastest.

You use a given SQL package to test the execution time for each configuration.

After running several times the SQL package in both configurations you want to take a decision.

Question: Is Conf. **B** faster than Conf. **A** ?

Conf. A exec. time	Conf. B exec. time
74	69
66	71
88	80
68	88
79	64
68	65
87	74
79	76
78	89
72	68
86	67
85	72
86	

$\mu_1 = 78.15$ $\mu_2 = 73.58$

$s_1 = 7.94$ $s_2 = 8.33$

$n = 13$ $n = 12$

Example 4: **t test** (two independent samples)

Step 1- State the hypothesis

- **H_0: $\mu_1 = \mu_2$**
  In words: configuration A and B are equivalent concerning the execution time of the SQL package

- **H_1: $\mu_1 > \mu_2$**
  Configuration B is faster than configuration A (i.e., the execution time of the SQL package is higher in configuration A)

Example 4: **t test** (two independent samples)

Step 2 - Set the criteria for a decision

- Assume similar variances ($0.5 < s_1^2 \,/\, s_2^2 < 2$)
- Consider the level of confidence of 95%→ α = 0.05
- Locate the **t score** (in the t table for the Student distribution, one-tailed) that represents the **critical value** (for α = 0.05 and df = 23)
- Look in the t table for:
  - α = 0.05 (α level for a confidence of 95%)
  - df = 23 (degree of freedom = 23)

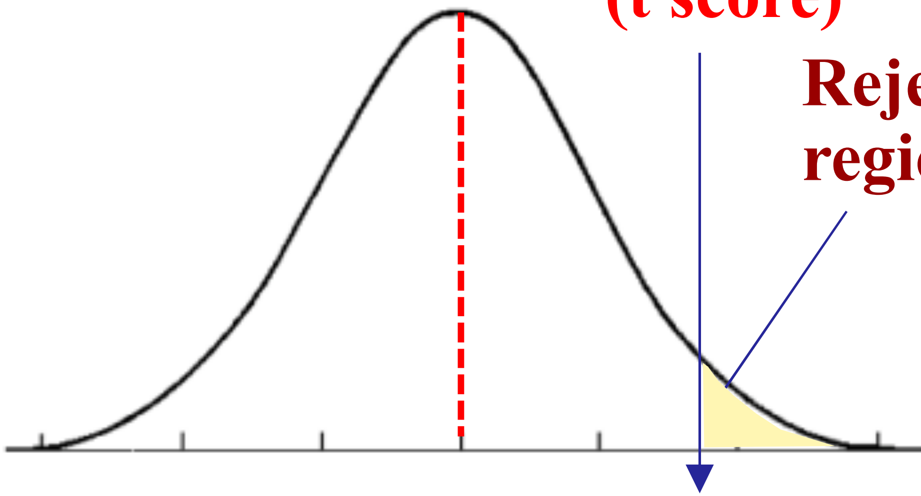
→ t score = 1.714

Example 4: **t test** (two independent samples)

Step 2 - Set the criteria for a decision

- Assume similar variances ($0.5 < s_1^2 / s_2^2 < 2$)
- Consider the level of confidence of 95%→ α = 0.05
- Locate the **t score** (in the t table for the Student distribution, one-tailed) that represents the **critical value** (for α = 0.05 and df = 23)
- Look in the t table for:
  - α = 0.05 (α level for a confiden
  - df = 23 (degree of freedom = 23)

→ **t score = 1.714**



Example 4: **t test** (two independent samples)

