Vall d'Hebron
Institut de Recerca

DESIGN OF EXPERIMENTS AND ANOVA

Estadística Biomèdica Avançada

Ricardo Gonzalo Sanz
ricardo.gonzalo@vhir.org

11/07/2016

Vall d'Hebron Hospital
Generalitat de Catalunya Departament de Salut
Institut Català de la Salut
UAB Universitat Autònoma de Barcelona
Generalitat de Catalunya Departament d'Economia i Coneixement
Institució CERCA Centres de Recerca de Catalunya
Instituto de Salud Carlos III

Vall d'Hebron
Institut de Recerca

Table of contents

Vall d'Hebron
Institut de Recerca

Table of contents

Vall d'Hebron
Institut de Recerca

Experimental design should be a mandatory step in every experiment

- Experimental design is a structured, organized method for **determining the relationship** between the different **factors** affecting an experimental process, and the **output** of that process.

Sir Ronald A. Fisher

Father of modern Mathematical Statistics and Developer of Experimental Design and ANOVA

"To consult the statistician after an experiment is finished, is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of. "

Vall d'Hebron
Institut de Recerca

Why are many life scientists so adverse to thinking about design?

It is common to think that time spent designing experiments would be better spent actually doing experiments

Some myths arise in biologists and statistical fields about that.

- *"It does not matter how you collect your data, there will always be a statistical 'fix' that will allow you to analyse them"*
- *"If you collect lots of data something interesting will come out, and you'll be able to detect even very subtle effects"*
- *"Garbage In, Garbage Out"*
- *"In science, as in life: more haste, less speed"*

Components of an experiment?

- An experiment is characterized by the ***treatments*** and ***experimental units*** to be used, the ***method*** treatments are assigned to units and the ***responses*** are measured.

- Treatments, units, and assignment method specify the *experimental design.*

- What about analysis? Analysis is not part of design, but SHOULD be consider during planning.

Vall d'Hebron
Institut de Recerca

Some important definitions:

- **Treatment**: The set of circumstances created for the experiment in response to the research hypothesis (15º + 50%).
- **Factor**: A set of treatments grouped together logically (Tª, humidity).
- **Level**: The several categories of a factor (15º, 25º, 37º).
- **Responses**: outcomes that we observe after applying a treatment to an experimental unit.

Types of variability that play role in an experiment:

- **Planned systematic variability**: These are the differences in response between treatments applied.

- **Noise variability:** random noise. Differences between two consecutives measures. We cannot avoid that.

- **Systematic variability not planned:** Produce a systematic variation in the results. A priori the reason is not known. It can be avoided with the *randomization* and the *local control*.

Vall d'Hebron
Institut de Recerca

Important steps to define before begin the experiment:

1. Establish the main **objectives** of the experiment. Avoid collateral problems
2. Identify all the **noise** sources: Treatment, experimental errors,…
3. **Allocate** each experimental unit which each treatment
4. Clarify the **type of response** expected in each treatment
5. Determinate the **number** of individuals in each group
6. Run a **pilot study**

Vall d'Hebron
Institut de Recerca

Important steps to define before begin the experiment:

1. Establish the main **objectives** of the experiment. Avoid collateral problems
2. Identify all the **noise** sources: Treatment, experimental errors,…
3. **Allocate** each experimental unit which each treatment
4. Clarify the **type of response** expected in each treatment
5. Determinate the **number** of individuals in each group
6. Run a **pilot study**

Vall d'Hebron
Institut de Recerca

Important steps to define before begin the experiment:

1. Establish the main **objectives** of the experiment. Avoid collateral problems
2. Identify all the **noise** sources: Treatment, experimental errors,…
3. **Allocate** each experimental unit which each treatment
4. Clarify the **type of response** expected in each treatment
5. Determinate the **number** of individuals in each group
6. Run a **pilot study**

Vall d'Hebron
Institut de Recerca

Important steps to define before begin the experiment:

1. Establish the main **objectives** of the experiment. Avoid collateral problems
2. Identify all the **noise** sources: Treatment, experimental errors,…
3. **Allocate** each experimental unit which each treatment
4. Clarify the **type of response** expected in each treatment
5. Determinate the **number** of individuals in each group
6. Run a **pilot study**
7. How the **data** will be statistically analysed.

Why Experimental Design?

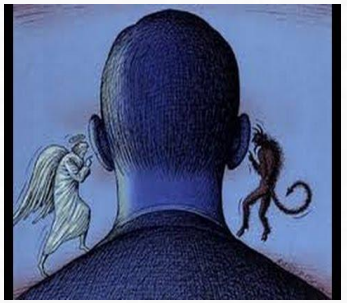
- We can design experiments to **minimize any bias** in the comparison
- We can design experiments so that the **error in the comparison is small**
- Very Important: **We are in control of experiments**, and having that control allow us to make stronger inferences about the nature of differences that we see in the experiment.

Vall d'Hebron
Institut de Recerca

Why Experimental Design?

- To **save** money and time
  - Less probability of bad results (mistakes,…)
  - Only analyze the samples strictly necessary (less reagents, time,…)
  - Only collect the data you need to answer the objectives

- **Ethical** issues
  - Treatment applied to the necessary animals
  - Careful when applying the treatments



1. Basic Ideas of Experimental Design

Vall d'Hebron
Institut de Recerca

RNA-seq Data: Challenges in and Recommendations for Experimental Design and Analysis

Alexander G. Williams,[1] Sean Thomas,[1,2] Stacia K. Wyman,[1] and Alisha K. Holloway[1,2]

Current Protocols in Human Genetics 11.13.1-11.13.2(
Published online October 2014 in Wiley Online Library
DOI: 10.1002/0471142905.hg1113s83

Clinical Chemistry 55:10
1816–1823 (2009)

Molecular Diagnostics and Genetics

Design and Optimization of Reverse-Transcription Quantitative PCR Experiments

Ales Tichopad,[1,2*] Rob Kitchen,[3] Irmgard Riedmaier,[1] Christiane Becker,[1] Anders Ståhlberg,[2,4] and Mikael Kubista[2,5]

Optimal designs for 2-color microarray experiments

P. S. SANCHEZ*, G. F. V. GLONEK
Discipline of Statistics, School of Mathematical Sciences, The University of Adelaide, SA 5005, Australia
penny.sanchez@adelaide.edu.au

Statistical aspects of quantitative real-time PCR experiment design

Robert R. Kitchen [a], Mikael Kubista [b,c], Ales Tichopad [c,d,*]

[a] *School of Physics, University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, UK*
[b] *TATAA Biocenter, Odinsgatan 28, 411 03 Göteborg, Sw*
[c] *Institute of Biotechnology AS CR, Lb Building, 2nd Floo*
[d] *Technical University Munich, Physiology Weihenstepha*

Statistical Design of Quantitative Mass Spectrometry-Based Proteomic Experiments

Ann L. Oberg[‡] and Olga Vitek[*‡]
Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, 200 First Street SW, Rochester, Minnesota 55905, and Department of Statistics and Department of Computer Science, Purdue University, 250 North University Street, West Lafayette, Indiana 47907

J. Proteome Res., **2009**, *8* (5), pp 2144–2156
DOI: 10.1021/pr8010099
Publication Date (Web): February 17, 2009

Vall d'Hebron
Institut de Recerca

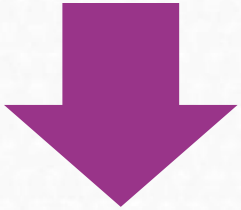
A good experimental design should…

- Avoid systematic error: e.g. samples from one group processed with instrument A and samples from the other group processed with instrument B.
- Be precise: try to maintain the random error as low as possible
- Allow estimation of error: enough replicates in each treatment
- Have broad validity: our experimental units should reflect the population about which we wish to draw inference

Vall d'Hebron
Institut de Recerca

A good experimental design should…

- Avoid systematic error: e.g. samples from one group processed with instrument A and samples from the other group processed with instrument B.
- Be precise: try to maintain the random error as low as possible
- Allow estimation of error: enough replicates in each treatment
- Have broad validity: our experimental units should reflect the population about which we wish to draw inference



Randomization | Replication | Local Control

Vall d'Hebron
Institut de Recerca

Table of contents

2. Experimental Design conditions. Randomization

- The method for assigning treatments to units involves *randomization ("any individual experimental subject has the same chance as any other individual of finding itself in each experimental group")*
- It is one of the most important elements of a well-designed experiment
- Made valid most of the statistical analysis usually performed

Haphazard is not randomized

e.g. 4 treatments to be assigned to 16 units

1. Use sixteen identical slips of paper, 4 marked with A, 4 with B, and so on to D. Put the slips of paper into a basket and mix them thoroughly. For each unit, we draw a slip of paper from the basket and use the treatment marked on the slip
2. Treatment A is assigned to the first four units we have encounter, treatment B to next four units, and so on.

2. Experimental Design conditions. Randomization

Randomization against confounding

The effect of the treatment cannot be distinguished from another

e.g. Consider a new drug treatment for coronary artery disease. 2 treatments to compare: this drug treatment with bypass surgery. We have 100 patients in our pool of volunteers; they need to be assigned to the two treatments. We then measure five-year survival as a response.

What sort of trouble can happen if we fail to randomize? Bypass surgery is a major operation, and patients with severe disease may not be strong enough to survive the operation:

- stronger patients to surgery and the weaker patients to the drug therapy.

This confounds strength of the patient with treatment differences. The drug therapy would likely have a lower survival rate because it is getting the weakest patients, even if the drug therapy is every bit as good as the surgery.

2. Experimental Design conditions. Randomization

Randomization against <u>confounding</u>

e.g. Cont.

Patient 1 to 100

Surgery

Drug therapy

2. Experimental Design conditions. Randomization

Saying “randomly assign…” is sometimes easier to say than to do, especially in complex designs.

Some tools may help

- Research Randomizer
  http://www.randomizer.org/
- Interactive Statistical Calculation pages

  http://statpages.org/

  (look por “Experimental design”)

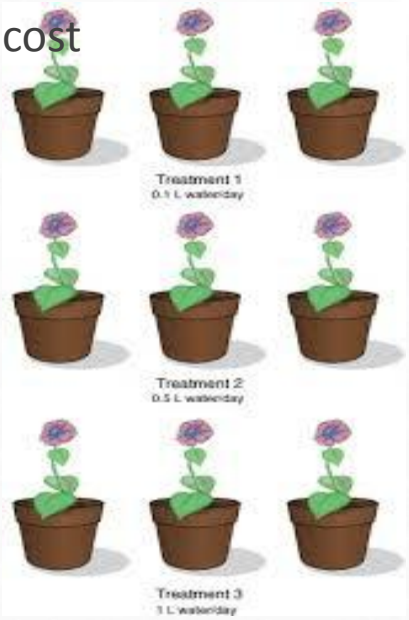
2. Experimental Design conditions. Replication

- It is the basis of all experimental design
- It is the repetition of the basic experiment with another experimental units

How many replicates I need?

the more replicates we have, the more confident we can be that differences between groups are real and not simply due to chance effects

More replicates increase in time/money cost



- To be in mind:
    1. To Know the variability of the technology used
    2. Previous works with similar technology
    3. Directly correlated with the precision of the experiment

2. Experimental Design conditions. Replication

- Which type of replicates?
  - Technical
  - Biological

- **Statistical power**: is the probability that a particular experiment will detect a difference, assuming that there really is a difference to be detected.
  1. Effect size
  2. amount of random variation
  3. number of replicates

  There are computer programs to calculate it (you must provide effect sizes and estimates of variation) (http://homepage.stat.uiowa.edu/~rlenth/Power/)

- **Pooling** samples: To pool or not to pool?

2. Experimental Design conditions. Local Control

- When the experimental units are not homogeneous or the process to analyze them neither are (Kits lot numbers, batch,…)

We are not interested in to find out the differences between the levels of the blocks

- It transforms **systematic variability not planned** in **planned systematic variability**.

- Differences among blocks could hide differences among treatments.

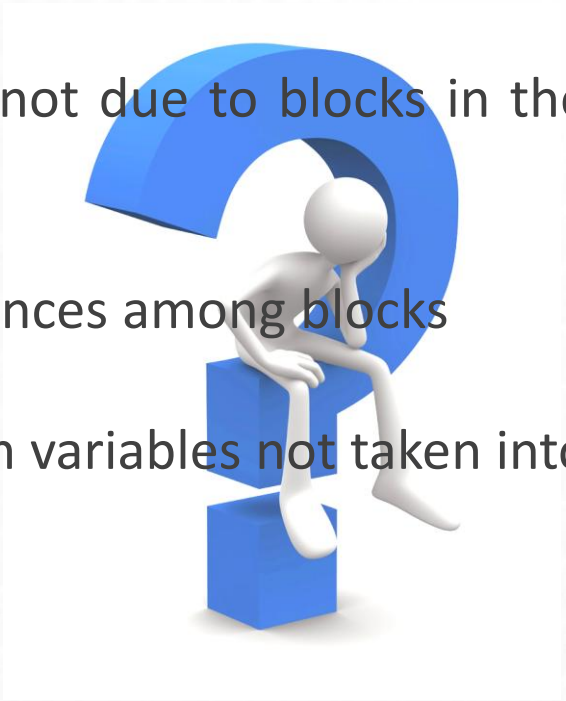
2. Experimental Design conditions. Local Control

Sample	Treatment	Sex	Batch
1	A	Male	1
2	A	Male	1
3	A	Male	1
4	A	Male	1
5	B	Female	2
6	B	Female	2
7	B	Female	2
8	B	Female	2

Sample	Treatment	Sex	Batch
1	A	Male	1
2	A	Female	2
3	A	Male	2
4	A	Female	1
5	B	Male	1
6	B	Female	2
7	B	Male	2
8	B	Female	1

Local control or randomize?

