Experimental Design and Statistical Methods

Workshop

GENERAL OVERVIEW

Jesús Piedrafita Arilla

jesus.piedrafita@uab.cat

Departament de Ciència Animal i dels Aliments

UAB

TWO TYPES OF STUDIES

- **EXPERIMENT:** it allow us to infer if a particular treatment ***causes*** changes in a variable of interest
- **DESCRIPTIVE** (*"Survey"*): information on a situation or process that cannot be controlled is collected and analysed

  → to establish ***associations***

STEPS TO DESIGN A GOOD EXPERIMENT

1. Formulation of an hypothesis or other objectives of the study
2. Choice of the "experimental unit" and the need for independent replication
3. Controlling variability
4. Choice of treatments
5. Choice of dependent variables to be measured
6. Choice of a design
7. Determination of sample size
8. Statistical analysis planning
9. Pilot study, if necessary
10. Protocols and standard operating procedures (SOPs)
11. Statistical analysis
12. Interpretation and presentation of the results

Formulating the hypothesis and other objectives of the study

- To define them clear and explicitly
  - To estimate treatment means
  - To estimate differences among the means
  - To establish the association among variables
  - To get new information for future studies
- Never start any experiment without knowing how the results will be statistically analysed

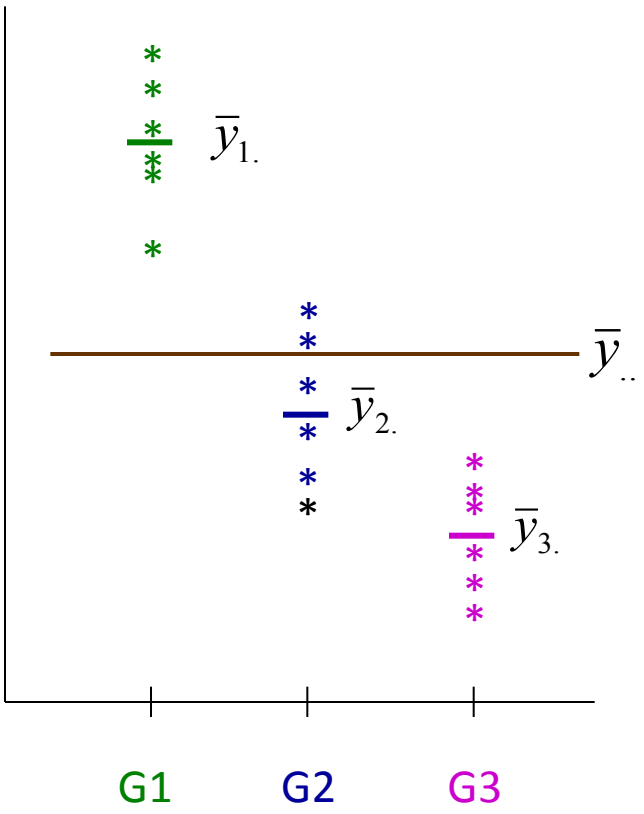
Experimental unit

- Unit of **replication**[1] that can be assigned **at random**[2] to a treatment
  - Animal
  - Box / pen (mean)
  - Petri plate
  - Skin areas
  - Some production

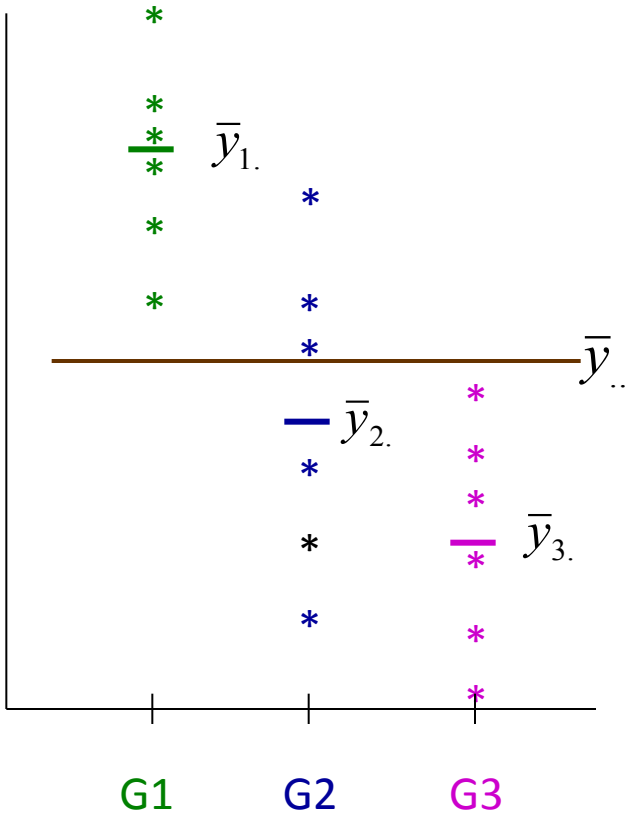
⇒ **Non significant *p*-values can be obtained in the statistical analysis if the experimental units are not established properly**

1 and 2, first and second principles of Experimental Design, according to R. A. Fisher

Two situations



Means differ clearly. Reduced within group variability: observations are alike, close to their means.



