A *Short* Introduction to Experimental Design

January 9, 2020
Helpful documents

- Statistiques pour Statophobes – Denis Poinsot (link)
- Points of significance – Nature website (link)
- Seeing Theory - A visual introduction to probability and statistics (link)
Why statistics ? Because …

• It takes into account the random nature of objects
• In biology every individual is different
• It helps making decisions on objective criteria rather than on a subjective perception

• ... think about girafes
Different aspects of a statistical study

1. Collecting data
   Sampling, experiment design

![Diagram of population and sample](image.png)

**Figure 1.1** Les méfaits des fluctuations d'échantillonnage. A: Deux échantillons, même fort différents, ne proviennent pas nécessairement de deux populations différentes.
B: Deux échantillons, même fort semblables, ne proviennent pas nécessairement de deux populations semblables.
Different aspects of a statistical study

2. Exploratory analysis
   Data structure, suggest hypotheses
Different aspects of a statistical analysis

3. Inferential statistics (statistical induction)
Hypothesis testing, estimation

Random sampling

Blue: 1/6 (8/41)
Black: 1/6 (5/41)
Red: 2/6 (6/41)
Green: 2/6 (22/41)
Different aspects of a statistical study

4. Statistical modeling
   Regression, Classification (machine learning),...
Measuring differences

Experimental design is necessary when one aims at measuring the effect of a given factor.

Effect of an experimental factor is measured with respect to a control. The point of the experiment is to measure the difference between a test condition and control.

Underlying assumption: the observed difference is attributed to the factor under study.

The experiment must be designed so that the effect observed is truly due to the test factor and not to other unwanted sources of variation.
Let’s sketch the problem...

Rearing → Experiment → Sample → Sequencing

Data analysis
Let’s sketch the problem...

1. Design of experiments
2. Suited statistical models
3. Power analysis
Let’s sketch the problem...

1. Experiment design
2. Suited statistical models
3. Power analysis

Data analysis

Various batch effects
Let’s sketch the problem...

1. Experiment design
2. Suited statistical models
3. Power analysis

Number of individuals
An important feature of lab-reared material

Cage effect may be huge!

Example: proportion of variance explained by the cage effect (real data)

at different times

in different organs

source: data from Pasteur anonymised
1. Design of experiments
The Importance of Experimental Design

Let's see if the subject responds to magnetic stimuli... ADMINISTER THE MAGNET!

Interesting... there seems to be a significant decrease in heart rate. The fish must sense the magnetic field.
Not a recent idea!

Ronald A. Fischer, 1890 - 1962

Design of Experiments (1935)
The lady tasting tea

She claimed she was able to tell whether the milk or the tea was poured first. R. Fisher designed an experiment to test whether she was right.
You said “experimental design”...

Why? Ensuring that the experiment actually addresses the biological question

What? Detailed description of each experimental step in order to minimize unwanted experimental effects with respect to the biological effects of interest

How? Answering the following:
1. What is the exact biological question?
2. How to optimally capture biological variation bound to this specific question?
3. How to control/minimise unwanted sources of variation
Outline

Addressing a biological question: different types of experiment design

Capturing biological variation: through replication

Avoiding confounding effects: randomisation and blocking

A word on pooling: definition and good practice
Basic: comparing 2 conditions (unpaired data)

I want to compare the viral load in two strains of virus (WT and mutant)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>WT</td>
</tr>
<tr>
<td>S2</td>
<td>WT</td>
</tr>
<tr>
<td>S3</td>
<td>WT</td>
</tr>
<tr>
<td>S4</td>
<td>mutant</td>
</tr>
<tr>
<td>S5</td>
<td>mutant</td>
</tr>
<tr>
<td>S6</td>
<td>mutant</td>
</tr>
</tbody>
</table>

This experiment involves:
- one **factor** of interest: the virus strain
- this factor has two **levels**: WT and mutant
Comparing 2 paired treatments (paired data)

I want to compare a damaged skin with a normal one on mice. Each pair of tissues is extracted from the same mouse.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Treatment</th>
<th>Mouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>control</td>
<td>m1</td>
</tr>
<tr>
<td>S2</td>
<td>control</td>
<td>m2</td>
</tr>
<tr>
<td>S3</td>
<td>control</td>
<td>m3</td>
</tr>
<tr>
<td>S4</td>
<td>wound</td>
<td>m1</td>
</tr>
<tr>
<td>S5</td>
<td>wound</td>
<td>m2</td>
</tr>
<tr>
<td>S6</td>
<td>wound</td>
<td>m3</td>
</tr>
</tbody>
</table>
Time course experiment (paired)