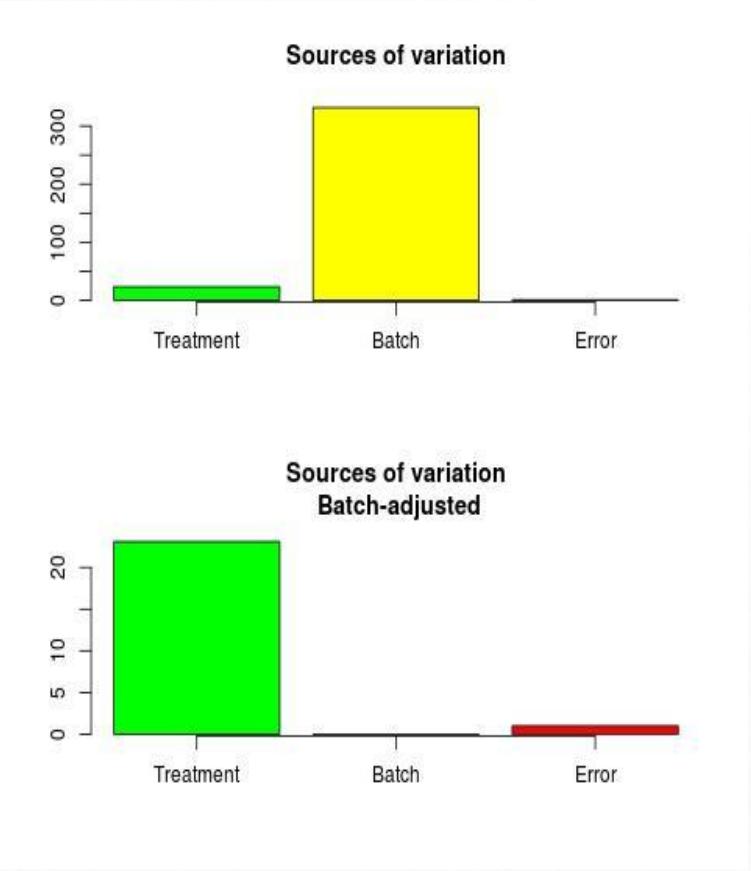
- Local control assure you that differences are not due to blocks in the sample
- Local control eliminate the noise due to differences among blocks
- Randomization is good for balance effects from variables not taken into account from the beginning.



"Block what you can, randomize what you cannot" (George Box, 1978)

2. Experimental Design conditions

Example of a bad experimental design (or perhaps an absence of...)



Vall d'Hebron
Institut de Recerca

Table of contents

3. Experimental design types

A key point in any experiment is the way that experimental units are allocated to treatments

- It must be chosen so that **random variability** is as **small** as possible.
- It must be chosen so that ***the best local control*** is achieved.
- It implicitly defines the **analysis model**, so it must be chosen so that the analysis can be performed and validity conditions hold.

More common are:

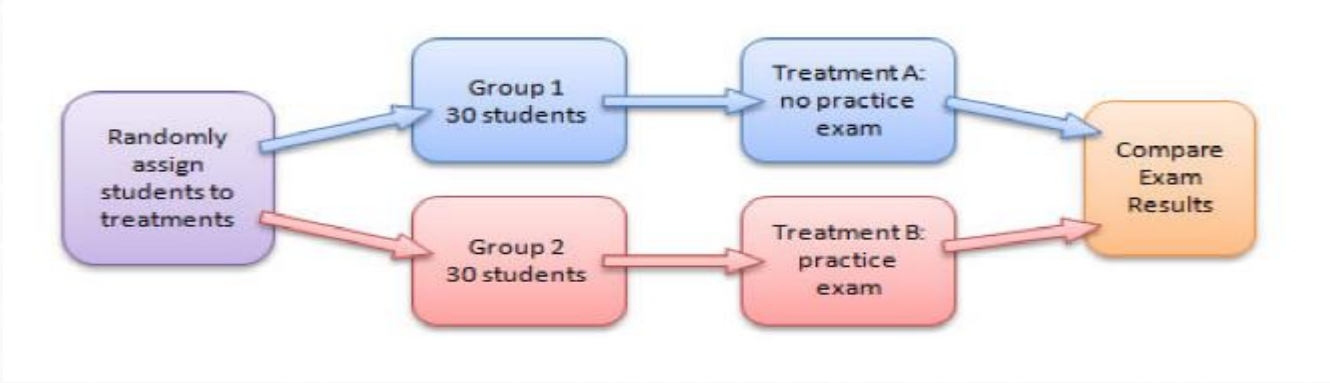
- Completely randomized design (CRD)
- Randomized Block design (RBD)
- Factorial designs
- Repeated measures designs

3. Experimental design types. COMPLETELY RANDOMIZED DESIGN

Vall d'Hebron Institut de Recerca

- Simplest of all designs
- Uses randomization and replication.
  - ✓Treatments are allocated at random to experimental units over the entire experimental material

Response = mean + treatment effect + error



Vall d'Hebron Institut de Recerca

EXAMPLE (File = Strains.txt):

A group of mice was inoculated with five strains of malaria organisms to observe the number of days each mouse survived so that a treatment strategy could be developed. Following table gives the number of days each mouse survived. Six mice were inoculated with each strain. *Find whether the effect of different strains of malaria organisms is the same.*

	1	2	3	4	5	6
Strain A	10	9	8	11	12	10
Strain B	6	7	8	6	5	5
Strain C	8	7	9	10	9	9
Strain D	11	10	12	13	11	12
Strain E	6	5	7	6	5	5

Vall d'Hebron Institut de Recerca

EXAMPLE:

Set up the null hypothesis as follows: $H_0 : \mu_A = \mu_B = \mu_C = \mu_D = \mu_E$

Load and look the data:

	Days	Strain
1	10	StrainA
2	9	StrainA
3	8	StrainA
4	11	StrainA
5	12	StrainA
6	10	StrainA
7	6	StrainB
8	7	StrainB
9	8	StrainB
10	6	StrainB
11	5	StrainB
12	5	StrainB
13	8	StrainC
14	7	StrainC
15	9	StrainC
16	10	StrainC
17	9	StrainC
18	9	StrainC
19	11	StrainD
20	10	StrainD
21	12	StrainD
22	13	StrainD
23	11	StrainD
24	12	StrainD
25	6	StrainE
26	5	StrainE
27	7	StrainE
28	6	StrainE
29	5	StrainE
30	5	StrainE

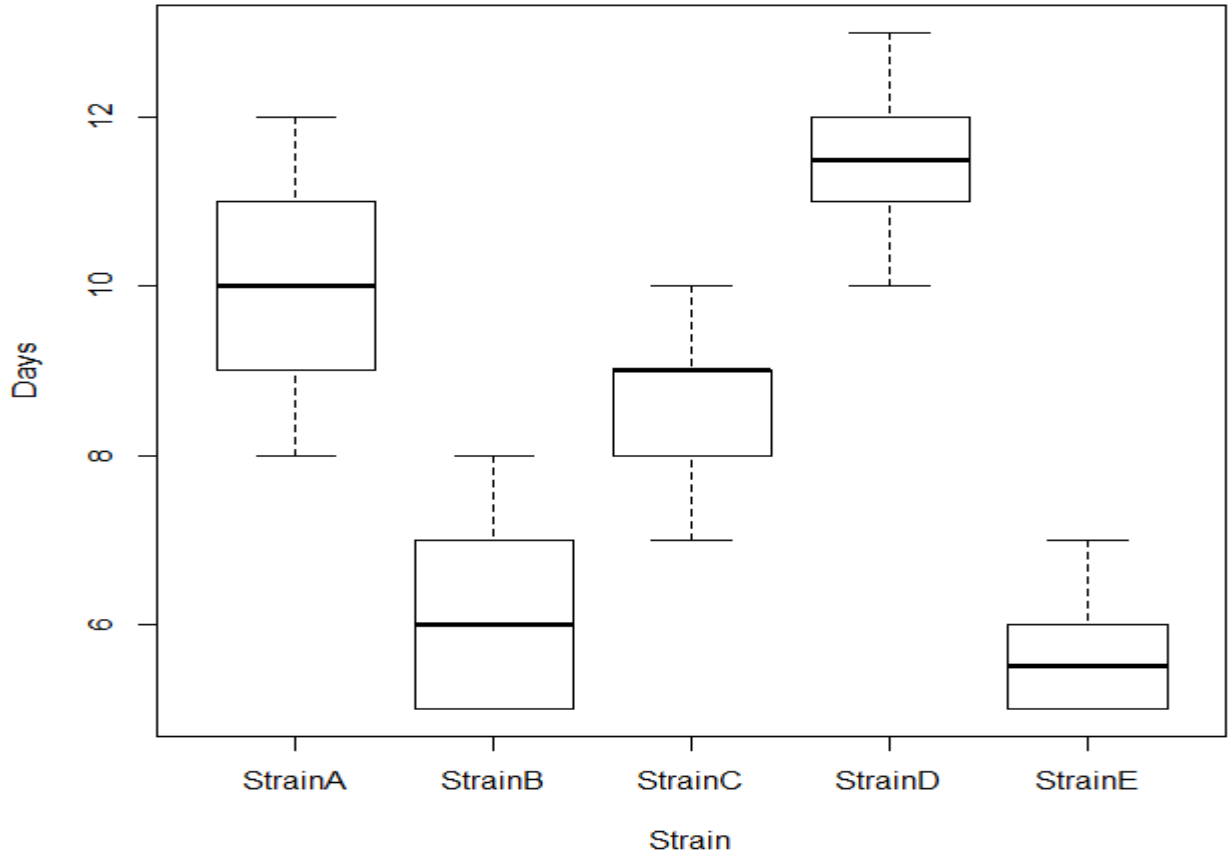
Numerical summaries:

	mean	sd	IQR	cv	0%	25%	50%	75%	100%	data:n
StrainA	10.000000	1.4142136	1.50	0.14142136	8	9.25	10.0	10.75	12	6
StrainB	6.166667	1.1690452	1.50	0.18957490	5	5.25	6.0	6.75	8	6
StrainC	8.666667	1.0327956	0.75	0.11916872	7	8.25	9.0	9.00	10	6
StrainD	11.500000	1.0488088	1.00	0.09120077	10	11.00	11.5	12.00	13	6
StrainE	5.666667	0.8164966	1.00	0.14408763	5	5.00	5.5	6.00	7	6

Vall d'Hebron
Institut de Recerca

EXAMPLE:

Graphical summaries



Vall d'Hebron
Institut de Recerca

EXAMPLE:

```
> AnovaModel.1 <- aov(Days ~ Strain, data=Strain)

> summary(AnovaModel.1)
             Df Sum Sq Mean Sq F value   Pr(>F)
Strain        4  148.2   37.05   29.88 3.39e-09 ***
Residuals    25   31.0    1.24
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> with(Strain, numSummary(Days, groups=Strain, statistics=c("mean", "sd")))
              mean        sd data:n
StrainA 10.000000 1.4142136      6
StrainB  6.166667 1.1690452      6
StrainC  8.666667 1.0327956      6
StrainD 11.500000 1.0488088      6
StrainE  5.666667 0.8164966      6
```

3. Experimental design types. RANDOMIZED BLOCK DESIGN

Vall d'Hebron Institut de Recerca

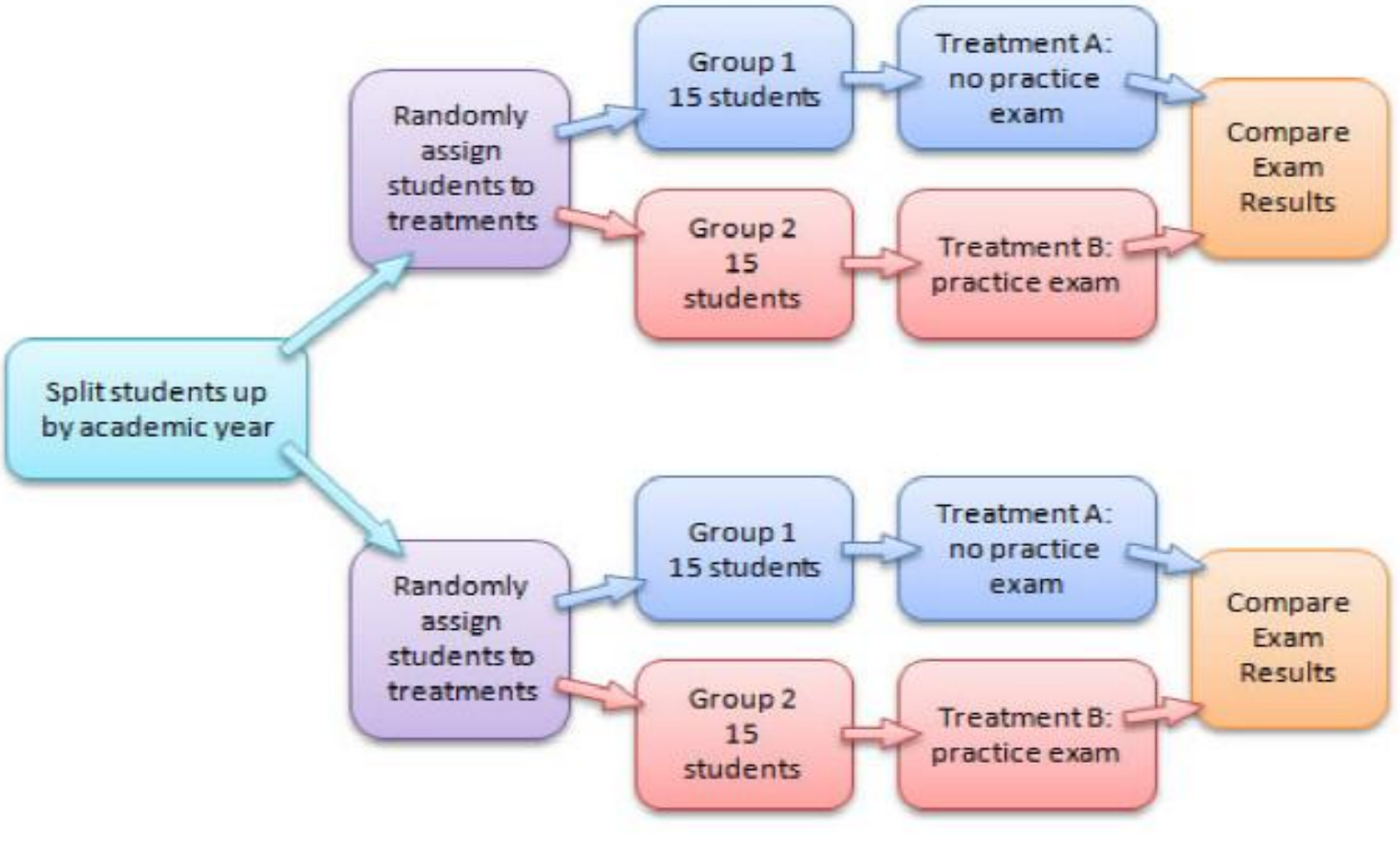
- CRD becomes less informative if the experimental material is not homogenous.
- Blocking may be used to divide the whole experimental material into homogeneous strata or sub-groups known as blocks.
- Blocking to “remove” the effect of nuisance factors
- Then, the experimental units are randomly assigned treatments.