Differences among means are not so clear. Larger within group variability and less similarity of the observations within their group.

Statistical tests

Quantitative data → Analysis of Variance (**ANOVA,** test t)

Differences between treatments: **Signal**
Differences within treatments: **Noise**

$$\frac{\textbf{Signal}}{\text{Noise}}$$

Strong statistical significance

$$\frac{\text{Signal}}{\text{Noise}}$$

NO statistical significance

Good experiments minimize random variation, i.e. NOISE

Variability control

- As the **animals** (in general, experimental units) are more **uniform**, a **lower number** is needed in the experiment
  - Genetically uniform animals
  - Freedom from clinic and sub-clinic diseases (SPF)
  - Uniform environmental conditions
  - Homogenous age and weight
- Use adequate statistical techniques when some of these conditions do not met: **blocking**[3], covariates, …
- Avoid biases: Code experimental animals in such a way that the caretakers do not know to which group they pertain

3, third principle of Experimental Design

Variability control (cont. 1)

Genetic types

- Genetically undefined outbred stocks
- Isogenic strains: monozygous twins, inbred strains, congenic strains, consomic strains, F_1 hybrids
- Partially defined strains: mutants and transgenics

Variability control (cont. 2)

Variability types

- Most of the statistical tests compare the magnitude of the (increase in) variability due to the effect -the biological **"signal"**- in relation to the magnitude of the variability of the data -the **"noise"**-.
- **Fixed effects**: **treatment**, strain, sex, age, …
- **Random variation** → contribute "noise": variation in weight, animal genotypes, environmental uncontrolled effects, measurement errors, inappropriate administration of the treatment, ...

Choice of treatments
(independent variables)

- Control groups can be used (non historical)
- Dosage groups can be spaced according to some scale
- Factorial designs allow us to explore the relationships among independent variables

Choice of dependent variables
(characters to be measured)

- Quantitative variables are generally more informative than categorical variables
- To measure as many as possible variables (of interest!), i.e. haematology and biochemical variables, fatty acids, milk characteristics, microbial characteristics and load …

  ⇒**It is important to obtain the maximum of information from the experiment**

- To plan measurement recording prior to start the experiment

Common experimental designs

- Completely randomized design
- Randomized block design
- Factorial designs
- Latin squares and crossover designs
- Repeated measures designs

Completely randomized design

- Random assignment of the animals or experimental units to treatments
- Easy to analyse but inefficient with heterogeneous material

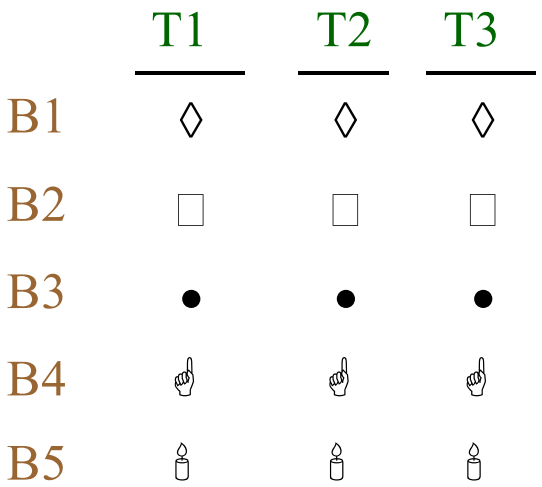
T1	T2	T3	T4
◊	●	°	□
◊	●	°	□
◊	●	°	□
◊	●	°	□

T_i : treatments

d.f.: Treat. = t - 1
error = t (n - 1)

Randomized block design

There is some systematic environmental factor –noise- (i.e. day or batch of laboratory analysis) or the experimental units have some structure (i.e. zones of the skin in a rabbit)

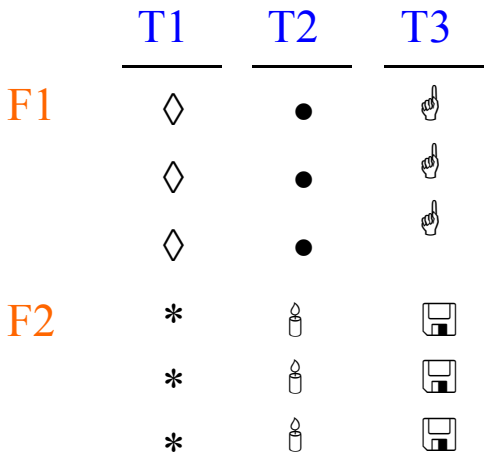


T_i : treatments; B_i : blocks

Note that there is not interaction of treatment × block. There are also RBDs with more than two observations per treatment-block combination that allow to estimate statistical interactions.

