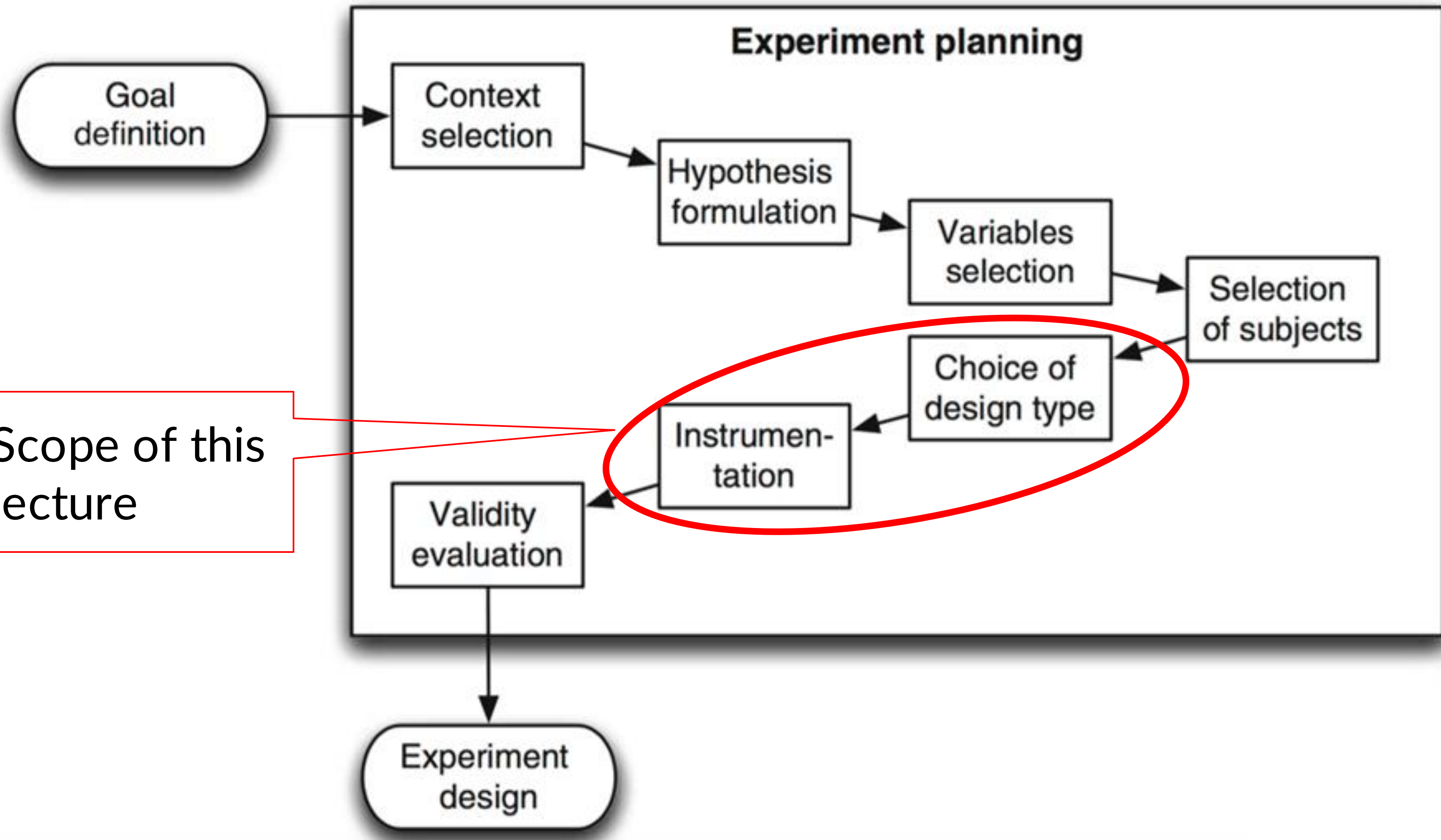
Experiment design (basics)