Response = mean + treatment effect + block effect + error

3. Experimental design types. RANDOMIZED BLOCK DESIGN

Vall d'Hebron
Institut de Recerca

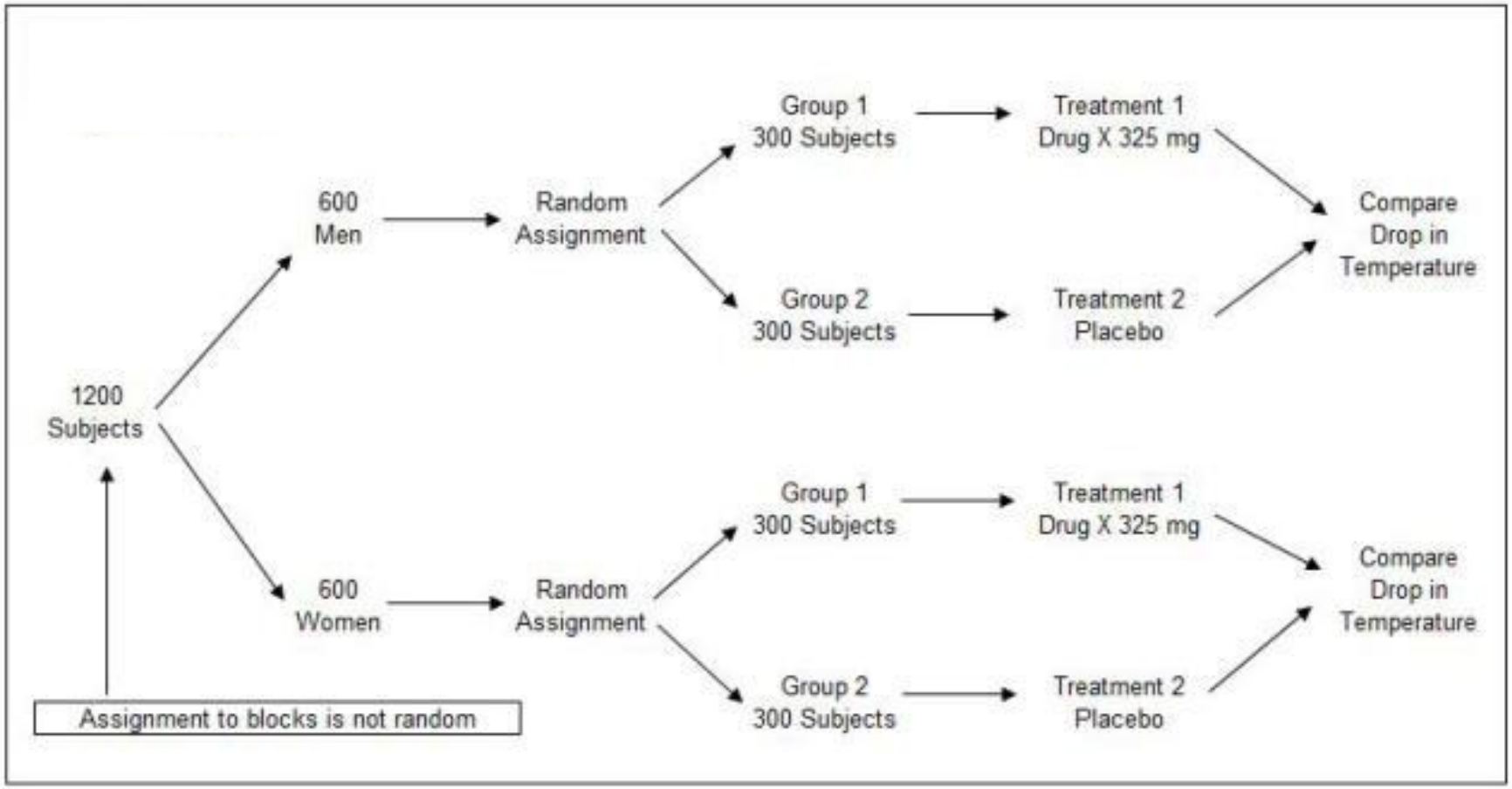
RBD example 1.



3. Experimental design types. RANDOMIZED BLOCK DESIGN

Vall d'Hebron
Institut de Recerca

RBD example 2.

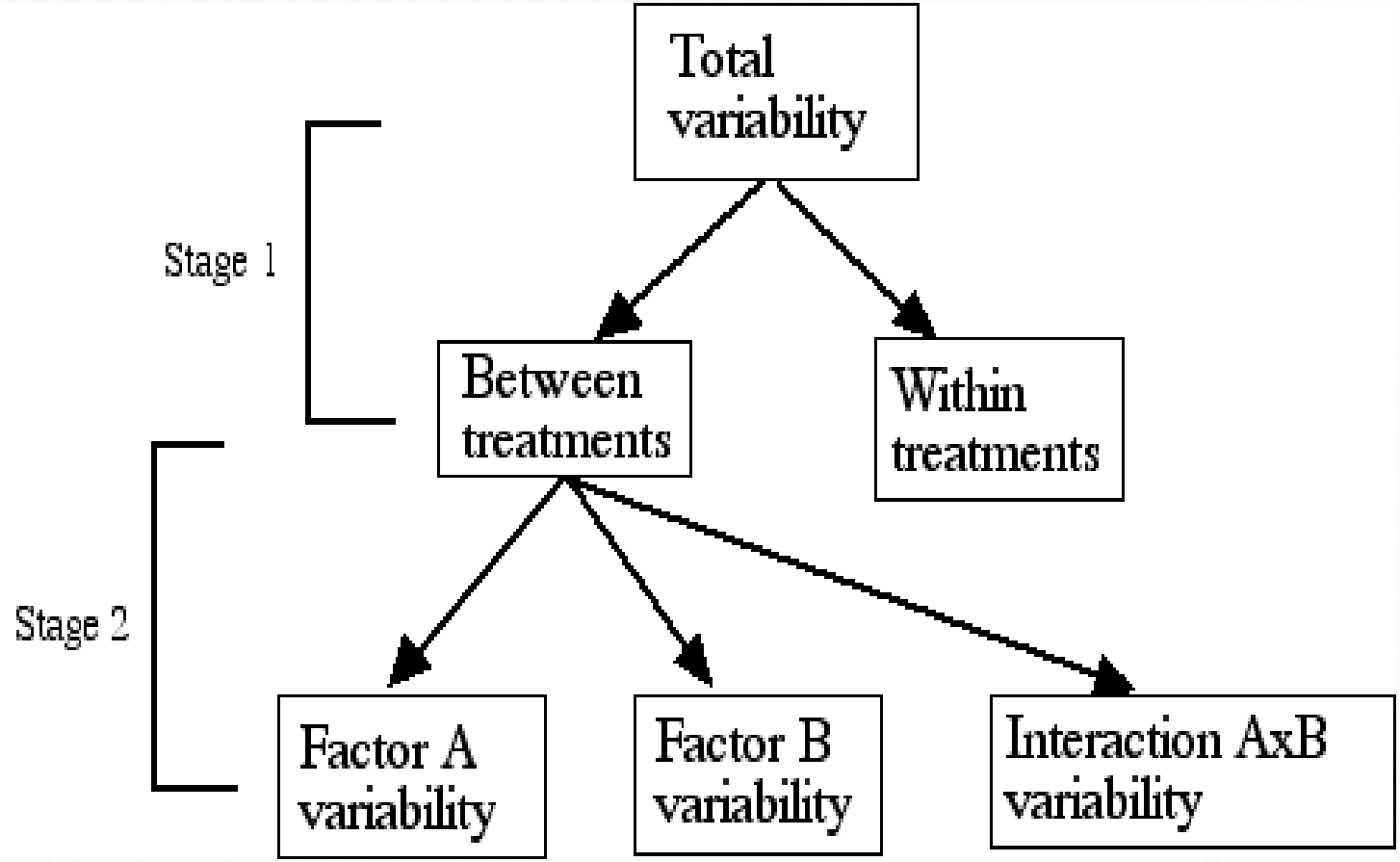


3. Experimental design types. FACTORIAL DESIGN

- Effects of several factors of variation are studied simultaneously.
- The treatments are all the combinations of different factors under study.
- The effects of each of the factors and the *interaction* effects, which are the variations in the effect of one factor as a result to different levels of other factors, are studied.

Response = mean+ treat-1 effect + treat-2 effect +treat-1:2-Interaction + error

3. Experimental design types. FACTORIAL DESIGN



3. Experimental design types. REPEATED MEASURES

When more than one measure is taken on each experimental unit one has a within subjects design:

- Measures on the same individual are correlated
- New source of variability to be accounted for:
  variability within subjects
- Same possibilities as with other designs
- With an extra source of variation (“time”)

3. Experimental design types. REPEATED MEASURES

Vall d'Hebron
Institut de Recerca

Repeated measures Example:

A study on plasma inorganic phosphate after oral glucose tolerance test (OGTT) was performed on 33 individuals.

– Two groups: Control (1) and obese (2)

–Phosphate levels were measured before test (time=0) and 0.5, 1.5, 2, 3, 4 and 5 hours after test.

DatasetWithin1

	GROUP	ID	T.0	T.0.5	T.1	T.1.5	T.2	T.3	T.4	T.5
1	2	1	4.3	3.3	3.0	2.6	2.2	2.5	3.4	4.4
2	2	2	3.7	2.6	2.6	1.9	2.9	3.2	3.1	3.9
3	2	3	4.0	4.1	3.1	2.3	2.9	3.1	3.9	4.0
4	2	4	3.6	3.0	2.2	2.8	2.9	3.9	3.8	4.0
5	2	5	4.1	3.8	2.1	3.0	3.6	3.4	3.6	3.7
6	2	6	3.8	2.2	2.0	2.6	3.8	3.6	3.0	3.5
7	2	7	3.8	3.0	2.4	2.5	3.1	3.4	3.5	3.7
8	2	8	4.4	3.9	2.8	2.1	3.6	3.8	4.0	3.9
9	2	9	5.0	4.0	3.4	3.4	3.3	3.6	4.0	4.3
10	2	10	3.7	3.1	2.9	2.2	1.5	2.3	2.7	2.8
11	2	11	3.7	2.6	2.6	2.3	2.9	2.2	3.1	3.9
12	2	12	4.4	3.7	3.1	3.2	3.7	4.3	3.9	4.8
13	2	13	4.7	3.1	3.2	3.3	3.2	4.2	3.7	4.3
14	1	14	4.3	3.3	3.0	2.6	2.2	2.5	2.4	3.4
15	1	15	5.0	4.9	4.1	3.7	3.7	4.1	4.7	4.9
16	1	16	4.6	4.4	3.9	3.9	3.7	4.2	4.8	5.0
17	1	17	4.3	3.9	3.1	3.1	3.1	3.1	3.6	4.0
18	1	18	3.1	3.1	3.3	2.6	2.6	1.9	2.3	2.7
19	1	19	4.8	5.0	2.9	2.8	2.2	3.1	3.5	3.6
20	1	20	3.7	3.1	3.3	2.8	2.9	3.6	4.3	4.4
21	1	21	5.4	4.7	3.9	4.1	2.8	3.7	3.5	3.7
22	1	22	3.0	2.5	2.3	2.2	2.1	2.6	3.2	3.5
23	1	23	4.9	5.0	4.1	3.7	3.7	4.1	4.7	4.9
24	1	24	4.8	4.3	4.7	4.6	4.7	3.7	3.6	3.9
25	1	25	4.4	4.2	4.2	3.4	3.5	3.4	3.8	4.0
26	1	26	4.9	4.3	4.0	4.0	3.3	4.1	4.2	4.3
27	1	27	5.1	4.1	4.6	4.1	3.4	4.2	4.4	4.9
28	1	28	4.8	4.6	4.6	4.4	4.1	4.0	3.8	3.8
29	1	29	4.2	3.5	3.8	3.6	3.3	3.1	3.5	3.9
30	1	30	6.6	6.1	5.2	4.1	4.3	3.8	4.2	4.8

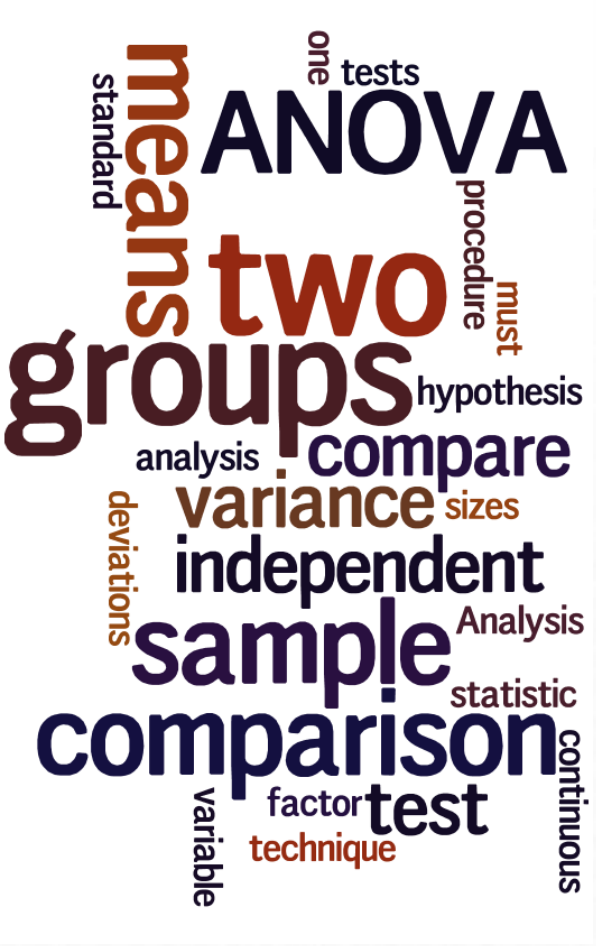
Vall d'Hebron
Institut de Recerca

Table of contents

4. From to sample t test to one factor ANOVA

Until now, we were comparing two samples...but what happens it we have more than two samples o two treatments to compare?