Factorial design

- Each cell includes replicates (animals, experimental units in general) that have been submitted to the specific combination of two factors
- It allows to detect **interactions** between factors
- The model can be expanded to more than two factors



T_i : treatment (first factor); F_i : second factor

Latin square

Besides the main effect, treatment, there are two potential sources of heterogeneity, i.e. laboratory and technician, hour of the day and body weight, …

	HD1	HD2	HD3
PC1	A	B	C
PC2	B	C	A
PC3	C	A	B

A, B, C : treatments

HD_i : hour of the day

PC_i : body weight

$N = rows^2 = cols.^2 = Treat.^2 = 9$

d.f.: HD = PC = Treat. = Error = 2

Note that this design is an efficient way to use few animals
But … observe the loss of degrees of freedom

Crossover designs

- Two or more treatments are applied in a particular sequence to the same animal. The experimental unit is the animal in a given time
- Risk of “carry over”, make some “washout”

	A1	A2	A3
P1	A	B	C
P2	B	C	A
P3	C	A	B

A, B, C : treatments

A_i : animals

P_i : periods

N = 3

Comparison of designs

Table 4.1 Summary of the most common experimental designs reviewed in this chapter

Type of design	When to use	Advantage	Disadvantage	Notes
The completely randomized design (fixed effects; random effects)	Fixed effects: when uncontrollable sources of variation are unlikely to be important Random effects: when only interested in quantifying source of random variation (e.g. surveys) rather than compare treatments. Useful for within-subject studies	Simplest 'designs' Easy to use and analyse Less affected by unequal sample size	No control of additional 'nuisance' variation caused by uncontrolled time and space variation etc. which might affect results	Typical designs used when doing simple *t-tests*
Block designs (including Latin square)	When wanting to remove the effects of known or suspected 'nuisance' variation. The evidence shows that it is underused in the biomedical literature, particularly in its simplest form (controlling at least *one* source of unwanted variation)	Deals with known or suspected 'nuisance 'variation in a systematic and powerful way by breaking the experiment into a series of subunits which are analysed as a whole. Unequal sample sizes present some problems	Can be sensitive to missing values Becomes increasingly complex when dealing with two (Latin square) or three sources (Graeco-Latin squares) of unwanted variation	These designs aim to improve the *precision* of the experiment. They can be used in conjunction with others (e.g. factorial designs). Analysis of covariance, another way of increasing precision is also discussed
Factorial designs	When wanting to increase the 'generality' of the results, by testing *simultaneously* the potential effect of several factors (e.g. treatment, sex, strain) and their interactions on the response (dependent variable[s])	A much more powerful alternative to doing several smaller experiments for each factor. Allows testing for interactions between these factors	Interaction between more than two factors are often difficult to interpret	These designs aim to increase the *amount of information* yielded by the experiment. They can be used in conjunction with other (e.g. blocking) designs
Repeated measure designs (crossover, split plot, mixed effects designs)	When using several measures on the same individual, either to study change over time or as a way of dealing with strong individual variation	A more powerful alternative to the dreaded repeated *t-tests*! The individual can act as its own control (crossover)	Crossover: not valid if there is a strong order effect. It may be dangerous to consider time as an independent variable in these experiments	High precision as a result of eliminating inter-individual variation
Sequential designs	When results with individual animals can be obtained quickly, and where the aim is to minimise animal use	Results are analysed as the experiment unravels, enabling the experimenter to stop as soon as significance is obtained	Requires expert advice. Logistics may be a problem	Up-and-down method may be replacing the classical LD50 test in the USA

(Festing et al., 2002)

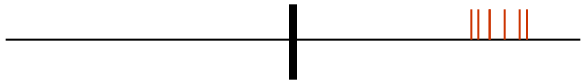
Precision of experimental designs

- **Accuracy:** how close the estimated mean of replicated measurements is to the true mean
  $\Rightarrow$ **absence of BIAS**
- **Precision:** how close together the measurements are regardless of how close they are to the true mean. Usually expressed as the **amount of information**:

$$I = \frac{n}{MS_{Error}} \quad \Rightarrow \quad I = \frac{1}{s_{\bar{y}}^2}$$

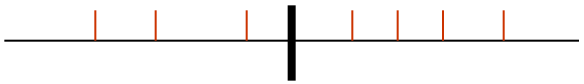
Accuracy and precision

Not accurate but precise



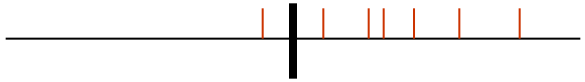
True value

Accurate but not precise



True value

Not accurate not precise



True value

Accurate and precise

True value

Sample size

➔Must be enough to obtain a good estimate of the within group standard deviation and contrast any treatment of interest

BUT

Without wasting scientific resources (animals, time, drugs, products, … MONEY!)