Ivano Malavolta

VU VRIJE UNIVERSITEIT AMSTERDAM

LOOKING FURTHER

Planning phases



Scope of this lecture

Roadmap

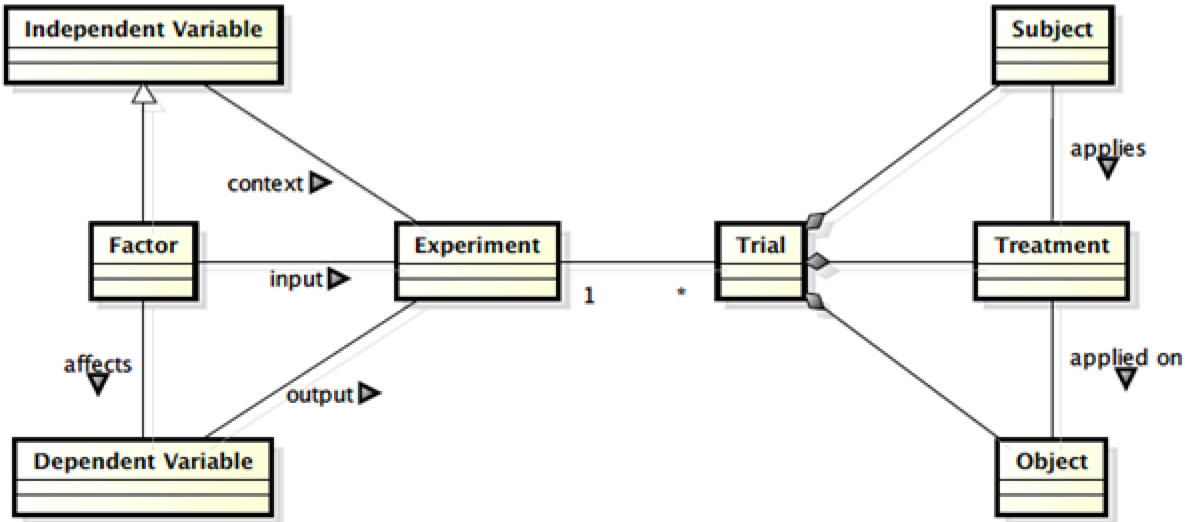
Experiment design

Design principles

Basic design types

Experiment design

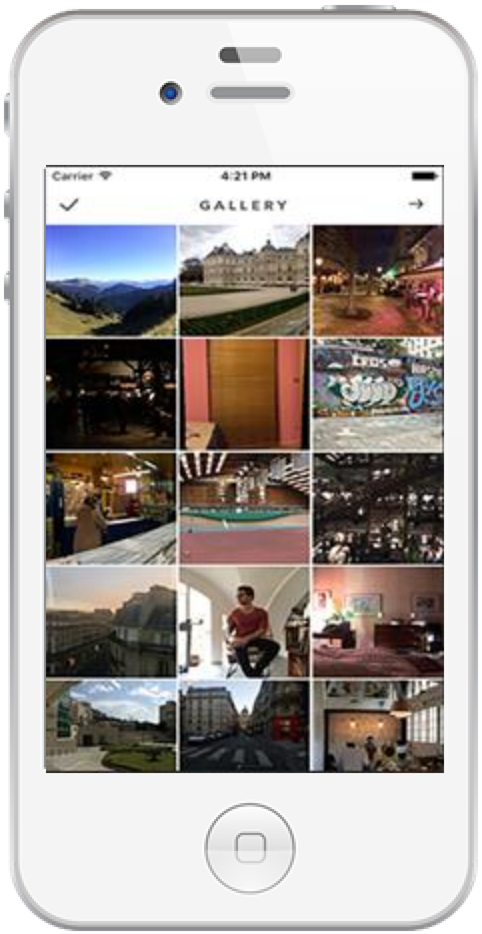
Experiment design: how to organize and execute the trials