- If we have more than two groups we will use the analysis of variance. Student's test is only valid for two groups



Vall d'Hebron
Institut de Recerca

Why not could we use Student's test?

To know which of the drugs is the best one, one could think to perform the following comparison using a t test:

- Drug 1 vs drug 2 (α=5%)
- Drug 1 vs drug 3 (α=5%)
- Drug 2 vs drug 3 (α=5%)

We could answer the question? **NO**

The new error type I for the whole contrast is not 5%...the new one would be: $1-(1-\alpha)^3 \longrightarrow 1-(1-0.05)^3=\mathbf{0.1426}$

Would be easier to accept the null hypothesis when it was wrong (more false positives)

4. What does ANOVA do?

ANOVA tests the following hypotheses:

H_0: The means of all the groups are equal

No effects among the different treatments

Another way to represent it: $H_0: \quad \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$

$$H_0: \quad \mu_1 = \mu_2 = \cdots = \mu_a$$

H_1: Not all the means are equal

4. What does ANOVA do?

ANOVA tests the following hypotheses:

H_1: Not all the means are equal

If we reject H_0 we only could say that there is a difference among the groups, but we couldn't say between which of them

Can follow up with "multiple comparisons"

4. How ANOVA works?

If we use ANOVA to compare means….why is it called “analysis of variance”?

H_0: The means of all the groups are equal
H_1: Not all the means are equal

ANOVA analyzes if there are differences between two variances to deduce if the are differences between the means

ANOVA partitions the total variability in the sample data into two component parts.

- Due to treatments: variation BETWEEN groups
- Due to the observations: variation WITHIN groups

4. How ANOVA works? Types of variability analyzed by ANOVA

Sample	Drug1	Drug2	Drug3
1	4	7	9
2	2	6	12
3	6	5	6
4	6	7	11
5	5	6	10
6	6	4	11
7	2	7	9
8	6	5	10
Mean	4.6	5.9	9.8

Variation among all the participants represents variation due to "treatment" and "individual differences"

Variations among the conditions represents variation due to "treatment effects"

Variation among participants within each condition represents "individual differences"

$$SS_{Total} = SS_{Effect} + SS_{Error}$$

4. How ANOVA works? Types of variability analyzed by ANOVA

Vall d'Hebron
Institut de Recerca

- Variation BETWEEN groups: for each data value looks at the difference between its group mean and the overall mean

- Variation WITHIN groups: for each data value looks at the difference between that value and the mean of its group

4. How ANOVA works? Types of variability analyzed by ANOVA

Differences will be found between groups if (plainly speaking):

Variation BETWEEN groups: > **Variation WITHIN groups**

- Strictly speaking to compare the two sum of squares they must be standardized.

↓

dividing each sum of squares by their *degrees of freedom*

↓

They are correlated with the # of terms of the sum

4. How ANOVA works? Types of variability analyzed by ANOVA

Vall d'Hebron Institut de Recerca

Source	DF	SS	MS	F
Treatments	$g-1$	SS_{Trt}	$SS_{\text{Trt}}/(g-1)$	MS_{Trt}/MS_E
Error	$N-g$	SS_E	$SS_E/(N-g)$	

```
> summary(AnovaModel.1)
            Df Sum Sq Mean Sq F value   Pr(>F)
TRX          2 114.25   57.12   22.11 6.79e-06 ***
Residuals   21  54.25    2.58
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4. How ANOVA works? Types of variability analyzed by ANOVA

The Table of ANOVA

Source	DF	SS	MS	F
Treatments	$g-1$	SS_{Trt}	$SS_{\text{Trt}}/(g-1)$	MS_{Trt}/MS_E
Error	$N-g$	SS_E	$SS_E/(N-g)$	

The ANOVA F-statistic is a ratio of the **Between group variation divided by the Within group variation**.

$$F = \frac{Between}{Within} = \frac{MSG}{MSE}$$

✓A **large F** is evidence against H_0, since it indicates that there are more difference between groups than within groups.

Vall d'Hebron Institut de Recerca

EXAMPLE

Let a gene be suspected to have some connection with blood cancer. There are four stages of blood cancer. For treating the cancer is crucial to identify the gene in the first three stages. The experiment is repeated six times. Find whether there is any difference in mean expression values in the three mRNA stages:

	1	2	3	4	5	6
mRNA Stage1	95	98	100	105	85	88
mRNA Stage2	94	92	78	88	92	91
mRNA Stage3	72	88	82	73	75	77

Vall d'Hebron Institut de Recerca

EXAMPLE

Set up the null hypothesis as follows: $H_0 : \mu_{mRNA1} = \mu_{mRNA2} = \mu_{mRNA3}$

We could calculate the rows totals:

	1	2	3	4	5	6	Total	Suma ^{2}
mRNA Stage1	95	98	100	105	85	88	**571**	**54623**
mRNA Stage2	94	92	78	88	92	91	**535**	**47873**
mRNA Stage3	72	88	82	73	75	77	**467**	**36535**
						TOTAL	**1573**	**139031**

And the summary of the data

	Count	Sum	Average	Variance
mRNA Stage1	6	571	95.17	56.57
mRNA Stage2	6	535	89.17	33.77
mRNA Stage3	6	467	77.83	37.37

Vall d'Hebron Institut de Recerca

EXAMPLE

And finally the ANOVA's Table:

Source of Variation	SS	df	MS	F	p-Value
Between Groups (Trx)	929.78	2	464.89	10.92	0.00118
Within Groups (Error)	638.5	15	42.57		
TOTAL	1568.28	17			

Since the p-value is 0.001183 $< \alpha = 0.05$, we reject the claim that the mean values of the mRNA samples are equal. Therefore, based on the given sample we conclude that the three stages have different mean expression values.

Vall d'Hebron Institut de Recerca

EXAMPLE (with RCommander. File=SBP.txt)

Which of the three drugs is the best to reduce the SBP? Three drugs are randomly assigned to 24 patients (same characteristics), and SBP is monitored after a month. Response variable is the difference between final and initial SBP value.

Sample	Drug1	Drug2	Drug3
1	4	7	9
2	2	6	12
3	6	5	6
4	6	7	11
5	5	6	10
6	6	4	11
7	2	7	9
8	6	5	10

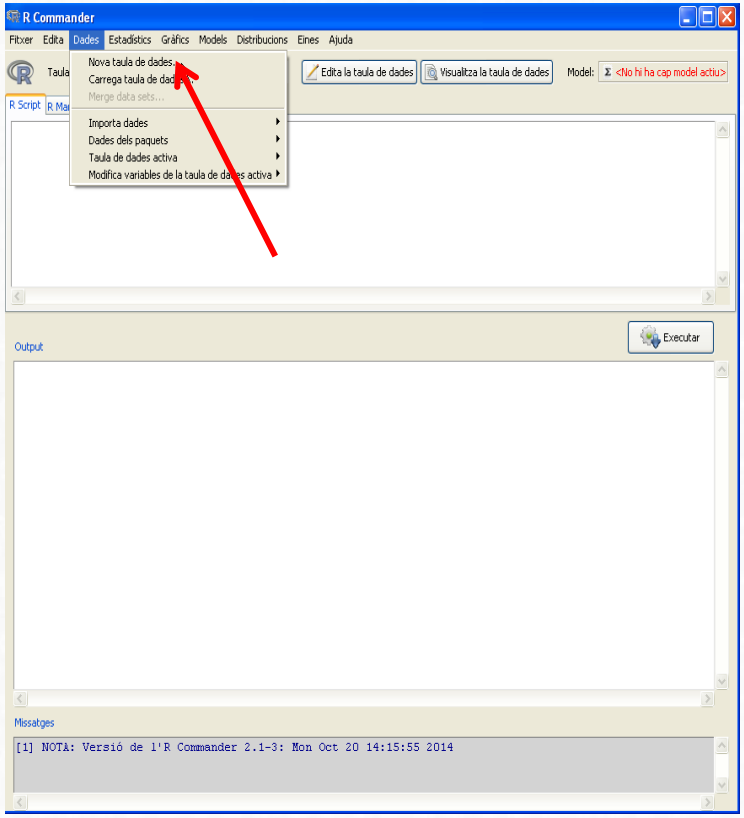
→ One factor

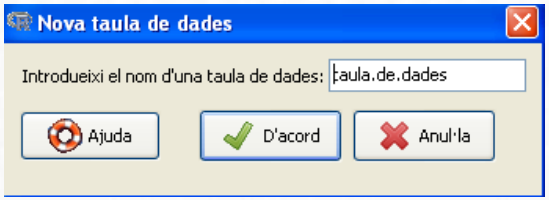
Three levels

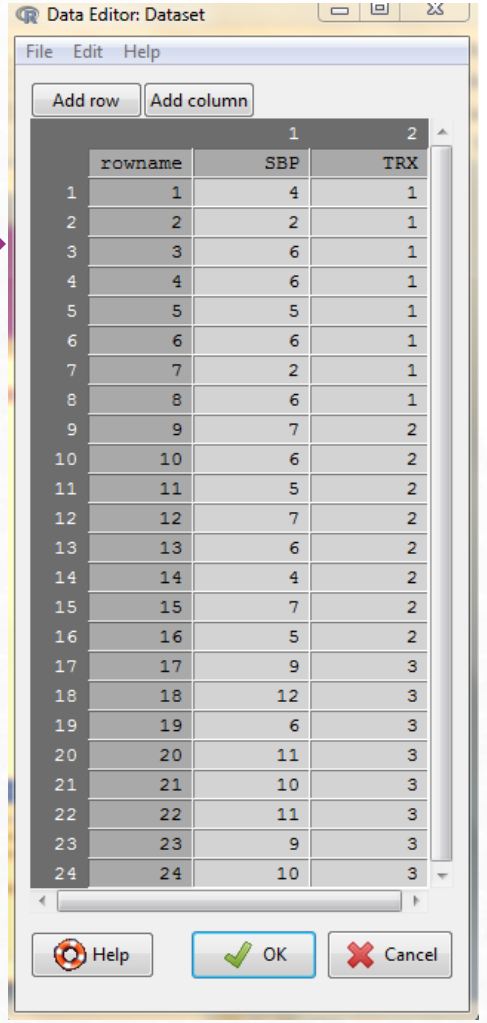
4. Analyze data from COMPLETELY RANDOMIZED DESIGN

Vall d'Hebron
Institut de Recerca

Always is good to “look” graphically the data. Let’s try with R-commander:





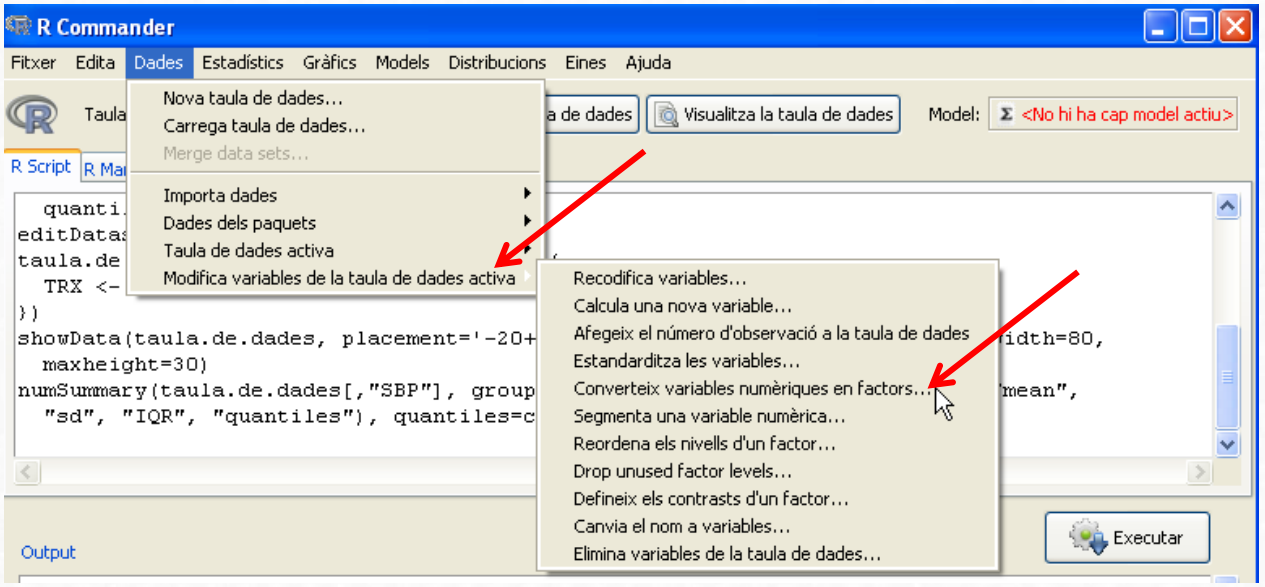


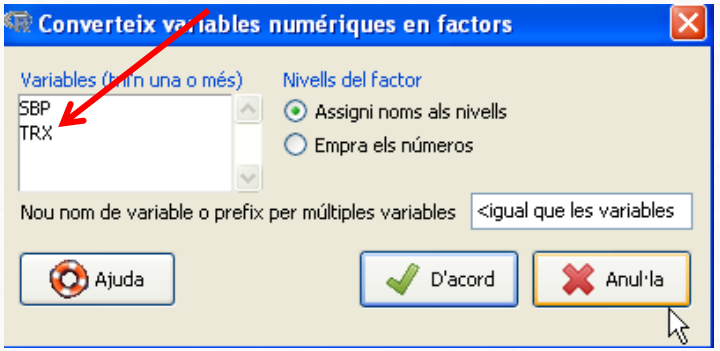
	rowname	SBP	TRX
		1	2
1	1	4	1
2	2	2	1
3	3	6	1
4	4	6	1
5	5	5	1
6	6	6	1
7	7	2	1
8	8	6	1
9	9	7	2
10	10	6	2
11	11	5	2
12	12	7	2
13	13	6	2
14	14	4	2
15	15	7	2
16	16	5	2
17	17	9	3
18	18	12	3
19	19	6	3
20	20	11	3
21	21	10	3
22	22	11	3
23	23	9	3
24	24	10	3

4. Analyze data from COMPLETELY RANDOMIZED DESIGN

Vall d'Hebron
Institut de Recerca

Recode Trx (numeric) variable to a factor:

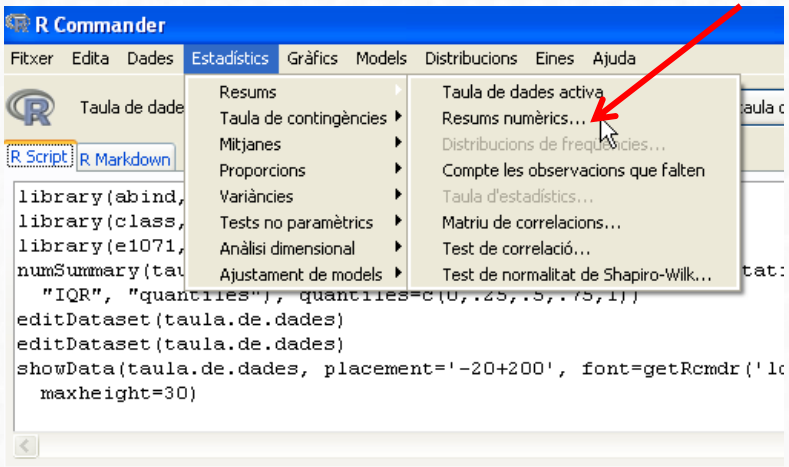


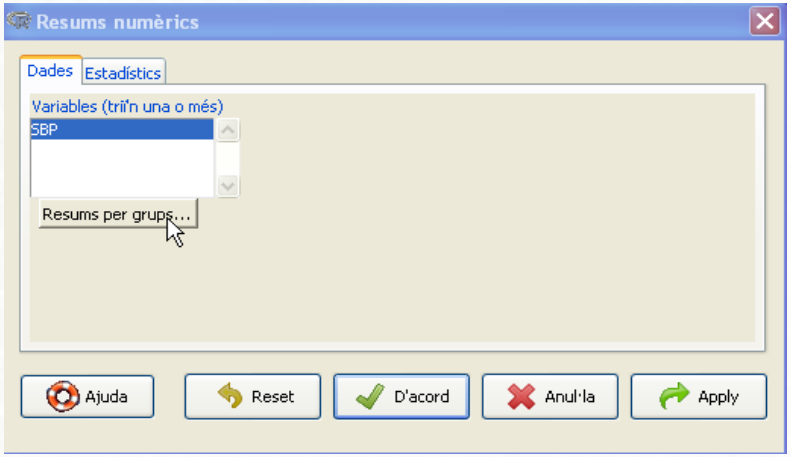


4. Analyze data from COMPLETELY RANDOMIZED DESIGN

Vall d'Hebron
Institut de Recerca

Calculate the means:

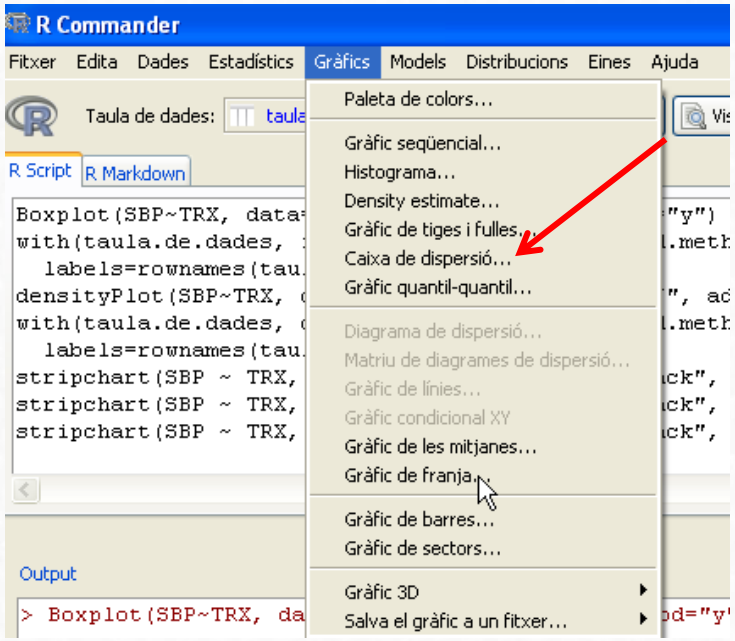


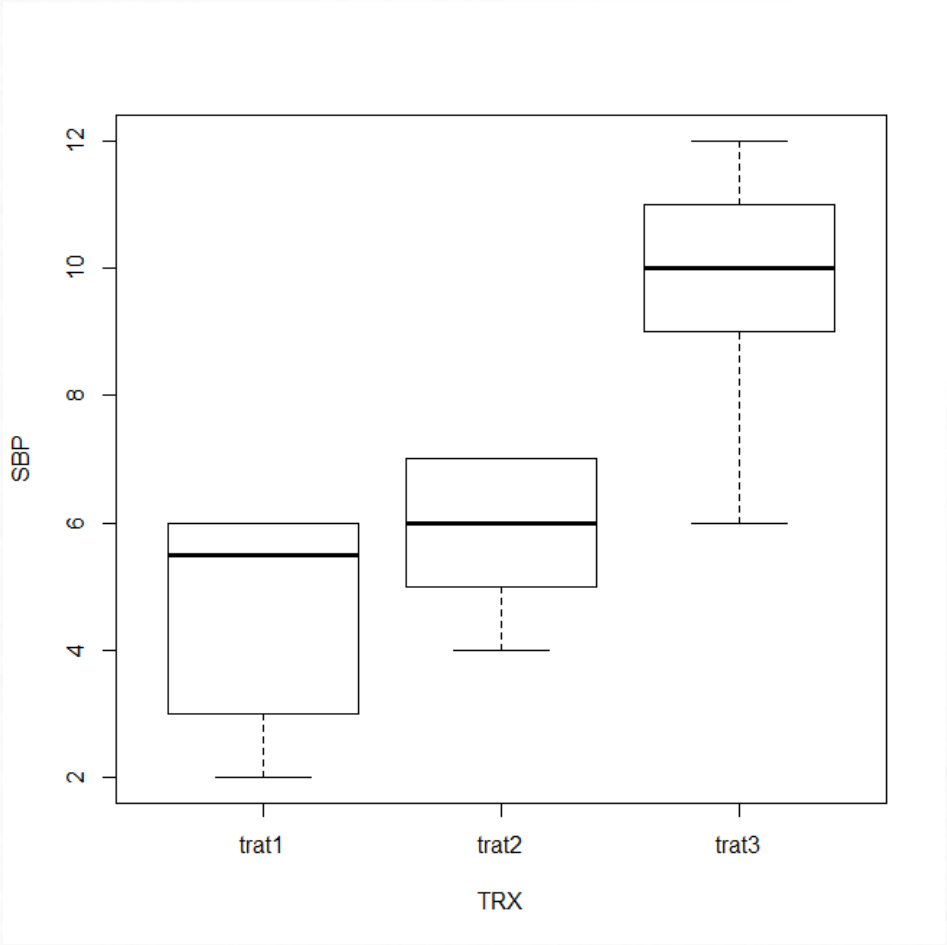


	mean	sd	IQR	0%	25%	50%	75%	100%	data:n
1	4.625	1.767767	2.5	2	3.5	5.5	6	6	8
2	5.875	1.125992	2.0	4	5.0	6.0	7	7	8
3	9.750	1.832251	2.0	6	9.0	10.0	11	12	8

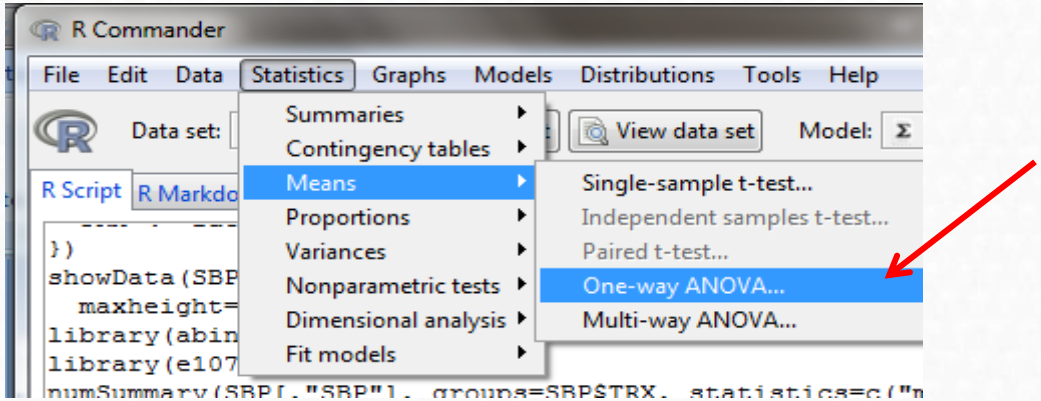
4. Analyze data from COMPLETELY RANDOMIZED DESIGN

Let's see graphically:





4. Analyze data from COMPLETELY RANDOMIZED DESIGN