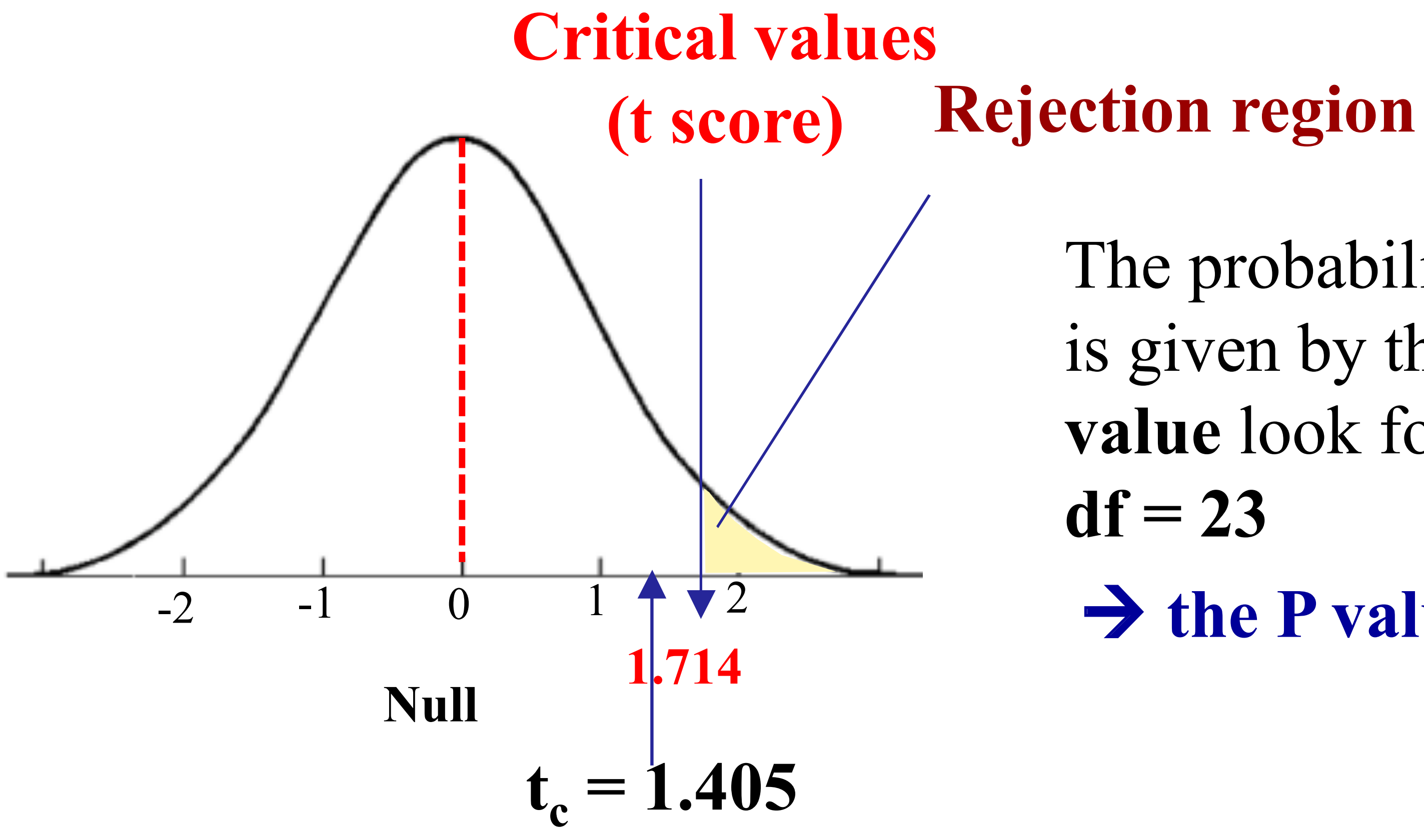
Step 3 - Compute the test statistic

Sample	**Configuration**	**n**	***x***	**s**
1	A	13	78.15	7.94
2	B	12	73.53	8.33

Test statistic:

$$t_c = \frac{\bar{x}_1 - \bar{x}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = 1.405$$

