An introduction to Experimental Design

With applications to bioinformatics

Alex Sanchez-Pla
Departamento de Genética, Microbiología y Estadística (UB)
Versión 2025-12-10

Outline

1) Experiments: What, why, how

2) Principles of experimental design

3) Basic types of experimental designs

4) Some resources

Experiments: What, Why, How

What is an Experiment?

An experiment is a systematic procedure conducted to **investigate a scientific hypothesis, validate a theory, or explore a phenomenon**. It involves manipulating variables under controlled conditions to observe and measure outcomes.

Through careful design and data analysis, experiments aim **to provide empirical evidence, establish causation, and contribute to the understanding of natural or social phenomena**.

The controlled nature of experiments allows researchers to draw meaningful conclusions and advance knowledge in various fields, including physics, psychology, biology, and social sciences.

Why experiment?

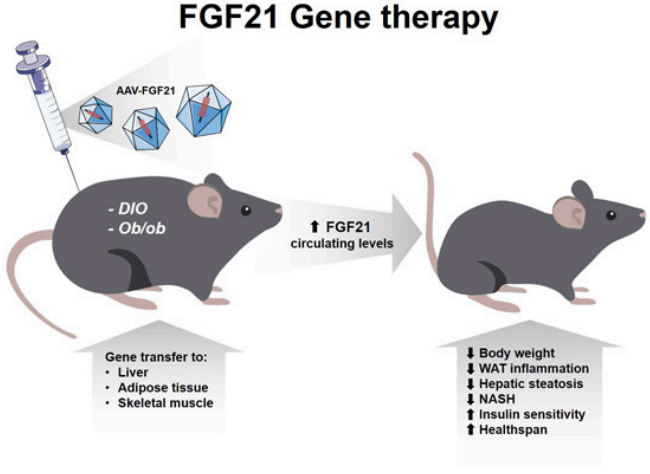
- The purposes of experimental studies are diverse. For example, we could propose experiments to:

    1. Compare responses to different treatments.
    2. Determine the cause (s) of the response variation.
    3. Find the conditions in which the optimal response is reached,
    4. Develop a model to predict responses.

- A characteristic of experimental versus observational studies is that, *under the right conditions, the former are the only ones that allow establishing causal relationships*.

An example and basic ideas

- Goal: investigate the effect of a drug/therapy to prevent the development of diabetes in genetically modified mice.
- Effect will be tested and compared with that of a placebo.
- Effectiveness of the treatment may be affected by the diet of the animal, so distinct diets ("D1", "D2", "D3") are considered
- Mice randomly selected will receive a distinct combination of drug/placebo and D1/D2/D3 diet.
- To determine the effect of each drug/diet combination, weight, and decrease in insulin resistance will be determined.



Columna1	*Placebo*	*Drug*
Diet 1	2, 5, 15	12, 14, 1
Diet 2	4, 7, 11	16, 8, 10
Diet 3	6, 15, 18	3, 9, 17

Some definitions

- An **experiment** is any investigation in which a particular set of conditions is applied to selected (groups of) individuals and the results are observed and evaluated. app.
  For example, effect of therapy in pre-diabetic mice

- Each group of experimental conditions is a **treatment** or **factor**.
  For example "Drug" and "Age" are two factors.

- Each particular condition within a factor is a **level** of that factor.
  For example Drug/Placebo is a factor with two levels each.
  Old/Aged cannot be assigned to a mouse. Not a factor, but a **block**.

- Results observed after applying a treatment are the **responses**.

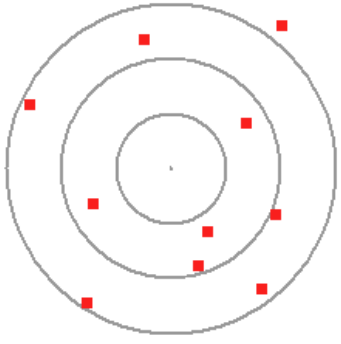
Some important definitions:

- **Experimental Unit (UE)** The physical entity or subject exposed to the treatment independently of other units.
- Each mouse is an experimental unit
- **Unit of observation (UO)** The unit in which they are carried out. observations, that is, measurements.
- It can be a sample of the EU or be identical to the EU.
- **Experimental error**: random variation observed between different repetitions carried out, or not, under the same experimental conditions.
- **Observation error**: The variation between multiple observations of the same experimental unit.