```
> summary(AnovaModel.1)
            Df Sum Sq Mean Sq F value   Pr(>F)
TRX          2 114.25   57.12   22.11 6.79e-06 ***
Residuals   21  54.25    2.58
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> with(SBP, numSummary(SBP, groups=TRX, statistics=c("mean", "sd")))
       mean       sd data:n
trat1 4.625 1.767767      8
trat2 5.875 1.125992      8
trat3 9.750 1.832251      8
```

4. Analyze data from COMPLETELY RANDOMIZED DESIGN

Vall d'Hebron
Institut de Recerca

Exercise with osteo dataset (*osteoporosis.csv*).

Check if variable *bua* (broadband ultrasound attenuation), is the same in the women of the study (classified by degree of illness)

taula.de.dades

	registro	area	f_nac	edad	grupedad	peso	talla	imc	bua	clasific	menarqui	edad_men	menop	t
1	3	10	11659420800	57	55 - 59	70.0	168.0	24.80	69	OSTEOPENIA	12	99	NO	NO MENOPAUSIA/NO
2	4	10	11671689600	46	45 - 49	53.0	152.0	22.94	73	OSTEOPENIA	13	99	NO	NO MENOPAUSIA/NO
3	10	10	11721024000	45	45 - 49	64.0	158.0	25.64	81	NORMAL	14	99	NO	NO MENOPAUSIA/NO
4	11	10	11464416000	53	50 - 54	78.0	161.0	30.09	58	OSTEOPENIA	10	50	SI	
5	12	10	11690784000	46	45 - 49	56.0	157.0	22.72	89	NORMAL	13	99	NO	NO MENOPAUSIA/NO
6	15	10	11716012800	45	45 - 49	63.5	170.0	21.97	76	NORMAL	14	99	NO	NO MENOPAUSIA/NO
7	16	10	11623737600	48	45 - 49	86.0	161.0	33.18	87	NORMAL	11	99	NO	NO MENOPAUSIA/NO
8	17	10	11562307200	50	50 - 54	61.5	164.0	22.87	74	NORMAL	10	99	NO	NO MENOPAUSIA/NO
9	18	10	11538028800	51	50 - 54	60.5	158.0	24.23	58	OSTEOPENIA	14	99	NO	NO MENOPAUSIA/NO
10	20	10	11332483200	57	55 - 59	64.0	149.0	28.83	61	OSTEOPENIA	13	50	SI	
11	21	10	11631945600	48	45 - 49	70.3	160.0	27.46	67	OSTEOPENIA	12	48	SI	OVARI
12	22	10	11425536000	55	55 - 59	74.4	160.0	29.06	68	OSTEOPENIA	14	50	SI	
13	23	10	11553235200	50	50 - 54	55.5	154.5	23.25	73	OSTEOPENIA	11	48	SI	
14	24	10	11367302400	56	55 - 59	89.0	166.0	32.30	61	OSTEOPENIA	14	47	SI	
15	25	10	11585635200	49	45 - 49	50.6	157.0	20.53	68	OSTEOPENIA	14	40	SI	
16	26	10	11572156800	50	50 - 54	71.4	152.0	30.90	74	NORMAL	14	48	SI	
17	27	10	11590992000	49	45 - 49	78.0	157.0	31.64	62	OSTEOPENIA	12	46	SI	
18	28	10	11293516800	58	55 - 59	72.0	162.0	27.43	65	OSTEOPENIA	11	54	SI	
19	29	10	11215238400	61	60 - 64	68.0	155.5	28.12	65	OSTEOPENIA	14	50	SI	
20	30	10	11405664000	55	55 - 59	75.0	161.0	28.93	92	NORMAL	13	50	SI	
21	31	10	11633155200	48	45 - 49	66.5	153.0	28.41	11	OSTEOPOROSIS	11	99	NO	NO MENOPAUSIA/NO
22	32	10	11287728000	59	55 - 59	101.0	156.0	41.50	82	NORMAL	12	45	SI	
23	34	10	10992758400	68	65 - 69	66.5	145.0	31.63	57	OSTEOPENIA	13	50	SI	
24	35	10	10909382400	69	65 - 69	70.0	168.0	24.80	48	OSTEOPOROSIS	13	45	SI	
25	36	10	11643868800	48	45 - 49	60.1	153.0	25.67	86	NORMAL	14	99	NO	NO MENOPAUSIA/NO
26	37	10	11551420800	50	50 - 54	67.0	159.0	26.50	105	NORMAL	12	45	SI	
27	38	10	11043907200	66	65 - 69	67.0	144.0	32.31	79	NORMAL	12	56	SI	
28	39	10	10948089600	69	65 - 69	70.5	148.5	31.97	40	OSTEOPOROSIS	11	43	SI	
29	40	10	11051251200	66	65 - 69	66.5	147.0	30.77	48	OSTEOPOROSIS	13	40	SI	
30	41	10	11333692800	57	55 - 59	58.5	142.0	29.01	80	NORMAL	15	50	SI	
31	45	10	11029651200	67	65 - 69	60.0	147.0	27.77	49	OSTEOPENIA	13	53	SI	
32	46	10	11544508800	50	50 - 54	70.0	160.0	27.34	119	NORMAL	9	48	SI	

4. Analyze data from RANDOMIZED BLOCK DESIGN

Vall d'Hebron
Institut de Recerca

EXAMPLE

D.Melanogaster BX-C serves as a model system for studying complex gene regulation. It is known that a cis-regulatory region of nearly 300Kb controls the expression of the three bithorax complex (BX-C) homeotic genes: Ubx, abd-A, and Abd-B1. Five flies of each genotype were collected, and DNA was extracted separately from each six sets of three adult heads or three adult abdomens. Expression of BX-C is measured in 6 different positions:

Using RBD , text the hypothesis that there is no significant difference between BX-C positions and the tissue types.

	BX-C position					
Tissue Type	**1**	**2**	**3**	**4**	**5**	**6**
Abdomen	0,21	0,35	0,65	0,97	1,25	1,01
Head	0,15	0,2	0,75	1,1	0,9	0,95

4. Analyze data from RANDOMIZED BLOCK DESIGN

Vall d'Hebron
Institut de Recerca

EXAMPLE

the summary of the data

	Count	Sum	Average	Variance
Abdomen	6	4,44	0,74	0,1654
Head	6	4,05	0,675	0,16275
BX-C-1	2	0,36	0,18	0,0018
BX-C-2	2	0,55	0,275	0,01125
BX-C-3	2	1,4	0,7	0,005
BX-C-4	2	2,07	1,035	0,00845
BX-C-5	2	2,15	1,075	0,06125
BX-C-6	2	1,96	0,98	0,0018

4. Analyze data from RANDOMIZED BLOCK DESIGN

Vall d'Hebron
Institut de Recerca

EXAMPLE

And finally the ANOVA's Table:

Source of Variation	SS	df	MS	F	p-Value
Tissue Type	0,01268	1	0,012675	0,82439	**0,405534**
BX-C	1,56388	5	0,312775	20,34309	**0,002453**
Error	0,07688	5	0,015375		
TOTAL	1,65343	11			

We fail to reject the claim that there is no significant effect of tissue type on gene expression measurements