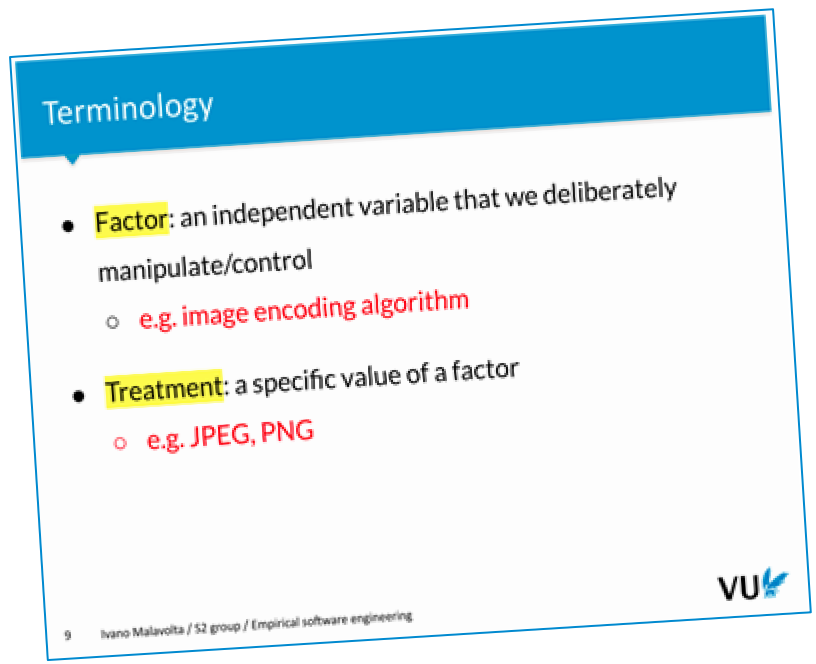


Goal: to determine the set of trials the experiment shall have to make sure that the effect of the treatments is <u>visible</u>

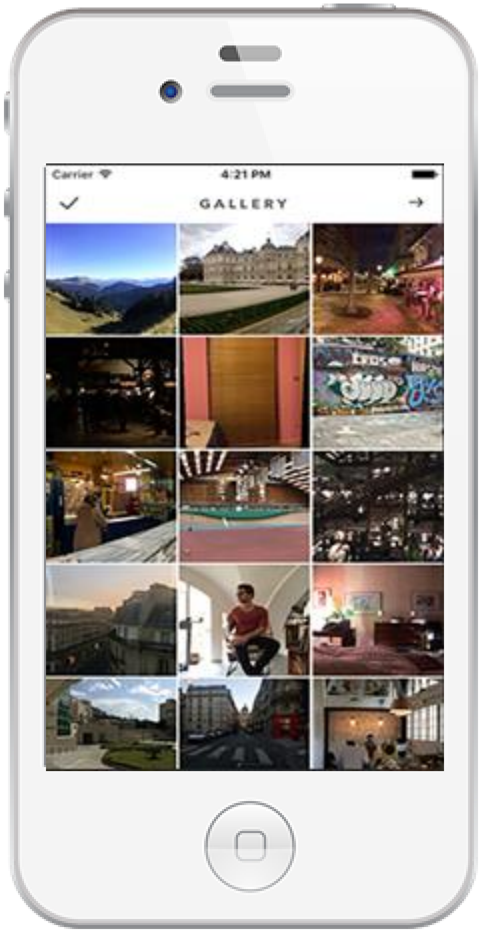
Example

- **How** do we combine subjects and treatments?
    - number of factors and treatments





What can you decide?



- **Factor**: image encoding algorithm
  - other factors?
- **Treatments**:
  - JPEG
  - PNG
- **Subjects**: mobile apps
  - one app vs many apps
- How to “cover” all the relevant combinations of treatments and subjects?

Design principles

- **Randomization**

Randomization

PROBLEM: when executing many trials, the chosen objects, subjects and execution ordering may lead to biased results

SOLUTION: randomize the involved objects and subjects, and the order in which trials are executed

Randomization

- **Aim**: *remove* the effects of a specific non-controlled independent variable

- Group the subjects/objects by that variable and then randomize

Examples:

- We **randomly** choose mobile apps from a dataset
- For each app we **randomly** assign a specific encoding algorithm for its images
- We execute the apps in **random** order