Types of variability

Random variability

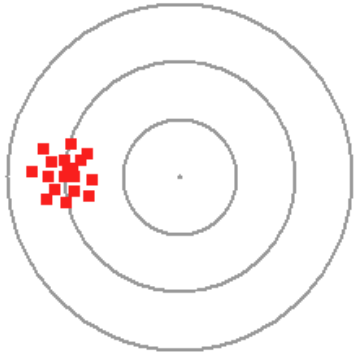
- Differences expected to be observed when different subjects from the same sample are measured.
- It is usually always present to a greater or lesser degree.
- It is usually reduced by increasing the sample size



Random Error

Systematic variability

- Differences between subjects or observations attributable to the measurement process or to a non-random selection of all the individuals in the sample.
- Usually can be corrected



Systematic Error

What characterizes an experiment?

1. The treatments to be used
2. The experimental units to be used
3. The way treatment levels are assigned to the experimental units.
    - This is precisely ***the experimental design***
4. Responses that are measured

A good experimental design ...

- **Avoids biases** or systematic errors
- **Allows a precise estimation of the response**, which implies that the random error is as low as possible.
- **Allows an adequate estimation of the error**.
- **Has wide validity**: the experimental units are a sample of the population in question, so it is possible to extrapolate the conclusions of the sample to the population.

Why do we need a good design

- Well designed experiments will be **scientifically sound**:
  - Smaller bias,
  - Smaller variability,
  - Higher generizability.
- They allow **saving money**
  - Budget can be done based on goals
  - Cost can be optimized
- Allows **avoiding unnecessary animal use**
  - The three Rs

How to get a good design

- Apply basic ideas which, together, use to guarantee a good result.

    1. Rely on a *checklist* of the experimental design.
    2. Apply *the scientific method in the appropriate study*.
    3. Be based on the basic principles of Experimental Design

        - Randomization,
        - Replication,
        - Local control.

- And also

    - Plan design and analysis **at the same time**,
    - Involve your favorite statistician from the beginning (or before) of the experiment.

Design checklist

1. Define the objectives of the experiment.
2. Identify all possible sources of variation.
3. Select an appropriate experimental design.
4. Specify the experimental process
5. Conduct a pilot study
6. Specify the hypothesized model
7. Describe the tests to be performed.
8. Estimate the required sample size using the results. of the pilot study
9. Review your decisions in 1 - 8 and make the necessary adjustments.

Principles of experimental design

Eperimental design principles

- Good experimental designs share common traits.
    - Based on the logic of experimentation and the scientific method,
    - Rely on *common principles* , whose application guarantees good designs, or, at least, better designs than those studies in which they are not taken into account explicitly.
- Some of these principles are:
    - *Randomization*
    - *Replication*
    - *Local control or blocking*

1. Randomization

- It is not possible to avoid random variations but,
  by randomly assigning treatments to units the effect of such variation can be compensated.

  This is precisely what ***randomization*** *is about.*

- This can be done in many different ways depending on the experiment:

  - Randomly assigning individuals to treatments and/or
  - Running the experiments in random order.
  - ...

Haphazard is not randomization

- **Randomizing does not mean doing everything at random**.
- E.G. Imagine we have 4 treatments to be assigned to 16 units
  - **A good strategy** :
    - 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
  - **A bad way to do it**
    - Treatment A is assigned to the first four units we have encounter,
    - treatment B to next four units, and so on.
    - *This approach does not protect against possible unknown groupings in experimental units*.

Randomization avoids confounding

- Consider a new drug treatment for coronary artery disease.
  - 2 treatments to compare: (1) drug treatment vs (2) bypass surgery.
  - 100 patients will be assigned to one of 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:
  - We decide to assign stronger patients to surgery and the weaker patients to the drug therapy.
- Patient strength **is confounded** 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. Replication

- There is general agreement on the need to apply each treatment independently to several experimental units.
- This ...
  - Helps to establish the reproducibility of the results.
  - Protects against eventual abnormal or unusual results.
  - Provides a way to estimate the variance of the experimental error in the absence of systematic differences between the experimental units.
  - Provides the ability to increase the precision of estimates of means of treatment.

