

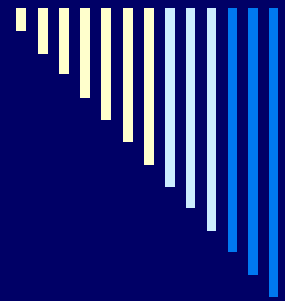
---



ESE

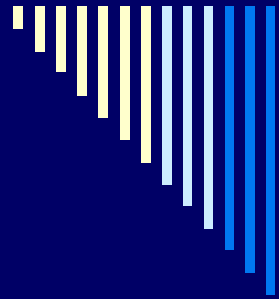
## 10. Data analysis and interpretation

### 10.2.1 Hypotheses Tests



# Data Description, Analysis and Interpretation

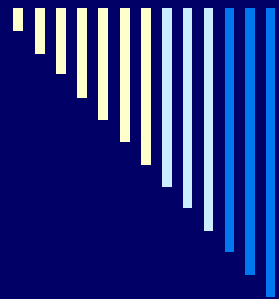
- Presentation purpose
- Background and definitions
- The great table
- Examples of hypothesis tests



# Presentation Purposes (1/3)

Development time	
Method 1	Method 2
5	17

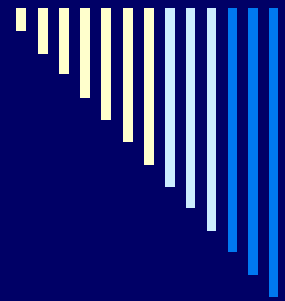
Development time	
Method 1	Method 2
5	17
12	30
50	100



# Presentation Purposes (2/3)

Development Time	
Method A	Method B
23	22
23	44
45	55
43	63
56	21
22	34
21	67
33	76
19	63
21	71
10	51
11	49
22	23
23	45
34	43
21	56
24	10

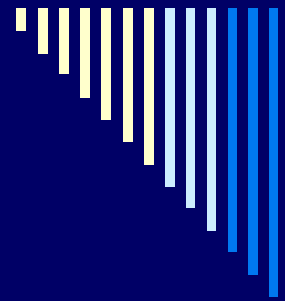




# Presentation Purposes (3/3)

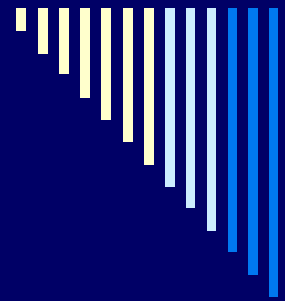
In practice:

- How to run the **right** statistical test.
- Which **conclusion** can be drawn.



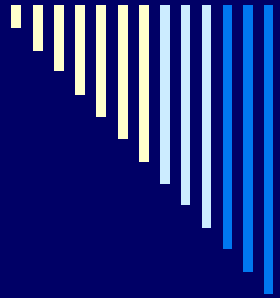
# The Concept of Test <sup>(1/2)</sup>

- Very frequently a test aims to **reject** the hypothesis that two (or more) treatments had the **same** output.



# The Concept of Test <sup>(2/2)</sup>

- Q: How can we judge two (or more) sets of data as **different** or **equivalent**?
  - Q1: In order to consider two sets of data as different, **how much** they have to be different?
  - Q2: **What** do we want to compare?
    - Means? Medians? Variances? ...



# How much they have to be different?



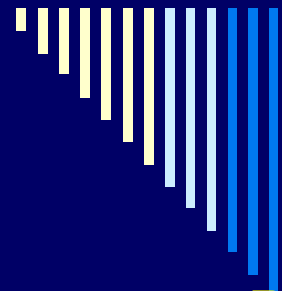
# Recalling the concepts of $H_0$ and $H_1$

- $H_0$ :: the **Null Hypothesis** is used to refer to the state in which the two distributions are **not** significantly **different**.
- $H_1$ :: the **Alternative Hypothesis** is used to refer to the state in which the two distributions are **significantly different**.



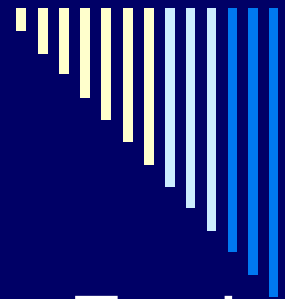
# Errors in evaluating the Null Hypothesis

- Error of **type I**::  $P(\text{type-I-error}) = P(\text{reject } H_0 \mid H_0 \text{ is } \mathbf{true})$
- Error of **type II**::  $P(\text{type-II-error}) = P(\mathbf{NOT} \text{ reject } H_0 \mid H_0 \text{ is } \mathbf{NOT true}), \text{ i.e., } P(\mathbf{NOT} \text{ reject } H_0 \mid H_0 \text{ is } \mathbf{false})$