We notice that the mean effect of the BXC-C position is statistically significant

4. Analyze data from Two way ANOVA

Exercise

Effect of Atorvastatin (Lipitor) on gene expression in people with vascular disease. Gene expression profiling of peripheral white blood cells provides information that may be predictive about vascular risk. Following table gives expression measurements, classified according to **age group** and **dose level** of Atorvastatin treatment:

		Dose Level			
		D1	D2	D3	D4
Age Group	1	100	99	87	98
	2	95	94	83	92
	3	102	80	86	85
	4	84	82	80	83
	5	77	78	90	76
	6	90	76	74	85

Test whether the dose level and age groups significantly affect the gene expression (Data from atorvastatin.csv)

Vall d'Hebron
Institut de Recerca

Table of contents

5. ANOVA assumptions

ANOVA's assumptions have to be checked in the data that they are true before run the model. The residues (therefore the response variable) should be:

1. *Normally distributed*. Each group is approximately normal

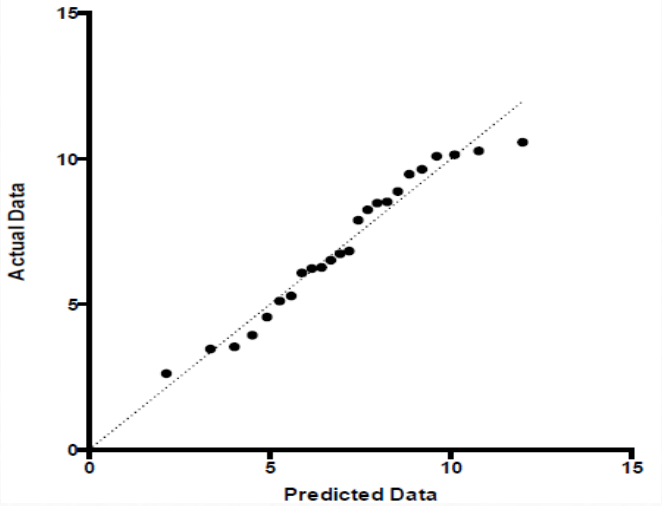
2. *Equally varied*. The variances of different samples are homogeneous (homocedasticity)
3. The errors *are independent* from observation to observation.

How to check them?

5. ANOVA assumptions

Normality of the data.

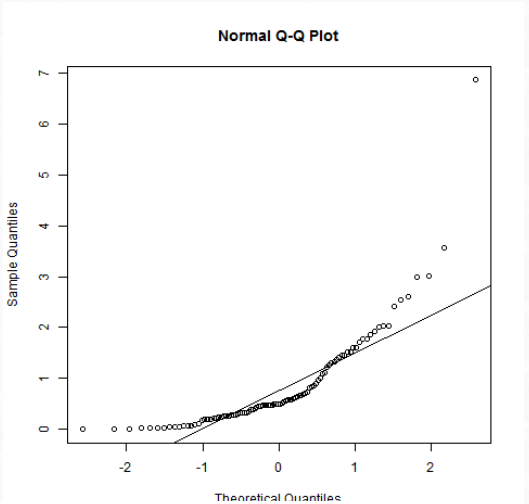
- Generally deviations from the normality do not seriously affect the validity of the assumptions. Not seriously outliers.
- F-test is very robust against non-normal data.
- The best way to check the normality of the data is a QQ plot (Shapiro-Wilk test is also valid):

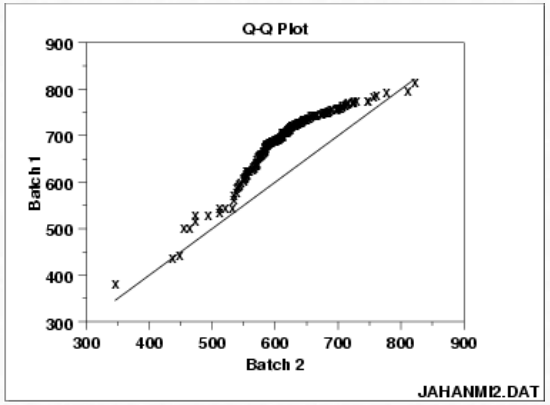


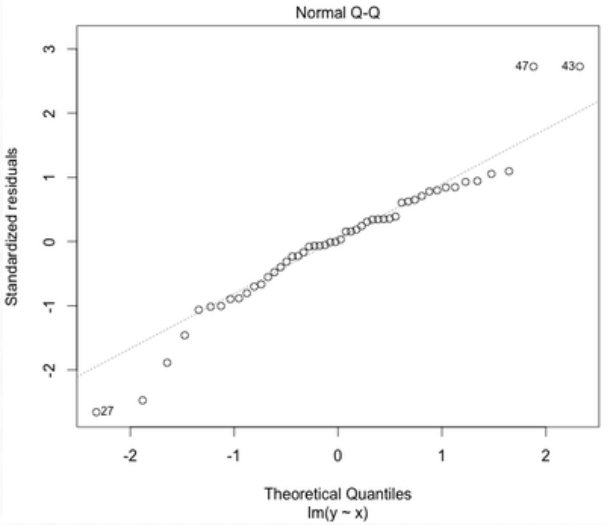
Normally distributed data is expected to line up on the line of identity

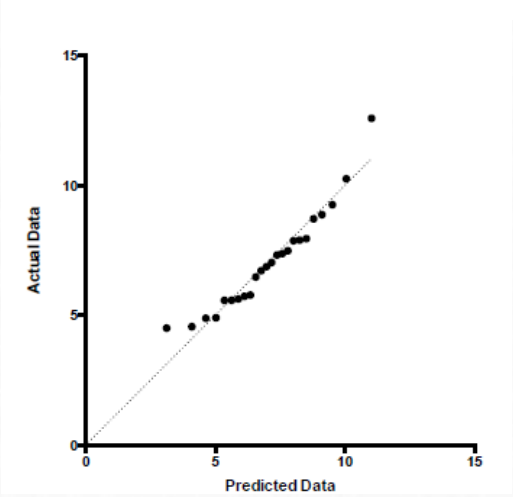
5. ANOVA assumptions

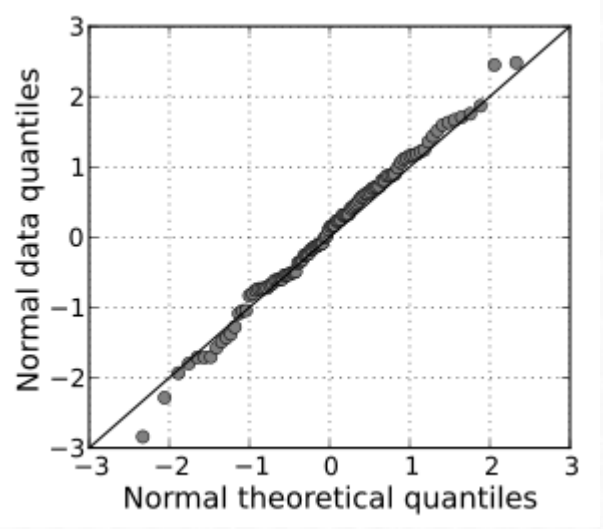
Normality of the data.

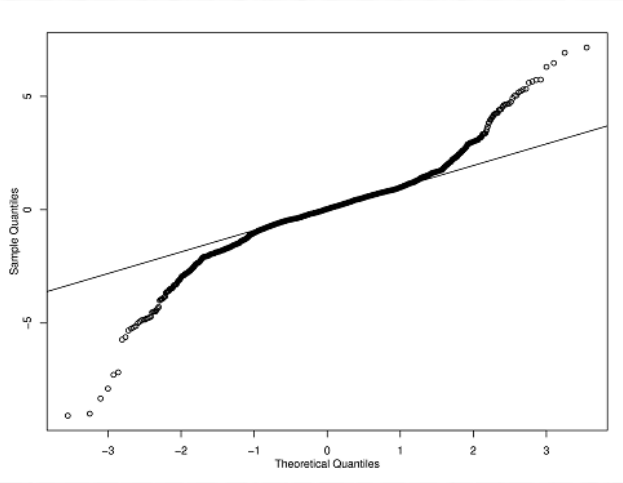






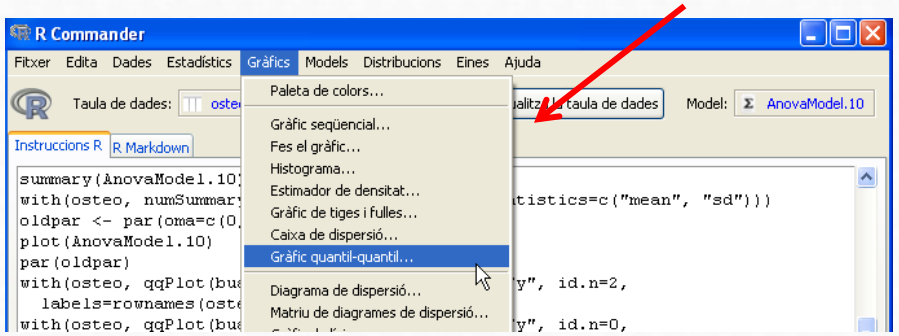


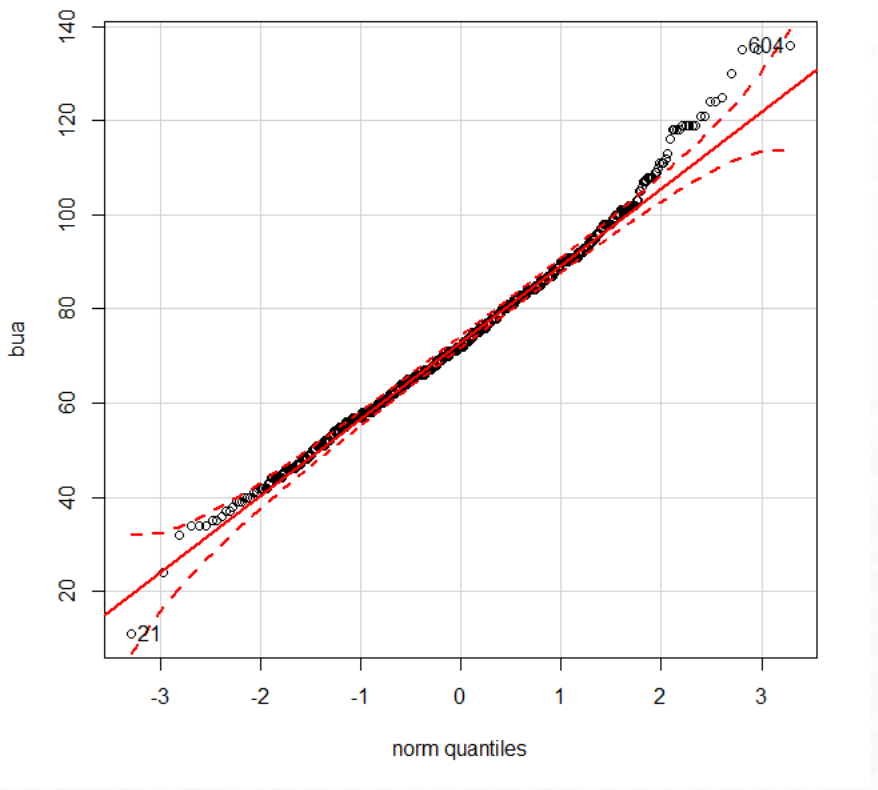




5. ANOVA assumptions

Normality of the data. (in Rcommander)





5. ANOVA assumptions

Homocedasticity of the variances.

- F-test is very robust against heterogeneity of variances, especially with fixed factors and equal sample sizes (*balanced* designs)

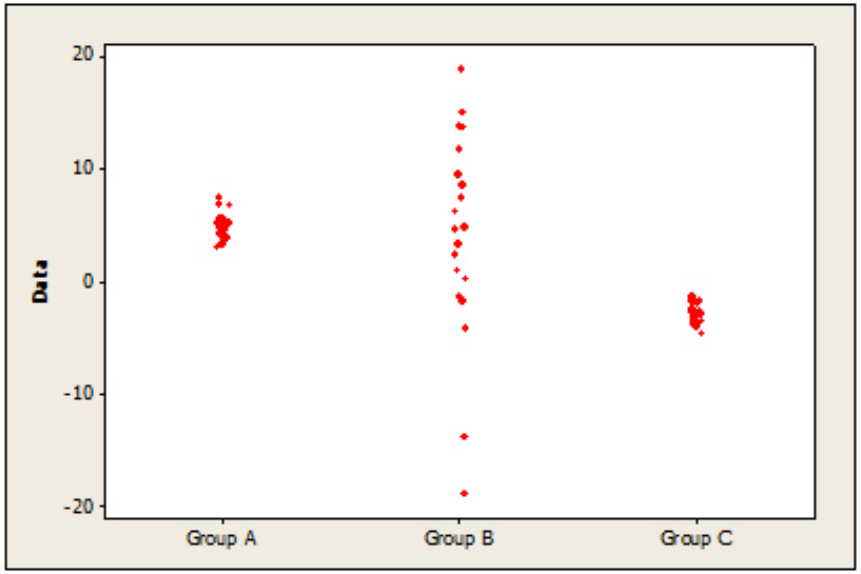
- There are some statistical test, like Breusch-Pagan and Levene's test, to check that…..but it is faster to see a scatter

- There are two tests that you can run that are applicable when the assumption of homogeneity of variances has been violated: (1) Welch or (2) Brown and Forsythe test. Alternatively, you could run a Kruskal-Wallis H Test.

5. ANOVA assumptions

homo → *homo* - same

scedastic → *skedannýnai* – to scatter



Group A and Group C exhibit homoscedasticity.

Group A and Group B exhibit heteroscedasticity

5. ANOVA assumptions

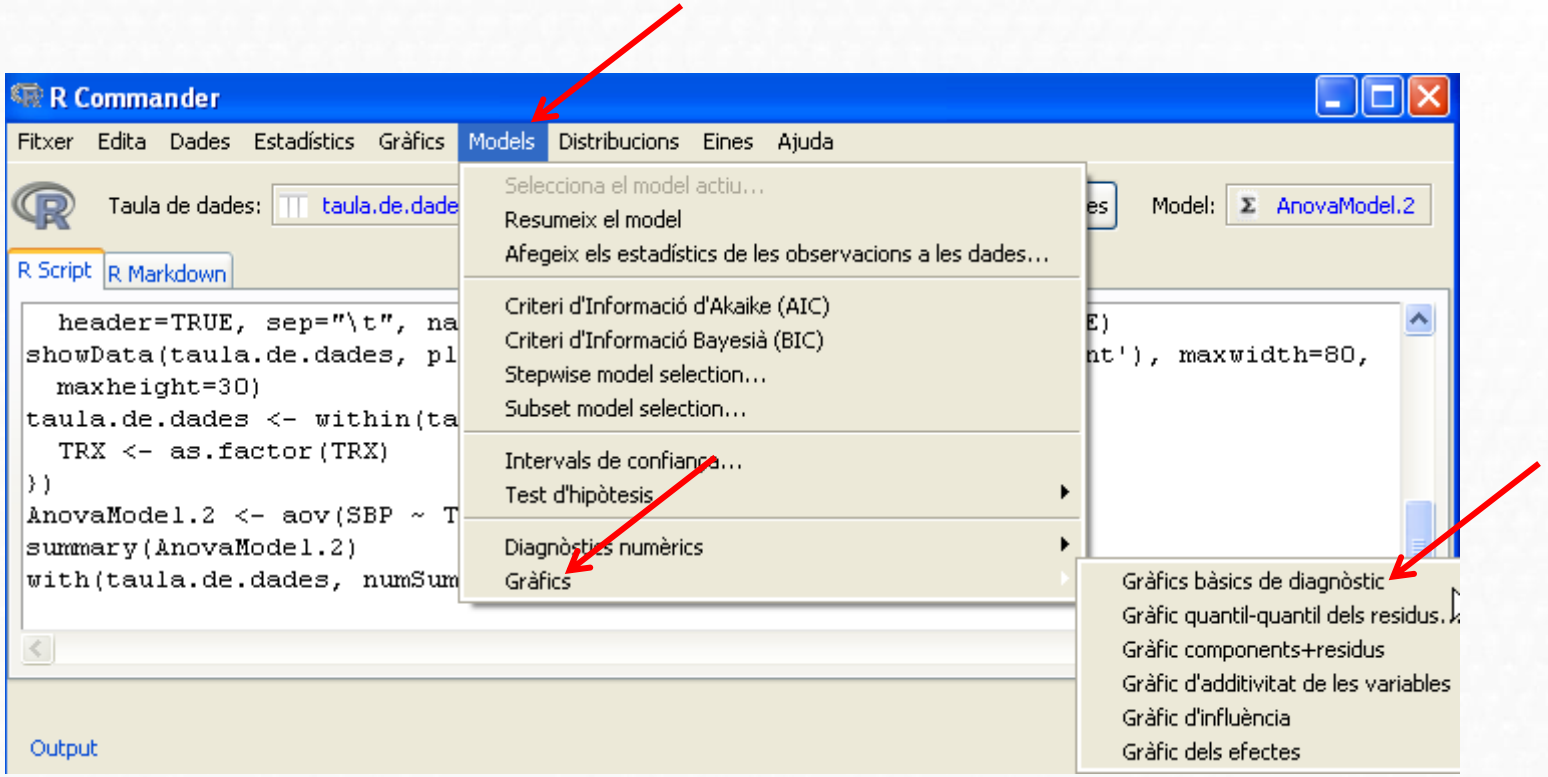
Independence of the residues.

- Independent observations
  - ✓No correlation between error terms
  - ✓No correlation between independent variables and error
- Positively correlated data inflates standard error
- If the residues are independent they won't have to follow any clear pattern when we observe them in a plot.
- It is difficult to determine the absence of a pattern. There are some statistics test to check that (Durbin–Watson), but it is not the scope of this course.

We observe a plot of the residues vs estimates values

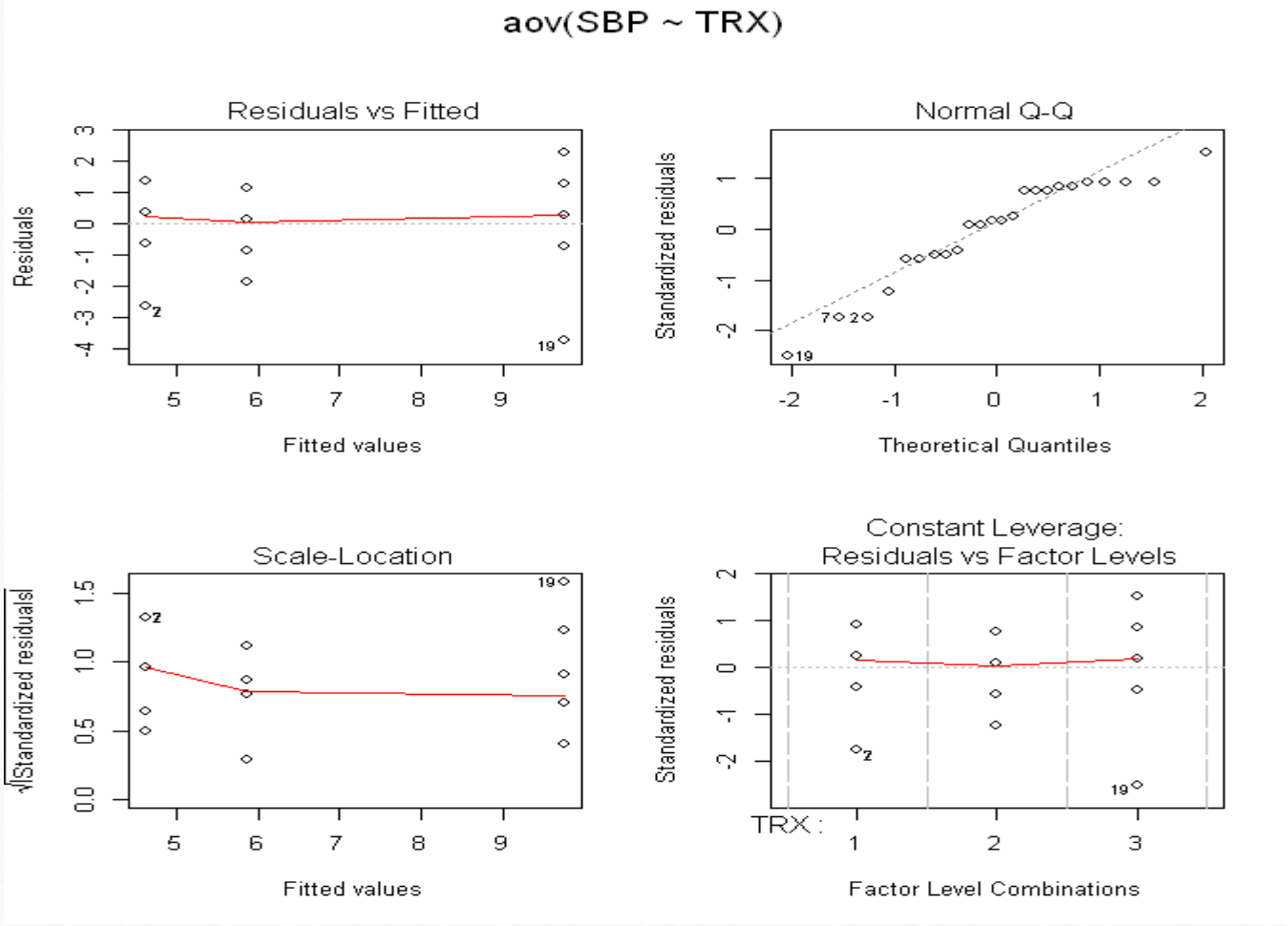
5. ANOVA assumptions

How to check in R-Commander



5. ANOVA assumptions

How to check in R-Commander.



5. ANOVA assumptions

Exercise: Check it with the osteo Data set (Model: bua ~ clasific).

Vall d'Hebron
Institut de Recerca

Table of contents

6. Multiple comparisons

Until now we only can say if there differences among the groups compared, but we don't now between which groups.

- Usually we are interested in the comparison of the samples or treatments two by two (remember we couldn't uses a t-test!!!)

- We need to adjust our p-value threshold because we are doing multiple tests with the same data (type I error probability increases). There some test that take into account this.

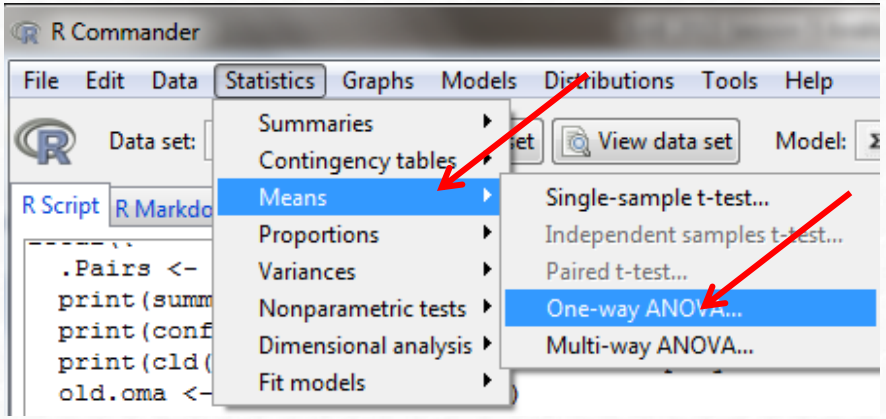
6. Multiple comparisons

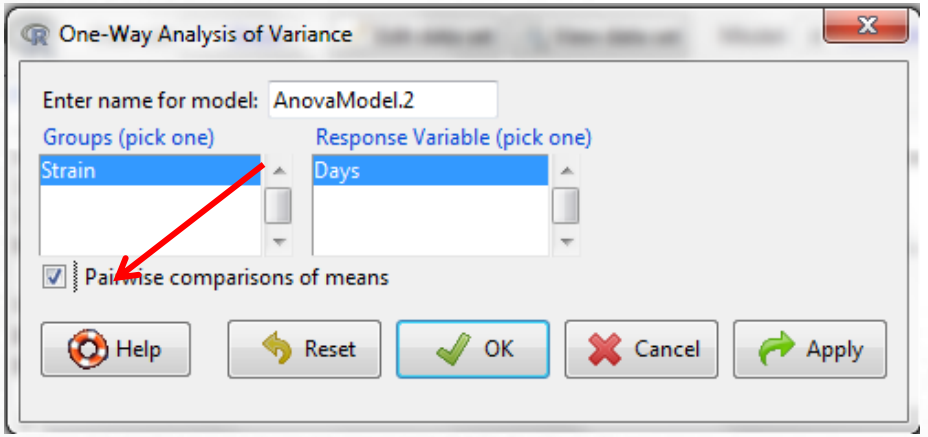
Some test for post hoc comparisons (only to remember the "name").

Bonferroni	$t_{\alpha/q,(r_i+r_j-2)}$	$\sqrt{S^2_{i,j}\left(\frac{1}{r_i}+\frac{1}{r_j}\right)}$
Tukey – Kramer	$q_{\alpha,(a,n-a)}$	$\sqrt{MS_E\left(\frac{1}{r_i}+\frac{1}{r_j}\right)}$
FPLSD	$t_{\alpha,(n-a)}$	$\sqrt{MS_E\left(\frac{1}{r_i}+\frac{1}{r_j}\right)}$

6. Multiple comparisons

Tukey-Kramer in R-Commander





A priori comparisons result is displayed again, plus…