Replication $\neq$ (necessarily) results!

- Replication is importanty but, *it does not necessarily guarantee valid estimates of experimental error* or, what is more:
- Having the appropriate sample size **does not guarantee the presence of an effect**.
- The -often heard sentence- *"we didn't detect any effect but if we can collect enough samples the effect will be seen"* can be considered a *Statistical Myth*.

Replicates, power and precision

- Number of repetitions r relates to the *precision* of the experiment
- Precision is inversely related with the Variability of the experiment.

$$\text{Precision} \sim 1/var(\overline{X}) = r/\sigma^2 \quad (*)$$

- While this is stated for estimating the sample mean, the rule can be easily extended to other characterristics.

- From (*) it follows that:

  - the greater the number of replicatess, r,
  - and the lower the variability, σ^2,
  - the greater the precision a design provides.

- From this relation, it is straightforward to derive formulae for the sample size needed for estimation.

How many replicates are needed?

- If the goal of an experiment is, not only estimating one characteristic, but also comparing groups, that is *detecting the effect of a treatment*, this can also be accounted for.

- Given the relation between:

  - (1) The variability
  - (2) The desired effect size,
  - (3) The level of significance ("alpha", type I error) of the test,
  - (4) The power (1-"beta", type II error) wished to attain

- One can compute the sample size needed given the previous four values or, one can fix any three and compute the other one (for instance the power given a sample size, etc.)

Sample size calculations

- Sample size to compare 2 (normal) samples with common variance σ, at a significance level α, a power β to detect an effect size $\mu_t - \mu_c$.

<Hypothesis>

$H_{01} : \mu_{t1} - \mu_c = 0$ **vs** $H_{11} : \mu_{t1} - \mu_c \neq 0$

$H_{02} : \mu_{t2} - \mu_c = 0$ **vs** $H_{12} : \mu_{t2} - \mu_c \neq 0$

Where μ_{t1}, μ_{t2} and μ_c are true means of test (EA1 and EA2) and control groups,

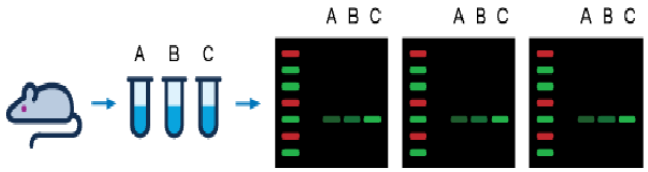
<Sample size calculation>

$$n = \frac{2\sigma^2(Z_{\alpha/2} + Z_\beta)^2}{(\mu_t - \mu_c)^2}$$

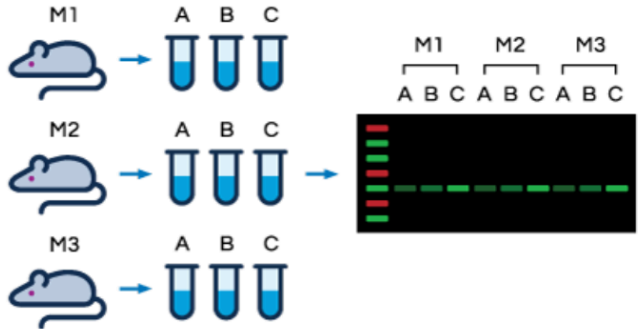
$Z_{\alpha/2}$: Z value from the standard normal distribution, 1.96
Z_β: Z value from the standard normal distribution, 1.282
$\mu_t - \mu_c$: true mean difference
σ: population standard deviation, assumed as $\sigma = \sigma_t = \sigma_c$

- Sample size online calculator example: [granmo] (https://www.imim.es/ofertadeserveis/software-public/granmo/)

Technical and biological replicates



Technical replications allow quantifying variability associated with the technique used.



Biological replications allow quantifying the variability associated with the study population.

- The total variability can be decomposed into various *components of the variance*.

$$\sigma(TOTAL)^2 = \sigma(TEC)^2 + \sigma(BIO)^2 + \sigma(ERR)^2$$

- *In general* :

$$\sigma(TEC)^2 < \sigma(BIO)^2$$

Replicates or *pools*?