Example 4: **t test** (two independent samples) Step 4 - Make a decision

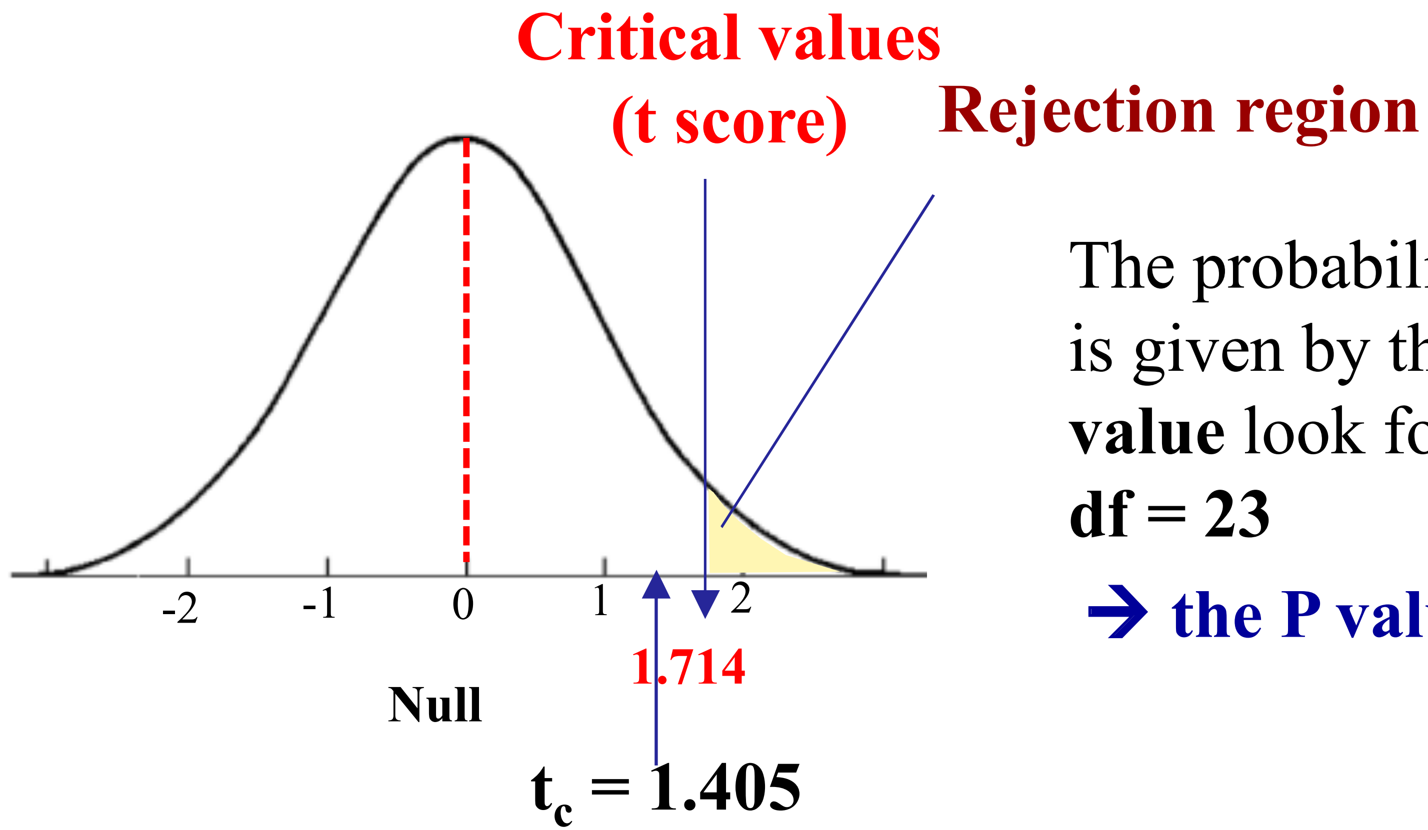


The probability of obtaining $\mathbf{t_c = 1.405}$ is given by the ***P* value**. To obtain ***P* value** look for 1.405 in the **t table**, for **df = 23**

→ the P value is 0.0867 (P = 8.67%)

As *P* > 5% → **Retain the null hypothesis (fail reaching significance)**

Example 4: **t test** (two independent samples)
Step 4 - Make a decision



The probability of obtaining $\mathbf{t_c = 1.405}$ is given by the ***P* value**. To obtain ***P* value** look for 1.405 in the **t table**, for **df = 23**

→ **the P value is 0.0867 (P = 8.67%)**

As *P* > 5% → **Retain the null hypothesis**

We cannot show that configuration B is faster than A

Hypothesis testing using T-test (two dependent samples)

Mean of the differences between the two samples

$$t = \frac{\bar{x}_D - \mu}{\frac{s_D}{\sqrt{n}}}$$

Standard deviation of the differences

The degree of freedom is

$n - 1$

Example 5 - Hypothesis testing using T-test (two dependent samples)

Assume you are the database administrator of a big information system. The database has just been installed and you are trying two tuning configurations: Conf. **A** and Conf. **B**. You believe that Conf. **B** should be the fastest.