Design principles

- **Randomization**

- **Blocking**

Blocking

PROBLEM: one factor influences our results but we want to mitigate its effects

SOLUTION: split the sample in blocks with same (or similar) level of this factor

Blocking

- The blocks are studied **separately**

eg, Main factor: image encoding algorithm {PNG, JPEG}

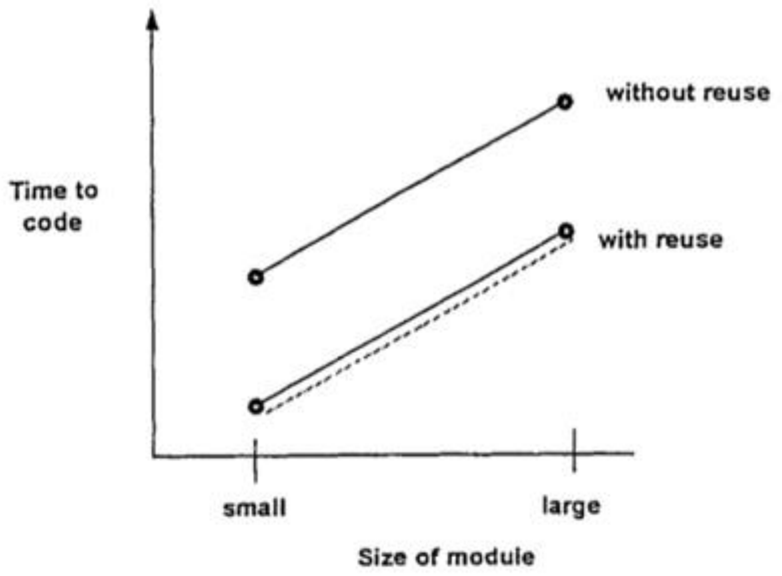
Blocked factor: type of device {low-end, high-end}

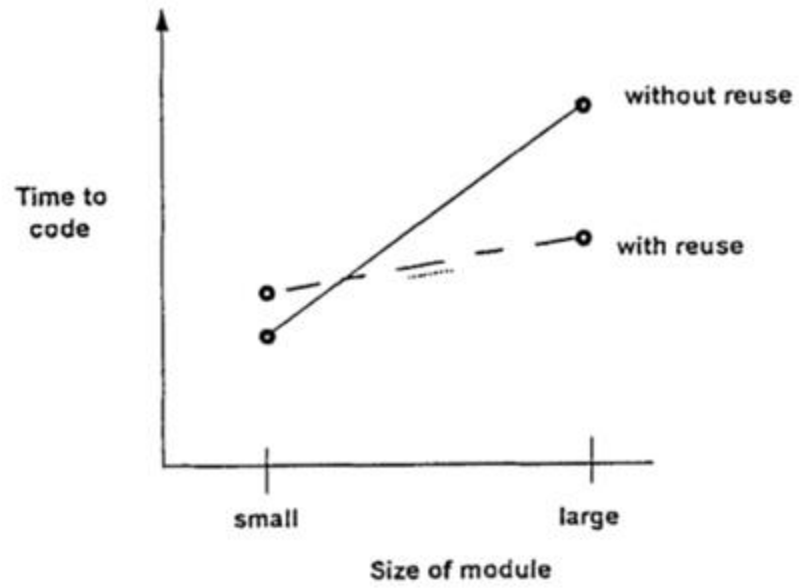
Block 1: run the apps only in low-end devices

Block 2: run the apps only in high-end devices

- We **DO NOT** study the effects between the groups
    - e.g. no conclusions on the correlation between the effects of image encoding and the type of device

Why more than 2 factors?





- {PNG, JPEG} → energy consumption
- {PNG, JPEG} and {single_rendering, par_rendering} → energy consumption

How to choose between factor or block?

If you are deciding between two methods or tools, then you should identify state variables that are likely to affect the results and sample over those variables using blocks to ensure an unbiased assignment of experimental units to the alternative methods or tools.

If you are deciding among methods or tools in a variety of circumstances, then you should identify state variables that define the different circumstances and treat each variable as a factor.

In other words, use blocks to eliminate bias; use factors to distinguish cases or circumstances.

How to choose between factors/blocks/etc?

In the literature, the independent variables correspond to main factors and end up in the research questions

Also co-factors end up in your RQs: these are at the same level of your main factor. You want to investigate how the main factor and the co-factors influence each other

The other factors can be:

- Uncontrolled factors: they are just “small details“ in your experiment, you do not want to investigate on their effect on the dependent variable
- Fixed factors: they are the aspects that you fix in your experiment

e.g., your subjects are Python libraries only, you load your web apps only in Chrome, you use only Wifi connection, etc.