- Old omics studies, previous to single-cell, often considered combining mRNA from different samples to form a "pooled sample" or *pool*
- This could be done when ...
  - Each separate sample does not provide enough mRNA
  - To compensate excess variability by "averaging" similar samples.
- This can be misleading, but correct if done apprpriately:
  - Combining several samples in each group but ...
  - Using several groups of different samples
- What not to do:
  - Don't use groups when individual information is important (e.g. paired designs).
  - A sample of 3 grouped individuals $\neq$ 3 individual samples!

3. Local control

- In many situations samples are not homogeneous.
- For example, in an experiment to compare two treatments using expression microarrays, there may be different types of subjects:
    - Animals from several litters
    - Samples processed in different days due to lab restrictions
    - and other sources of known but unavoidable variability.
- Systematic differences between groups of samples ("blocks") can obscure the effects of interest (for example the effect of a treatment).
- That is, it may not be clear if the observed differences are attributable to the effect of the treatment or other *confounding* factors.
- This undesirable effect may be decreased by *blocking* or local control, that is *distributing each treatment evenly among the different blocks* .

How to apply 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

- This design does not apply good local control.
- Treatment effect can be confused with
  - the effect of age or
  - that of the production batch.

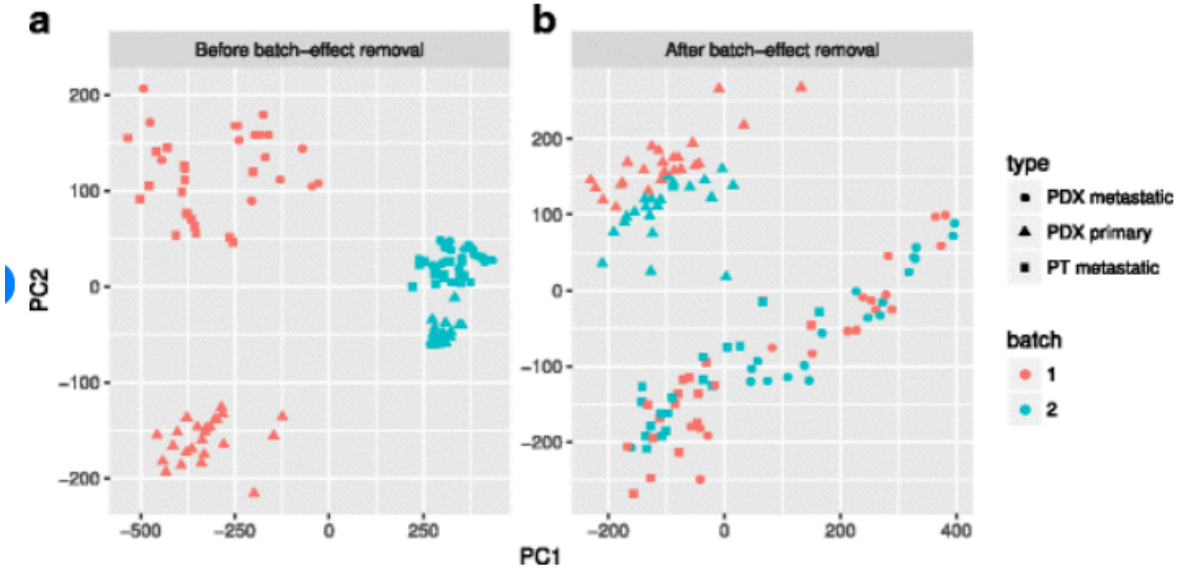
Sample	Treatment	Sex	Batch
1	A	Male	1
2	A	Female	2
3	A	Male	2
4	A	Female	1
5	B	Male	1
6	B	Female	2
7	B	Male	2
8	B	Female	1

- This design applies good local control.
- The possible effect of sex or of the production batch is distributed among the different levels of treatment, which will allow them to be analyzed separately.

Batch effect

- A **batch effect** is a non-biological, systematic source of variation introduced when samples are processed in different groups (e.g., different days, operators, reagent lots, instruments, or runs).
- Batch effects can create apparent differences between samples that **mimic or obscure true biological signal**.
- Good experimental design aims to minimize batch effects through **balanced processing and randomization**.