I want to find genes that are differentially expressed between time 0 and time 24h on cultures of *E. Coli*
Time course experiment (paired)

I want to find genes that are differentially expressed between time 0 and time 24h on cultures of *E. Coli*

**Same plate at different time points**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Time</th>
<th>Plate nb</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0h</td>
<td>P1</td>
</tr>
<tr>
<td>S2</td>
<td>0h</td>
<td>P2</td>
</tr>
<tr>
<td>S3</td>
<td>0h</td>
<td>P3</td>
</tr>
<tr>
<td>S4</td>
<td>24h</td>
<td>P1</td>
</tr>
<tr>
<td>S5</td>
<td>24h</td>
<td>P2</td>
</tr>
<tr>
<td>S6</td>
<td>24h</td>
<td>P3</td>
</tr>
</tbody>
</table>

**Different plates**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Time</th>
<th>Plate nb</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0h</td>
<td>P1</td>
</tr>
<tr>
<td>S2</td>
<td>0h</td>
<td>P2</td>
</tr>
<tr>
<td>S3</td>
<td>0h</td>
<td>P3</td>
</tr>
<tr>
<td>S4</td>
<td>24h</td>
<td>P1</td>
</tr>
<tr>
<td>S5</td>
<td>24h</td>
<td>P2</td>
</tr>
<tr>
<td>S6</td>
<td>24h</td>
<td>P3</td>
</tr>
</tbody>
</table>


Time course experiment (unpaired)
Time course experiment (unpaired)

I want to find genes that are differentially expressed between time 0 and time 24h on cultures of *E. Coli*

<table>
<thead>
<tr>
<th>Sample</th>
<th>Time</th>
<th>Plate nb</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0h</td>
<td>P1</td>
</tr>
<tr>
<td>S2</td>
<td>0h</td>
<td>P2</td>
</tr>
<tr>
<td>S3</td>
<td>0h</td>
<td>P3</td>
</tr>
<tr>
<td>S4</td>
<td>24h</td>
<td>P1</td>
</tr>
<tr>
<td>S5</td>
<td>24h</td>
<td>P2</td>
</tr>
<tr>
<td>S6</td>
<td>24h</td>
<td>P3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>Time</th>
<th>Plate nb</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0h</td>
<td>P1</td>
</tr>
<tr>
<td>S2</td>
<td>0h</td>
<td>P2</td>
</tr>
<tr>
<td>S3</td>
<td>0h</td>
<td>P3</td>
</tr>
<tr>
<td>S4</td>
<td>24h</td>
<td>P4</td>
</tr>
<tr>
<td>S5</td>
<td>24h</td>
<td>P5</td>
</tr>
<tr>
<td>S6</td>
<td>24h</td>
<td>P6</td>
</tr>
</tbody>
</table>
Paired vs. unpaired: impact on the analysis
Paired vs. unpaired: impact on the analysis
Paired vs. unpaired: impact on the analysis
Crossed factors designs

Effect of a treatment on the viral load in two different mouse strains (B6 vs SEG)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Treatment</th>
<th>Mouse strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>no</td>
<td>B6</td>
</tr>
<tr>
<td>S2</td>
<td>no</td>
<td>B6</td>
</tr>
<tr>
<td>S3</td>
<td>no</td>
<td>B6</td>
</tr>
<tr>
<td>S4</td>
<td>yes</td>
<td>B6</td>
</tr>
<tr>
<td>S5</td>
<td>yes</td>
<td>B6</td>
</tr>
<tr>
<td>S6</td>
<td>yes</td>
<td>B6</td>
</tr>
<tr>
<td>S7</td>
<td>no</td>
<td>SEG</td>
</tr>
<tr>
<td>S8</td>
<td>no</td>
<td>SEG</td>
</tr>
<tr>
<td>S9</td>
<td>no</td>
<td>SEG</td>
</tr>
<tr>
<td>S10</td>
<td>yes</td>
<td>SEG</td>
</tr>
<tr>
<td>S11</td>
<td>yes</td>
<td>SEG</td>
</tr>
<tr>
<td>S12</td>
<td>yes</td>
<td>SEG</td>
</tr>
</tbody>
</table>
Crossed factors designs