➔**Do not** use **6 replicates** –or another arbitrary number- because anybody does!

Factors affecting sample size

- Objectives of the study
- Type of data to be collected
  - Categorical
  - Numerical: discrete or continuous
  - Percentages
- Uniformity of the experimental material
- Design of the experiment

(we will develop in depth this topic later)

Some packages to determine sample size

- Granmo (http://www.imim.es/ofertadeserveis/es_softwarep_blic.html)
- Minitab
- SAS
- nQuery Advisor
- Power and Precision
- Specific programs in R
- PASS 13 (http://www.ncss.com)

Significance *vs* relevance

- Small experiments do not have power, but increasing sample size increases power only to a limit, from which the increase is only marginal (power curves)
- There must be an equilibrium between power and cost
- A **statistically significant** difference (for example when sample size is very large) can be **irrelevant** from a biological, clinical or productive point of view

Statistical analysis planning

- Before starting the experiment we must know how to perform the statistical analysis
- Each experiment must be statistically analysed before starting the next experiment
- Anticipate complications or losses of experimental animals: **sample size is referred to data available for the statistical analysis**

Pilot study

- To ensure that the study is feasible even if technical or logistic difficulties are anticipated

BUT

- The estimates of the within group standard deviations (σ) are usually imprecise

  $\Rightarrow$ use them with care in determining sample size

Protocols

- Discuss the protocol with some other colleagues after designing the experiment. Explain also the protocol to the caretakers.
- Advisable practices:
  - Daily record, including notes of ever event
  - Register details on the loss of animals or measurements and their causes: that can be explained in Material and Methods.
  - Suspects on a given record: possible ***OUTLIER***?
    → retain, discard, use non parametric methods?

Objective of the statistical analysis

To give some quantitative estimate of the probability that any observed differences between treatment groups could have arisen by chance sampling variation due to differences among individuals, rather than being caused by the treatments

Making the statistical analysis

- Discard all possible erroneous data
- Verify the applicability conditions (of ANOVA)
  - Normality of errors
  - Homogeneity of within group variances
  - Independence of the data
- If needed, transform the data to an appropriate scale to meet the previous conditions

Making the statistical analysis (2)

- ANOVA tell us that there exist differences between treatments, but between which ones?
  - Decide which are the comparisons of interest
  - Multiple comparisons (*post hoc*): they control the risk of false positives (Type I error)
- **DO NOT MAKE** comparisons between means if the *F* value is not significant

Some statistical packages

- Excel
- R
- Genstat
- Minitab
- SAS
- SPSS
- StatXact

Interpreting results

- Where it is unlikely that a difference could have arisen by chance, it is possible to obtain some idea of the likely magnitude of any differences by estimating a confidence interval for any difference.

BUT

- We do not know if the differences will arise in any other conditions $\rightarrow$ factorial design, randomized blocks?

$\Rightarrow$ The interpretation of the results is largely a matter of **scientific judgement**, though the statistical analysis should prevent claims of important treatment effects when these could be due entirely to sampling variation

Presentation of the results

- **As clearly and briefly as possible**
- Describe the experimental design and the statistical analysis in the Materials and Methods section of the paper
  - The form of describing this section varies among journals
- Sufficient numerical data should be given to allow other investigators to reach the same conclusions as the authors
- Tables and figures should be clearly laid out
- Sometimes results are not statistically significant, but this is, in itself, of interest: non significance may occur because there is no true effect, or because the experiment was too small to be able to detect a biologically important difference

SOME INTERESTING BOOKS

Festing M.F.W., Overend P., Gaines Das R., Cortina Borja M., Berdoy M. 2002. *The Design of Animal Experiments. Reducing the use of animals in research through better experimental design.* Laboratory Animals Ltd., London. (www.lal.org.uk)

Morris T.R. 1999. *Experimental design and analysis in Animal Sciences*. CABI Publishing, Wallingford UK.

Ruxton G.D., Colegrave N. 2006. *Experimental designs for the life sciences*. Oxford Univ. Press, Oxford.

Una definición de Estadística

Para tener siempre presente …

Lo esencial es invisible a los ojos; no se ve bien sino con el corazón.

Saint-Exupéri

The three Rs

Legal requirement in the EU: researchers that plan a project involving vertebrates must consider using alternative methods

- **Replacement:** to use alternatives to animals, such as tissue culture or simulation studies
- **Refinement:** to minimise animal pain through management
- **Reduction (main topic of this course):** to minimize sample size while reaching the desired scientific objectives:
  - To understand the objectives of the study
  - To control variation with efficient experimental designs
  - To perform an adequate statistical analysis
  - To interpret results