Batch effect detection



Batch-effect removal. The PCA plots show the before/after median alignment comparison. The colors indicate the two batches 1 and 2, and the shapes indicate the three cell types reported from the original data. a Before batch-effect removal; b after batch-effect removal

- Principal Component Analysis can reveal hidden structure in the data.
  - Can be used to show undetected blocks in the design.
- If samples separate primarily by **batch** rather than by treatment or phenotype, this indicates a batch effect.

Adjusting batch effects

- Batch effects can be corrected statistically by including **batch as a factor** in a linear model or using specialized methods such as **ComBat**, linear regression of unwanted variation (**RUV**), or mixed-effects models.
- Correction methods require that **batch and treatment are not confounded** otherwise the true biological signal cannot be recovered.
- Prevention is even more important:
  - **randomize** sample processing order,
  - **balance** treatment groups within each batch, and
  - document metadata to allow correction.

Basic types of experimental designs

Experimental designs

- A key point in any experiment is the way in which *the experimental units are assigned to the treatments*.
- This assignment must be done in such a way that:
  - it is possible to estimate the effects that interest the researcher
  - the random variability is as small as possible ("maximum precision") with the available resources.
  - the best possible local control is achieved, given the circumstances of the experiment.
- To achieve the best possible design, we will take into account the components that define each design.

Design components

When choosing a design for an experiment we must consider:

- The **design of the treatments**.
  - Which and how many treatments are included in the study?
  - Are they considered separately or in combination?
  - What are the levels of each treatment?
- The **error control design**.
  - How are treatments assigned to experimental units ?.
    This depends on the resources, the available units, the desired precision, the heterogeneity between UEs.
- The **Observational Design**
  - At what level are the observations made?
    Is each EU an OU or are there several OUs per EU (subsampling)?

From components to design

	Treatments Design	**Error control Design**	**Observational Design**
Completely Randomized	1 factor, (k levels); replicates/level='r'	Assign treatments 1 ... k to EU	1 EU = 1 OU
Completely randomized block design	1 factor (k levels); 1 block (l level), k*l EU*	Assign treatments 1 ... k to EU, in each block	1 EU = 1 OU
Two-factor design with interaction	2 factors (k, l levels); r replicates of each combination	Assign each combination 1 ... kl of treatments to EU	1 EU = 1 OU
Repeated Measures Design	1 factor (k levels); l repeated measures; replicates/combination = 'r'	Assign treatments 1 ... k to individuals at time 1	For each EU there are 'l' OU (temporal measurements)

Designs, Models and Analysis

- The design of the treatments and the observational design help us to choose the appropriate design for an experiment.
- Each design can be represented by a *linear model*. This model:
  - Represents the relationships between responses, treatments and experimental and observational units.
  - Is the basis for the analysis of the data once it has been collected.
- Error control design defines how the randomization is carried out, that is, the assignment of individuals to the treatments.
  - This should be done when *planning the investigation*.

Experimental design and ANOVA

- Sometimes the design of the experiment is confused with its analysis, which is done using "Analysis of the Variance" or ANOVA techniques.
    - This is understandable, because when one defines the experimental design the way it will be analyzed is set. That is *they are related, but they are not the same*.
    - It is a common problem, in some books or statistics courses, which do not pay attention to how treatments were allocated between individuals and provide the data already collected.
    - This makes it difficult for students to realize that experimental design had been carried out before the data were collected.
- Summarizing: Although the treatment design suggests a certain analysis model *the experimental design should not be confused with the analysis of the data collected in the experiment!*

Experimental design and ANOVA

	Treatments Design	**Error control Design**	**Analysis**
Completely Randomized	1 factor, (k levels); replicates/level='r'	Assign treatments 1 ... k to EU	1-way ANOVA
Completely randomized block design	1 factor (k levels); 1 block (l level), k*l EU	Assign treatments 1 ... k to EU, in each block	2-way ANOVA without interaction
Two-factor design with interaction	2 factors (k, l levels); r replicates of each combination	Assign each combination 1 ... k*l of treatments to EU	3-way ANOVA with interaction
Repeated Measures Design	1 factor (k levels); l repeated measures; replicates/combination = 'r'	Assign treatments 1 ... k to individuals at time 1	ANOVA of repeated measures