Interaction:
Is the effect of treatment similar across strains?
Is the difference between strains consistent among treatments?
The viral load is higher in SEG than in B6 but the difference between SEG and B6 is constant irrespective of the treatment.
Crossed factors designs: reinforcement (treatment effect)

The difference in viral load between treatment and no treatment is higher for SEG than for B6.
Crossed factors designs: reduction (of the treatment effect)

The difference in viral load between treatment and no treatment is lower for SEG than for B6
Crossed factors designs: inversion (of the treatment effect)

The treatment reduces the viral load for B6 but it increases it for SEG
Nested designs

Used when it is impossible to cross all of the factors of interest in an experimental design.

E.g., when the factor is observational and/or cannot be manipulated
Nested design: take home

Unless you can apply both modalities of the treatment to individuals from a same cage (crossed design), you are dealing with nested designs. It crucially matters to statistical analysis.
To recap...

Express the biological question as accurately as possible:

- what is (are) the factor(s) of interest?
- for each factor of interest, how many levels are there?
- paired/unpaired design, nested design?
- If > 1 factor (complicated design), is there any interaction between the factors?

Golden rule: the simpler, the better. When the experiment includes more than two factors, the results may be very difficult to interpret (too many comparisons and potential interactions).
Outline

Addressing a biological question: different types of experiment design

Capturing biological variation: through replication

Avoiding confounding effects: randomisation and blocking

A word on pooling: definition and good practice
Biological and technical replicates

Sample 1

Sample 2

Technical replicates

Biological replicates

Sample 1

Sample 2

Technical replicates

Biological replicates
Biological and technical replicates

**Biological replicates**

Parallel measurements of biologically distinct *samples* that capture random biological variation, which may itself be a subject of study or a noise source.

**Technical replicates**

Repeated measurements on the same *sample* that represent independent measures of the random noise associated with protocols or equipment.

*Blainey et al. 2014 - Nature Methods*
Avoid biased sampling

It may occur that the healthiest mice are chosen preferentially for experiment. How are they assigned to a particular group?
Replication and pseudo-replication

**Replication**

Independent repetition of an experimental condition so that variation bound to the variable of interest may be estimated properly.

**Pseudo-replication**

Use of inferential statistics to test for treatment effects with data from experiments where either treatments are not replicated (though samples may be) or replicates are not statistically independent. (Hurlbert, 1984).
Replication and pseudo-replication

Identifying the level of (independent) experimental units
To recap...

Experiments must be replicated
to precisely measure the biological variability associated with
the condition under study.

Sampling must be representative
of the whole population under study

The higher the within group variability ...
the higher the number of replicates, in order to make sure that
the whole range of variation is covered (e.g. exp. with humans)
Outline

Addressing a biological question: different types of experiment design

Capturing biological variation: through replication

Avoiding confounding effects: randomisation and blocking

A word on pooling: definition and good practice
How to control for unwanted (technical or biological) sources of variation?

Avoid confounding effects between the biological factor of interest and other sources of variation

Ideally, use blocking...

to ensure that the biological conditions are evenly distributed among factors that are important (unwanted) sources of variability.

It helps to avoid confusion

or randomization

when blocking is not possible
## Technical confounding effects

Comparing the effects of two virus strains in mice (e.g. immune genes)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Treatment</th>
<th>RNA extraction date</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>strain WT</td>
<td>July 12th, 2018</td>
</tr>
<tr>
<td>S2</td>
<td>strain WT</td>
<td>July 12th, 2018</td>
</tr>
<tr>
<td>S3</td>
<td>strain WT</td>
<td>July 12th, 2018</td>
</tr>
<tr>
<td>S4</td>
<td>mutant</td>
<td>July 20th, 2018</td>
</tr>
<tr>
<td>S5</td>
<td>mutant</td>
<td>July 20th, 2018</td>
</tr>
<tr>
<td>S6</td>
<td>mutant</td>
<td>July 20th, 2018</td>
</tr>
</tbody>
</table>

Full association between Treatment and extraction date. Analysis cannot distinguish the effects of viral strain and dates of RNA extraction.
Technical confounding effects