You use a given SQL package to test the execution time for each configuration.

After running several times the **same** SQL queries in both configurations you want to take a decision.

Question: : Is Conf. **B** faster than Conf. **A** ?

Conf. A exec. time	Conf. B exec. time
74	69
66	71
88	80
68	88
79	64
68	65
87	74
79	76
78	89
72	68
86	67
85	72

$\mu_1 = 77.5$ $\mu_2 = 73.58$

$n = 12$ $n = 12$

Example 5: **t test** (two dependent samples)

Step 1- State the hypothesis

- **H_0: $\mu_1 = \mu_2$**
  In words: configuration A and B are equivalent concerning the execution time of the SQL package
- **H_1: $\mu_1 > \mu_2$**
  Configuration B is faster than configuration A (i.e., the execution time of the SQL package is higher in configuration A)

Example 5: **t test** (two dependent samples)

Step 2 - Set the criteria for a decision

- Consider the level of confidence of 95%→ $\alpha = 0.05$
- Locate the **t score** (in the t table for the Student distribution, one-tailed) that represents the **critical value** (for $\alpha = 0.05$ and df = 11)
- Look in the t table for:
  - $\alpha = 0.05$ (α level for a confidence of 95%)
  - df = 11 (degree of freedom = 11)

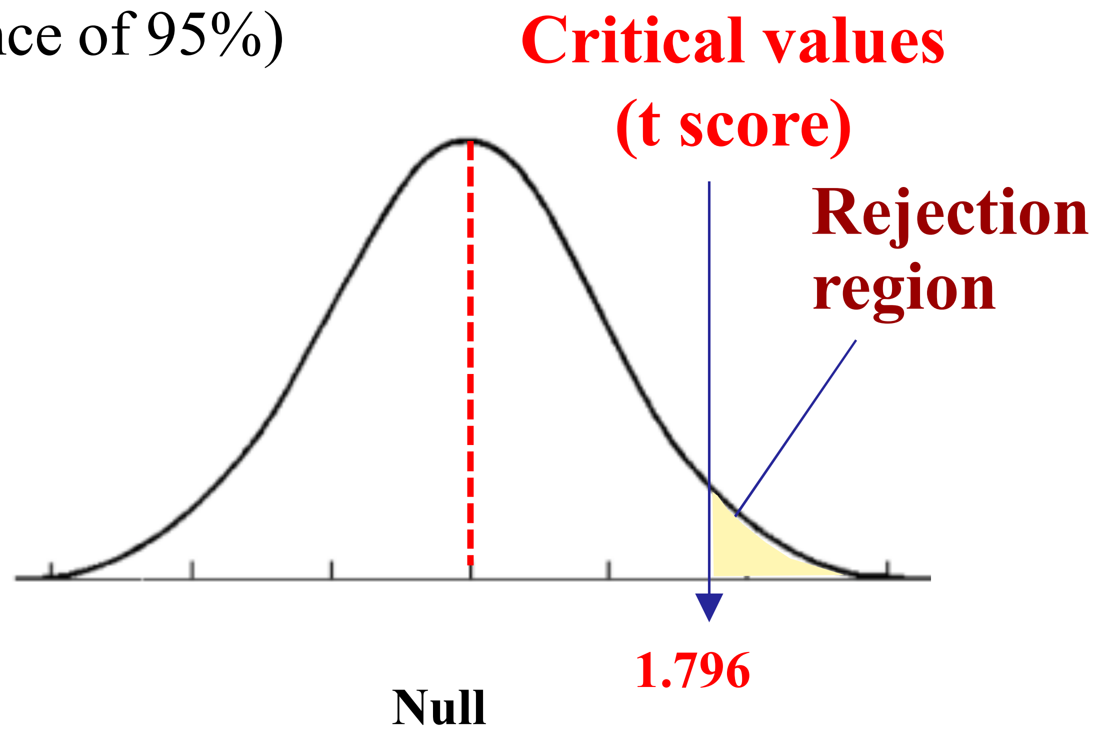
→ t score = 1.796

Example 5: **t test** (two dependent samples)