Completely randomized design

- Gene therapy experiment: compare four techniques to correct faulty genes
    - A: Normal gene inserted in a nonspecific location.
    - B: Abnormal gene exchanged for a normal gene.
    - C: Abnormal gene repaired by selective reversion mutation.
    - D: Regulation of a particular altered gene.
- 20 genetically identical and modified mice, affected by the disease to be treated, are selected.
- Treatments are randomized between the mice.
- The response variable is *gene expression*.
    - May be for a single or for multiple genes.

Completely randomized design

- The simplest design, suitable for comparing several treatments on a homogeneous sample.
- Randomization is performed by randomly assigning each of the 1 .... k treatments to individuals out of a total of N = k * r
- The basic linear model for one-factor experiments is as follows:

$$Y_{ij} = \mu_i + e_{ij} = \mu + \tau_i + e_{ij}, \quad i = 1 \ldots k, \quad j = 1 \ldots r.$$

- The analysis will usually be carried out by means of a one-way analysis of variance (ANOVA).

Randomization in a DCA

- There are many libraries that allow randomization, but it can also be done easily with a small script.
- Randomization is carried out *before* the experiment and it only indicates which treatment will receive each experimental unit
- Once the experiment is carried out, it is usual to present the data ordered by the treatments received, which *eliminates the evidence* that the assignment has been made randomly.

```
> n <- 20
> k <- 4
> numRep <- n/k
> trat <- rep(LETTERS[1:4], numRep)
> aleat <- sample(trat, n, replace=FALSE)
> asignacion <- paste (paste0('UE',1:n),aleat, sep='-')
> print(asignacion)
 [1] "UE1-A"  "UE2-C"  "UE3-B"  "UE4-C"  "UE5-B"  "UE6-D"  "UE7-A"
 [8] "UE8-C"  "UE9-A"  "UE10-A" "UE11-D" "UE12-B" "UE13-C" "UE14-C"
[15] "UE15-D" "UE16-D" "UE17-B" "UE18-D" "UE19-B" "UE20-A"
```

Random block design

- After exposure to a poison, cells can be treated by different substances that accelerate regeneration.
- A study wants to compare six of these growth factors (5 are treatments and 1 is a control).
- A problem has caused that there is not enough culture medium to grow all treatments with replicates. Instead, there are 4 culture media available.
- Since a complete randomization is not possible, it is decided to *block by type of culture medium*.
    - We prepare 4 groups of 6 plates, each group of a type of culture
    - Within each group a different treatment is randomly assigned to each of the six plates.

Random block design

- The completely randomized design loses utility if the experimental material is not homogeneous.
- Her, we can apply local control (blocking) and divide the experimental material in homogeneous subgroups, which we will call blocks.
- Distribute samples among blocks and apply treatments to experimental units randomly and independently of the other blocks.
- This design is called *Random Block Design (RBD)*.
- The linear model that describes the experiment is the following:

$$Y_{ij} = \mu + \rho_i + \tau_j + e_{ij}, i = 1 \dots k,\ j = 1 \dots l.$$

- The analysis will usually be carried out by means of an analysis of variance (ANOVA) of two factors without interaction.
- If it is not possible to distribute samples evenly between blocks, we are often faced with *unbalanced designs*

Block or randomize?

- *Block what you can and randomize what you cannot - Box, Hunter & Hunter (1978)*
- Randomization provides a rough balance between variables that have not been taken into account.
- Local control eliminates the effect of differences between blocks, thereby ensuring that differences between treatments cannot be due to differences between blocks.

Factorial design

- A study was conducted to study the effect of a drug and a diet on systolic blood pressure.
- 20 people with high blood pressure were randomized to one of four treatment conditions.
    - Control group (neither diet nor drug modification)
    - Diet modification only
    - Drug only
    - Modification of both drugs and diet
- At the end of the treatment period, systolic blood pressure was assessed.
- It is a factorial design in which each of the two treatments (drug, diet) can be randomly assigned to each individual.
- By having 20 individuals, there can be replicates of each treatment combination.