Comparing the effects of two virus strains in mice (e.g. immune genes)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Treatment</th>
<th>RNA extraction date</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>strain WT</td>
<td>July 12th, 2018</td>
</tr>
<tr>
<td>S2</td>
<td>strain WT</td>
<td>July 20th, 2018</td>
</tr>
<tr>
<td>S3</td>
<td>strain WT</td>
<td>July 25th, 2018</td>
</tr>
<tr>
<td>S4</td>
<td>mutant</td>
<td>July 12th, 2018</td>
</tr>
<tr>
<td>S5</td>
<td>mutant</td>
<td>July 20th, 2018</td>
</tr>
<tr>
<td>S6</td>
<td>mutant</td>
<td>July 25th, 2018</td>
</tr>
</tbody>
</table>

A practical solution: evenly distributing the putative “date effect” across treatments.
Technical confounding effects

Comparing the effects of two virus strains in mice (e.g. immune genes)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Treatment</th>
<th>RNA extraction date</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>strain WT</td>
<td>July 12th, 2018</td>
</tr>
<tr>
<td>S2</td>
<td>strain WT</td>
<td>July 20th, 2018</td>
</tr>
<tr>
<td>S3</td>
<td>strain WT</td>
<td>July 25th, 2018</td>
</tr>
<tr>
<td>S4</td>
<td>mutant</td>
<td>July 12th, 2018</td>
</tr>
<tr>
<td>S5</td>
<td>mutant</td>
<td>July 20th, 2018</td>
</tr>
<tr>
<td>S6</td>
<td>mutant</td>
<td>July 25th, 2018</td>
</tr>
</tbody>
</table>

In case of paired data: the pairing may be confounded with the day effect. But those effects are NOT confounded with the biological effect of interest.
Examples of technical confounding effects
Focus on a particular case: RNA-seq

- Compare gene expression across conditions
- Multiplexing allows to sequence several samples on the same lane, depending on the genome size
- Bacterial genomes can be highly multiplexed (more than 10 samples per lane), human genomes can not (2 to 4 samples per lane)
- Technical sources of variability: lane, run, multiplexing rate (among others)
Focus on a particular case: RNA-seq

Taking into account the lane effect

Problem: lanes and treatment are confounded
Focus on a particular case: RNA-seq

Taking into account the lane effect

Problem: lanes and treatment are confounded
Solution: distribute the lane effect evenly on both conditions
Focus on a particular case: RNA-seq

Taking into account the lane effect: a tricky example

Problem: partial confusion between lane and condition
Focus on a particular case: RNA-seq

Taking into account the lane effect: a tricky example

Problem: confusion between lane and condition
Solution: distribute the lane and cage effects in all conditions
Randomisation

Whenever it is not possible to block (to distribute each condition evenly across all sources of technical variation), then randomize

Example
Consider the following RNA-seq experiment:
- 3 different biological conditions
- 7 time points
- 4 biological replicates

= 84 samples to be sequenced on 6 lanes (14 samples per lane).

How should I arrange the samples on the lanes?
Biological and technical confounding effects

Comparison of lung cells in healthy and cystic fibrosis patients

<table>
<thead>
<tr>
<th>id</th>
<th>status</th>
<th>age</th>
<th>gender</th>
<th>lib prep day</th>
<th>experimentalist</th>
</tr>
</thead>
<tbody>
<tr>
<td>s1</td>
<td>healthy</td>
<td>56</td>
<td>female</td>
<td>May 12th, 2019</td>
<td>John</td>
</tr>
<tr>
<td>s2</td>
<td>healthy</td>
<td>48</td>
<td>female</td>
<td>May 12th, 2019</td>
<td>John</td>
</tr>
<tr>
<td>s3</td>
<td>healthy</td>
<td>51</td>
<td>female</td>
<td>May 12th, 2019</td>
<td>John</td>
</tr>
<tr>
<td>s4</td>
<td>cf</td>
<td>17</td>
<td>male</td>
<td>July 12th, 2019</td>
<td>Jennyfer</td>
</tr>
<tr>
<td>s5</td>
<td>cf</td>
<td>22</td>
<td>male</td>
<td>July 12th, 2019</td>
<td>Jennyfer</td>
</tr>
<tr>
<td>s6</td>
<td>cf</td>
<td>23</td>
<td>male</td>
<td>July 12th, 2019</td>
<td>Jennyfer</td>
</tr>
</tbody>
</table>

Confusions between patient status, patient age, patient gender, day of library preparation and experimentalist
To recap...