```
> summary(AnovaModel.2)
            Df Sum Sq Mean Sq F value    Pr(>F)
Strain       4  148.2   37.05   29.88 3.39e-09 ***
Residuals   25   31.0    1.24
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> with(Strain, numSummary(Days, groups=Strain, statistics=c("mean", "sd")))
              mean        sd data:n
StrainA 10.000000 1.4142136      6
StrainB  6.166667 1.1690452      6
StrainC  8.666667 1.0327956      6
StrainD 11.500000 1.0488088      6
StrainE  5.666667 0.8164966      6
```

6. Multiple comparisons

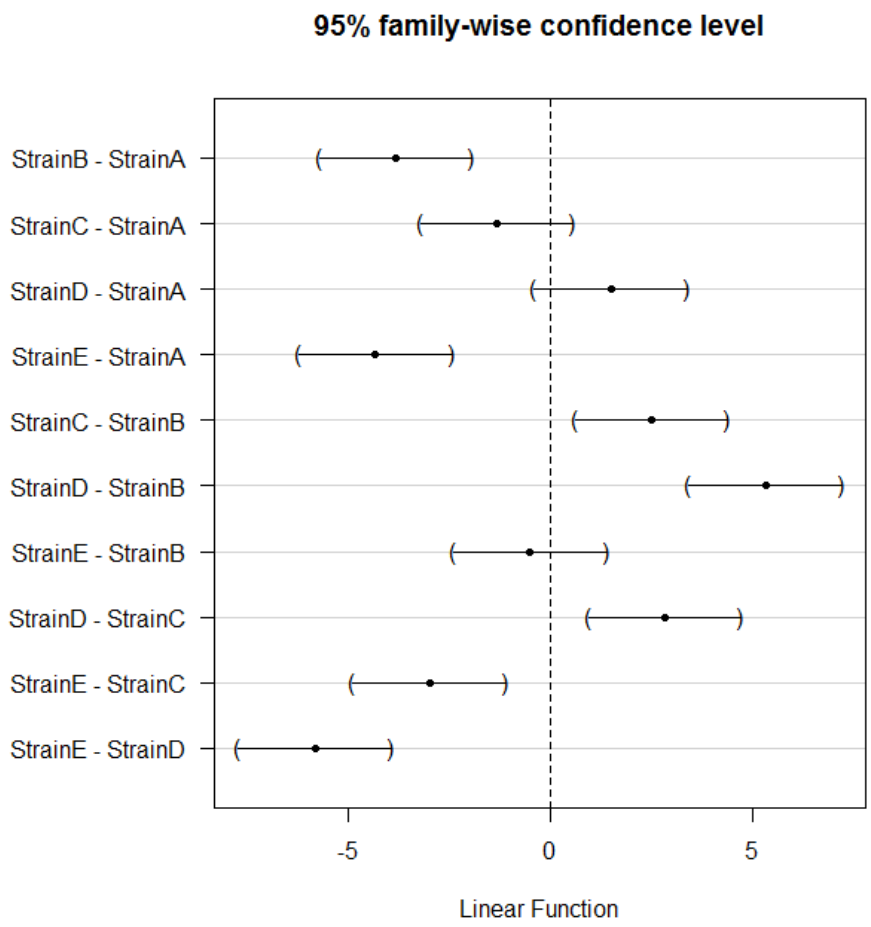
Tukey-Kramer in R-Commander