Factorial design

- Useful to study the effects of several factors simultaneously.
- "Treatments" are *all combinations of the different factors* under study.
- Randomization similar to completely randomized design: each combination of treatments randomly assigned to independent *r* EUs.
- The fact that each combination is replicated makes it possible to study, not only the effects of each factor separately, but also the interaction between them.
- The linear model that describes a two-factor design with interaction with *t* and *s* levels and *r* replicates respectively is the following:

$$Y_{ijk} = \mu + \rho_i + \tau_j + \tau\rho_{ij} + e_{ijk},\ i = 1 \dots t,\ j = 1 \dots l,\ k = 1 \dots r.$$

- The analysis will usually be carried out by means of an analysis of variance (ANOVA) of two factors with interaction.

Repeated measures design

- A study wanted to measure the concentration of certain metabolites in plasma after two dietary interventions consisting of adding an amount of olive oil or an equivalent amount of walnuts to the standard diet.
- 21 mice subjected to the same diet were taken and an intervention (water, olive oil or nuts) was randomly assigned.
- The concentration of the metabolite in blood was measured after before the intervention and at 24h, 48h and one week.

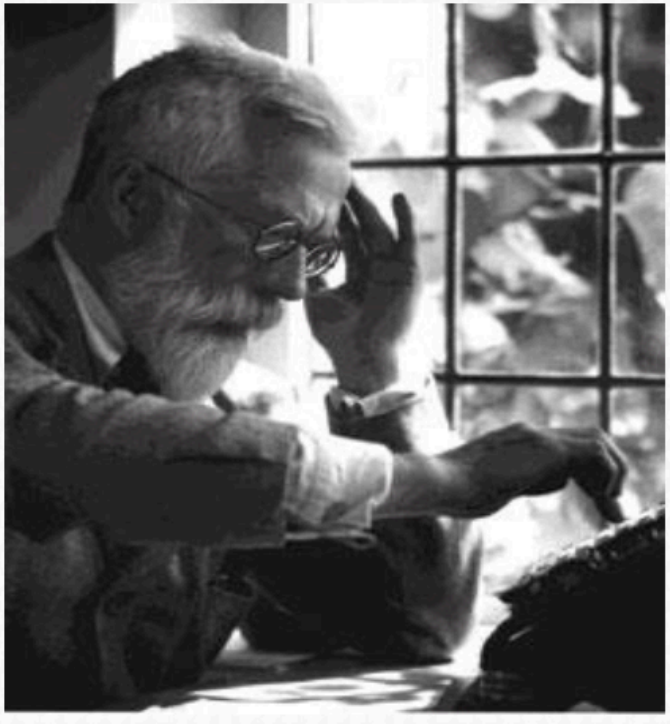
Repeated measures design

- When we take more than one measurement in each experimental unit, we have a *within-subjects design*.
- In this case, the data have different characteristics from the previous ones.
    - The measurements taken on the same individual are correlated.
    - There is a new source of variation that must be taken into account: variability *within* subjects.
- Apart from this, they offer the same possibilities as with other designs, but with an additional source of variability, "time".
- The analysis of repeated measures data is a whole world. Although the *ANOVA of repeated measures* is traditionally used, the current trend is to perform the analyzes using *linear mixed models* which are much more flexible.

Summarizing ...

- A good experimental design is essential to carry out good experiments.
- Experimental design means *planning in advance*, that is, before and not after the experiment.
- The experimental design must consider all steps: from sampling to data analysis.
- Applying grounded principles such as *randomization*, *replication* and *local control* is key to obtain good experimental designs.
- The analysis of designed experiments is carried out with the Analysis of the Variance (ANOVA). While each design can be asociated with an ANOVA model they should not be confused.
- Whenever possible we should have statistical support *from the beginning of the study*

And, as the master said ...



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.

References and resources

References and resources

- An introduction to Experimental Design
  - The github repository for these materials with some complements including R code examples.
- The three Rs
  - A short and interactive introductory course on the design of experiments focused on the benefits derived from an adequate design for the reduction of suffering in experimental animals.
- A First Course in Design and Analysis of Experiments
  - A book for an introductory course to design of experiments that the author decided to provide freely on the internet.
  - The author updated the examples by implementing them in R.