Step 2 - Set the criteria for a decision

- Consider the level of confidence of 95% → $\alpha = 0.05$
- Locate the **t score** (in the t table for the Student distribution, one-tailed) that represents the **critical value** (for $\alpha = 0.05$ and df = 11)
- Look in the t table for:
  - $\alpha = 0.05$ (α level for a confidence of 95%)
  - df = 11 (degree of freedom = 11)

→ t score = 1.796

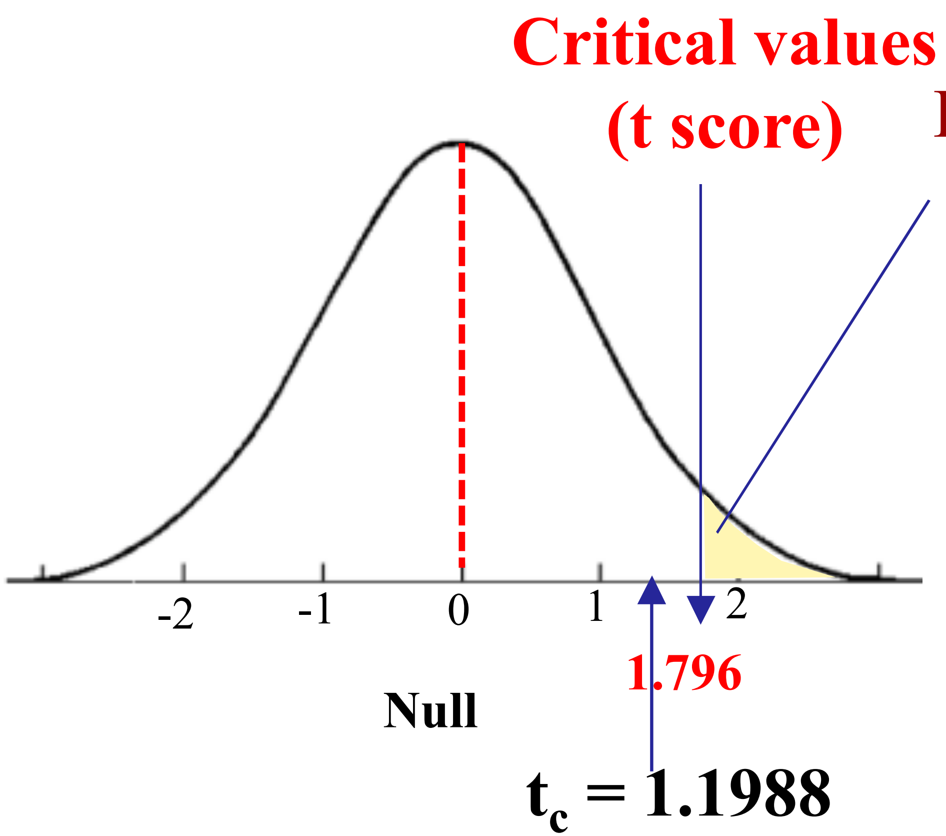


Example 5: **t test** (two dependent samples)

Step 3 - Compute the test statistic

Test statistic: $t_c = \frac{x_D - 0}{\frac{s_D}{\sqrt{n}}}$ **= 1.1988**

Example 5: **t test** (two dependent samples)
Step 4 - Make a decision



The probability of obtaining $\mathbf{t_c = 1.1988}$ is given by the ***P* value**. To obtain ***P* value** look for 1.1988 in the **t table**, for **df = 11**

→ the P value is 0.1279 (P = 12.79%)

As *P* > 5% → **Retain the null hypothesis (fail reaching significance)**