There may be many unwanted sources of technical and biological variation interfering with the phenomenon of interest.

It is of primary importance to account for them when designing the experiment…

... in order to tell apart the biological effects from any other technical or biological factors.

Any doubts? Contact us!
Outline

Addressing a biological question: different types of experiment design

Capturing biological variation: through replication

Avoiding confounding effects: randomisation and blocking

A word on pooling: definition and good practice
Pooling

Pooling is used when one individual can not provide enough biological material to form a biological replicate. It consists in grouping several individuals for each independent biological sample.
Pooling

Pooling is used when one individual cannot provide enough biological material to form a biological replicate. It consists in grouping several individuals for each independent biological sample.
Pooling is averaging

Pooling individuals

- reduces the noise due to individual-specific features
- strengthens features that are in common to all individuals
- decreases the variation expressed within condition
  (i.e. variance across biological replicates)
Pooling: mean and variance

Reminder: mean and variance of a sample mean (central limit theorem)

Simulated distribution with $k = 10$ with samples (pools)
Each sample being a pool of $n = \{1, 2, 3\}$ individuals
Pooling: good practice

- Use pooling only when **it is necessary or when** you want to shrink the effects of inter-individual variation
- Include the same number of individuals in each pool (i.e., biological replicate, here)
Pooling: good practice

- Use pooling only when it is necessary or when you want to shrink the effects of inter-individual variation
- Include the same number of individuals in each pool (i.e., biological replicate, here)
2. Models

Cf. track *R programming and statistics*
3. Statistical power
Statistical power — the basics

The power of a statistical test is the probability of rejecting the null hypothesis when we should — that is, when the alternative hypothesis is actually true.

$$\text{power} = 1 - P(\text{type II error}) = 1 - \beta$$

The power is a probability and is very often expressed as a percentage.
Statistical power — the basics

The power of a statistical test is the probability of rejecting the null hypothesis when we should — that is, when the alternative hypothesis is actually true.

\[
power = 1 - P(\text{type II error}) = 1 - \beta
\]

The power is a probability and is very often expressed as a percentage.
The power increases with sample size

Increasing $n$ decreases the spread of the distribution of sample averages in proportion to $1/\sqrt{n}$

Krzywinski & Naomi Altman - Nat. Meth. 2013
The power increases with effect size $d$

Power increases with $d$, making it easier to detect larger effects

$d = (\mu_1 - \mu_0) / (\sigma)$

Krzywinski & Naomi Altman - Nat. Meth. 2013
Summary

Power, sample size, effect size relative to noise, and α level are interrelated and constrained by a mathematical relationship involving the four quantities.

This relationship is often complicated, and can’t always be explicitly written down as a formula, but it does exist.

For any statistical test, you can (at least in theory) determine any one of the four quantities if you know the other three.

In practice, you may not have access to some of these information (in particular to the effect size and the variance of the distribution). In addition you may have strong cost constraints: do your best!
3. Live experiment
Bransford and Jonhson (1972)

1. Take an image
2. Listen to the text
3. When prompted, rise 1 to 5 fingers according to how well you feel you can remember the text.
   ◆ 1 = not well at all
   ◆ 2 = badly
   ◆ 3 = some of the text but not all
   ◆ 4 = well
   ◆ 5 = very well
The context

How would you fix my badly flawed design?
4. Conclusion
"To consult a 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."

A nice reference dedicated to experimentalists

12 practical recommendations which cover 3 aspects/steps of the analysis
Design, Analysis and Reporting (reproducibility concerns)

Review Article

Guidelines on statistics for researchers using laboratory animals: the essentials

Romain-Daniel Gosselin1,2

Abstract
There is growing concern that the omnipresence of flawed statistics and the deficient reproducibility that arises therefrom results in an unethical waste of animals in research. The present review aims at providing guidelines in biostatistics for researchers, based on observed frequent mistakes, misuses and misconceptions as well as on the specificities of animal experimentation. Twelve recommendations are formulated that cover sampling, sample size optimisation, choice of statistical tests, understanding p-values and reporting results. The objective is to expose important statistical issues that one should consider for the correct design, execution and reporting of experiments.

Keywords
Experimental design, statistics, sample size, reproducibility, guidelines
Thank you for your attention!