- Blocking factors: they are the factors that you suspect might have an influence on your dependent variables, but you you use them only for “compartmentalizing” your experiment

e.g., the type of device (or OS) since you do not want to directly compare the energy consumed across two different devices since they might have a completely different architecture

Design principles

- **Randomization**

- **Blocking**

- **Balancing**

Balancing

PROBLEM: many statistical analyses are more powerful and simple when performed on balanced data

SOLUTION: consider the same (or similar) number of subjects for each type of treatment

eg, Block 1: 20 apps

Block 2: 20 apps

Basic design types

Basic design types

We can have the following cases:

- 1 factor and 2 treatments (1F-2T)
- 1 factor and >2 treatments (1F-MT)
- 2 factors and 2 treatments (2F-2T)
- >2 factors, each one with >=2 treatments (MF-MT)

Basic

1 factor and 2 treatments

We assume to have 1 dependent variable P

Notation:

- μ_i : dependent variable mean for treatment i
  - μ_i = avg(P)
- y_{ij}: j-th measure of the dependent variable for treatment i

Example:

- We are seeing whether different image encoding algorithms impact energy consumption of mobile apps
- **Factor**: encoding algorithm
- **Treatments**:
  - PNG
  - JPEG
- **Dependent variable**: consumed energy during common usage scenarios

1F-2T: fully randomized design

- Each treatment is randomly assigned to the experimental objects

- Same number of objects for each treatment (balancing)

Object (Application)	Treatment 1 (PNG)	Treatment 2 (JPEG)
1	X	
2		X
3		X
4	X	

Examples of hypotheses:

$H_0 : \mu_1 = \mu_2$
$H_a : \mu_1 \mathrel{!=} \mu_2$ or $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$

Analyses:

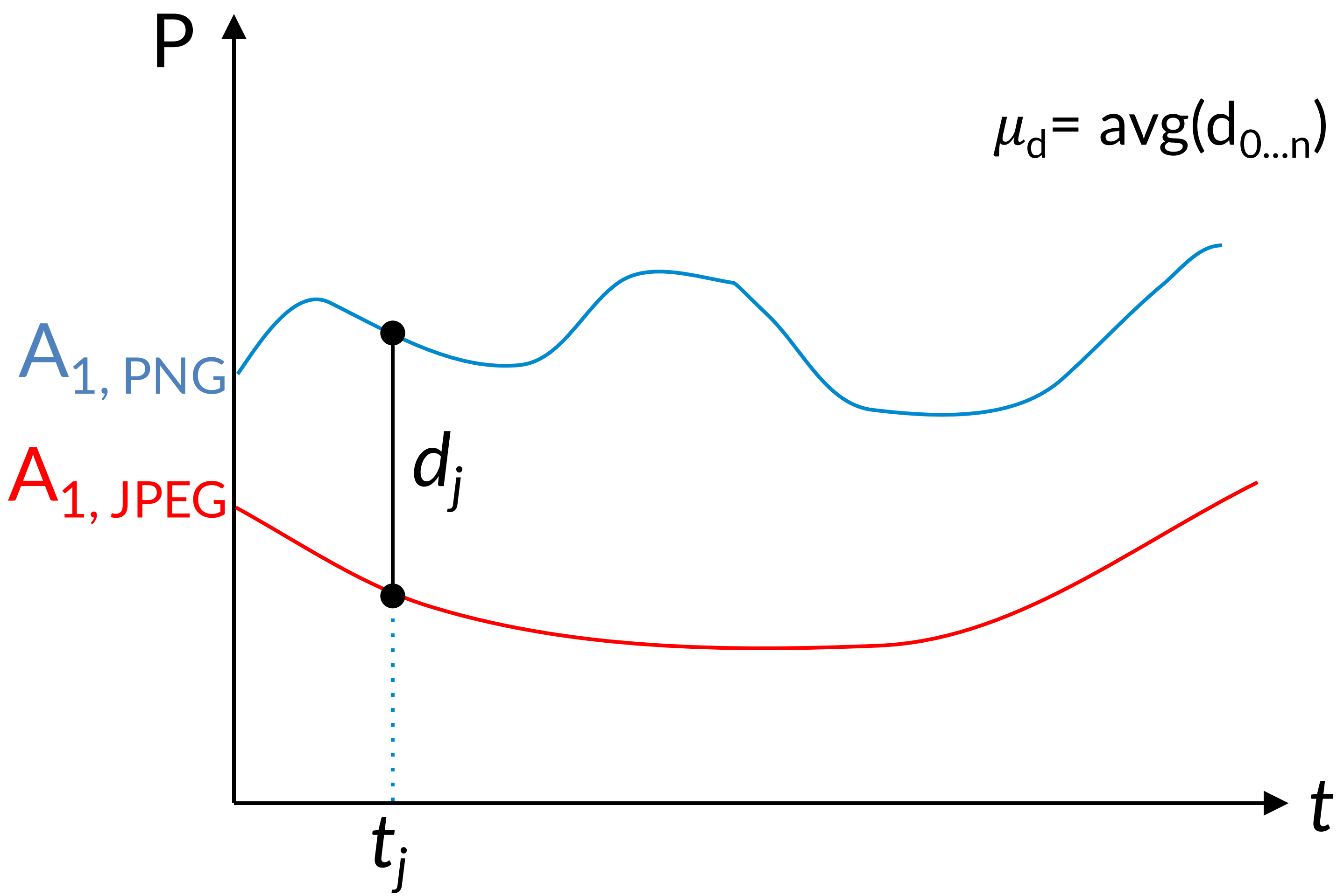
- t-test (unpaired)
- Mann-Whitney test

1F-2T: paired comparison design

- Each treatment is applied on each object (crossover design)
- The order of the treatments is random

Object (Application)	Treatment 1 (PNG)	Treatment 2 (JPEG)
1	1st	2nd
2	2nd	1st
3	2nd	1st
4	1st	2nd

1F-2T: paired comparison design



1F-2T: paired comparison design

Example of hypotheses:

$H_0 : \mu_d = 0$
$H_a : \mu_d \mathrel{!=} 0$ *or* $\mu_d > 0$ *or* $\mu_d < 0$

Analyses:

- Paired t-test
- Sign test
- Wilcoxon

How to choose between the two design?

Object (Application)	Treatment 1 (PNG)	Treatment 2 (JPEG)
1	X	
2		X
3		X
4	X	

Object (Application)	Treatment 1 (PNG)	Treatment 2 (JPEG)
1	1st	2nd
2	2nd	1st
3	2nd	1st
4	1st	2nd

1 factor and >2 treatments

In this case the factor can have more than 2 values

Example:

- **Factor**: encoding algorithms
- **Treatments**:
  - PNG
  - JPEG
  - TIFF
- **Dependent variable**: consumed energy during common usage scenarios

1F-MT: fully randomized design

- Each treatment is randomly assigned to the experimental objects

- Same number of objects per each treatment (balancing)

Object (Application)	Treatment 1 (PNG)	Treatment 2 (JPEG)	Treatment 3 (TIFF)
1	X		
2		X	
3	X		
4		X	
5			X
6			X

Example of hypotheses:

$H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_t$
$H_a : \mu_i \mathrel{!=} \mu_j$ for at least one pair of (i,j)

Analyses:

- ANOVA (ANalysis Of VAriance)
- Kruskal-Wallis

1F-MT: Randomized complete block design

- Each treatment is applied on each object (crossover design)
- The order of the treatments is random

Object (Application)	Treatment 1 (PNG)	Treatment 2 (JPEG)	Treatment 3 (TIFF)
1	1st	3rd	2nd
2	3rd	1st	2nd
3	2nd	3rd	1st
4	2nd	1st	3rd
5	3rd	2nd	1st
6	1st	2nd	3rd

Example of hypotheses:

$H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_t$
$H_a : \mu_i != \mu_j$ for at least one pair of (i,j)

Analyses:

- ANOVA
- Kruskal-Wallis
- Repeated Measures ANOVA

Acknowledgements

Some contents of lecture extracted from:

- Giuseppe Procaccianti's lectures at VU