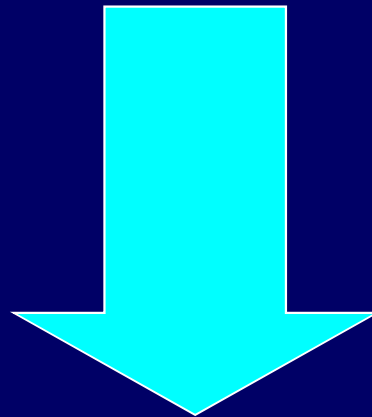
# Error Notations $\alpha$ and $\beta$

- $\alpha :: P(\text{Type I Error})$ ; **Level of significance** of a test (result):  
The probability of a statistical test of incorrectly rejecting the (**true**) null hypothesis.  
The maximum probability with which we are prepared to run the risk of making a type I error.
  - $1-\alpha$ : **Level of confidence** we have on rejection.
- $\beta :: P(\text{Type II Error})$ . The probability of a statistical test of incorrectly accepting the (**false**) null hypothesis.
  - $1-\beta$ : **Power of the statistical test**: the probability of a statistical test of correctly rejecting the (**false**) null hypothesis =  $1-P(\text{Type II Error})$ .

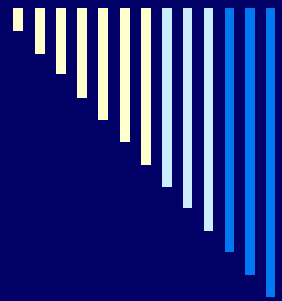


# The Concept of Test

For instance, we want an error probability  $\alpha$  to be not greater than a given value, e.g.:



$$\alpha \leq 0,10$$



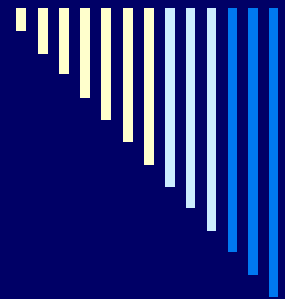
# The Concept of Test

**Warning !!!**

Which is the minimum level of  $\alpha$  required to judge a distribution as different in respect to another?

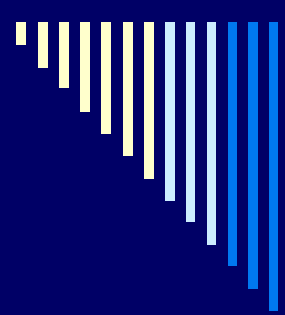
It depends on our level of knowledge about the domain: e.g.,

$$\alpha \leq 0.05$$



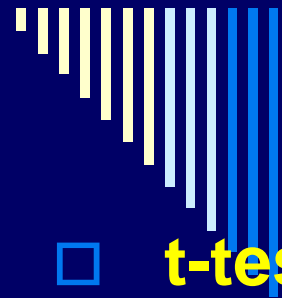
# Normal vs. Not Normal Distribution

- We cannot reject that a distribution can be **considered** as normal only when there is **no one test** able to reject the null hypothesis of similarity in respect to a normal distribution.



# Parametric vs. Not Parametric Tests

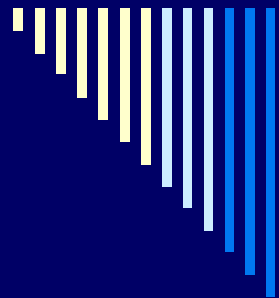
- ❑ Parametric tests can be used only when we cannot reject that **each** distribution object of the study is normal.
- ❑ Non parametric tests **can be used** instead of parametric tests but they are **less powerful!**



# Overview of Tests

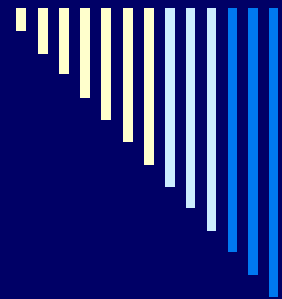
- **t-test**: Often used parametric test to compare two samples means, i.e., the design is one factor with two treatments.
- **Mann-Whitney**: A non-parametric alternative to t-test.
- **F-test**: A parametric test that can be used to compare two sample distributions, in particular their variance.
- **Paired-t-test**: A t-test for paired comparison design.
- **Wilcoxon**: A non-parametric alternative to Paired t-test. It is based on ranks of the samples.
- **Sign-test**: A simple alternative to the Wilcoxon test. It is based on the sign of differences between samples.
-





# Overview of Tests

- **ANOVA**: A family of parametric tests that can be used for designs with more than two levels of a factor. ANOVA test can, for example, be used in the following designs: One factor with more than two treatments, one factor with blocking variables, factorial design, and nested design.
- **Kruskal-Wallis**: A non parametric-test alternative to ANOVA in the case of one factor with more than two treatments.
- **Chi-2**: A family of non-parametric tests that can be used when data are in the form of frequencies.



# The Great Table

Type of experiment	Parametric	Non Parametric
<b>Two Treatments, randomized<sup>1</sup></b>	T-test	Mann-Whitney
	F-test	Chi Square
<b>Two treatments, paired<sup>2</sup></b>	Paired T-test	Wilcoxon
		Sign test
<b>Tree or more tratments</b>	ANOVA	Kruskal-Wallis
		Chi Square



# Paired vs. Randomized Design

1. **Randomized**: subjects, objects, treatments, and their orders are randomly designed. Each subject uses **one treatment on one object**.
  - Be sure of using a randomized technique! Not something vaguely casual!
2. **Paired**: each subject uses **both treatments on each object**.
  - Be sure to balance the **order** and that the **learning effect** is not destructive.