```
           Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = Days ~ Strain, data = Strain)

Linear Hypotheses:
                       Estimate Std. Error t value Pr(>|t|)
StrainB - StrainA == 0  -3.8333     0.6429  -5.962  < 1e-04 ***
StrainC - StrainA == 0  -1.3333     0.6429  -2.074 0.262331
StrainD - StrainA == 0   1.5000     0.6429   2.333 0.167875
StrainE - StrainA == 0  -4.3333     0.6429  -6.740  < 1e-04 ***
StrainC - StrainB == 0   2.5000     0.6429   3.889 0.005444 **
StrainD - StrainB == 0   5.3333     0.6429   8.296  < 1e-04 ***
StrainE - StrainB == 0  -0.5000     0.6429  -0.778 0.934683
StrainD - StrainC == 0   2.8333     0.6429   4.407 0.001467 **
StrainE - StrainC == 0  -3.0000     0.6429  -4.666 0.000793 ***
StrainE - StrainD == 0  -5.8333     0.6429  -9.073  < 1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

6. Multiple comparisons

Tukey-Kramer in R-Commander

```
  Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = Days ~ Strain, data = Strain)

Quantile = 2.9369
95% family-wise confidence level


Linear Hypotheses:
                       Estimate lwr     upr
StrainB - StrainA == 0 -3.8333  -5.7215 -1.9452
StrainC - StrainA == 0 -1.3333  -3.2215  0.5548
StrainD - StrainA == 0  1.5000  -0.3881  3.3881
StrainE - StrainA == 0 -4.3333  -6.2215 -2.4452
StrainC - StrainB == 0  2.5000   0.6119  4.3881
StrainD - StrainB == 0  5.3333   3.4452  7.2215
StrainE - StrainB == 0 -0.5000  -2.3881  1.3881
StrainD - StrainC == 0  2.8333   0.9452  4.7215
StrainE - StrainC == 0 -3.0000  -4.8881 -1.1119
StrainE - StrainD == 0 -5.8333  -7.7215 -3.9452

StrainA StrainB StrainC StrainD StrainE
   "bc"     "a"     "b"     "c"     "a"
```



6. Multiple comparisons

Vall d'Hebron
Institut de Recerca

Let’s do the same with osteo dataset. (Model: bua ~ clasific).

6. Non-parametric ANOVA

Kruskal-Wallis Test

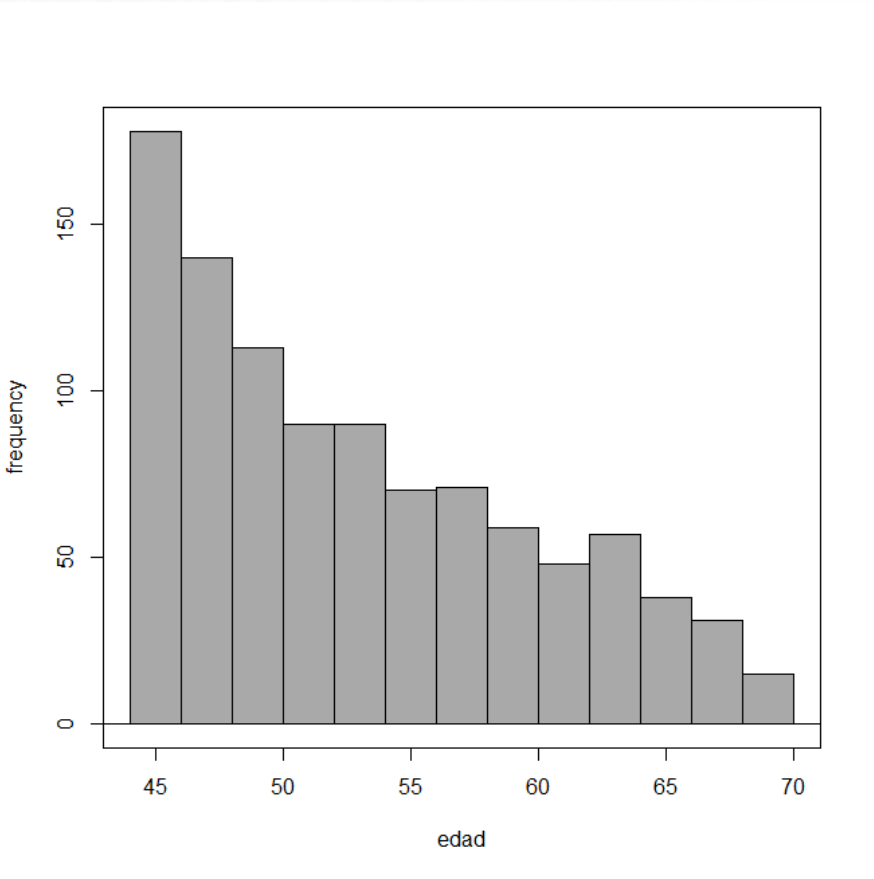
- Kruskal-Wallis test is used to find differences among *k* experimental samples or treatments
- This test don’t assume the normality of the variables.
- It is similar to ANOVA, but it uses the ranges
- Used when number of observations are little and unbalanced designs.

6. Non-parametric ANOVA

Kruskal-Wallis Test in Rcommander

"Osteo" data set. Check if "age" is the same in the women of the study (classified by degree of illness)

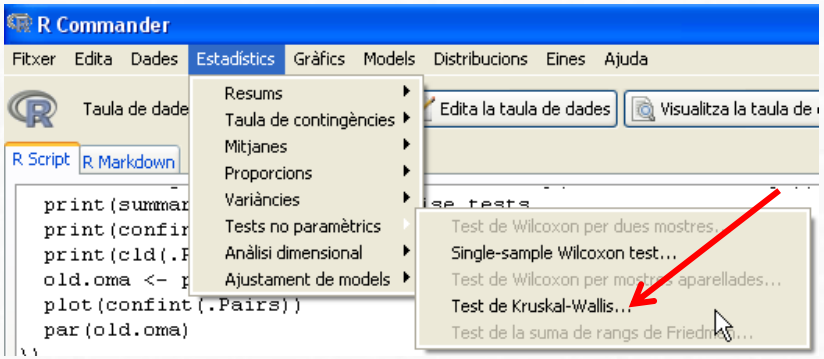
```
> with(osteo, shapiro.test(edad))

        Shapiro-Wilk normality test

data:  edad
W = 0.9237, p-value < 2.2e-16
```

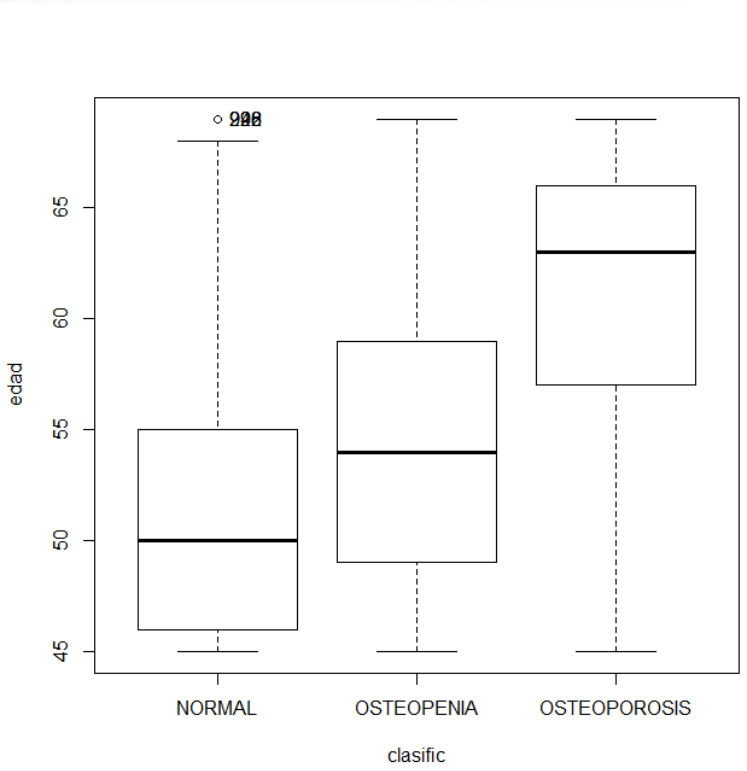


6. Non-parametric ANOVA

Kruskal-Wallis Test in Rcommander

Perform the test.






```
> kruskal.test(edad ~ clasific, data=osteo)

        Kruskal-Wallis rank sum test

data:  edad by clasific
Kruskal-Wallis chi-squared = 112.0766, df = 2, p-value < 2.2e-16
```

Vall d'Hebron
Institut de Recerca

Table of contents

7. Exercises

Exercise 1.

With the dataset osteo, find out if there are some relation between the age of the women and the body mass index (*imc*). After that try to analyze among which groups. The desirable steps to follow would be:

1. Exploratory data analysis (numerically and graphically)
2. Run the model.
3. Check for assumptions of the model (graphically)
4. Run post hoc comparison if necessary

Vall d'Hebron
Institut de Recerca

Exercise 2.

A biotechnology company is developing a new growth factor additive for cell culture, to increase the growth rate in the fibroblast cultures. They are supplementing the typical bovine serum with four different kinds of vitamins(A, B, C, D). They are checking the confluence of the culture plates twelve hours after the addition. Are there any difference among the four vitamins? Which is the more effective? Try to follow the same steps as in the Exercise 1.(dataset = *vitamin.csv*)

Vitamin			
A	**B**	**C**	**D**
75	30	60	50
60	25	55	55
81	33	43	58
53	22	38	59
74	31	62	62
68	18	58	64
82	27	41	58
85	15	48	51
72	35	68	66
70	20	54	68

Exercise 3.

High cholesterol level in people can be reduced by exercise or by drug treatment. A pharmaceutical company has developed a new cholesterol-reducing drug. Researchers would like to compare its effect to the effects of the cholesterol-reducing drug that is currently on the market. Volunteers who have a history of high cholesterol and who are currently not on medication will be recruited to participate in the study.

A. Explain how you would carry out a completely randomized experiment for the study.
B. Describe an experimental design that would improve the design in part A by incorporating blocking.
C. Can the experiment in part B be carried out in a double blind manner? Explain.