Day 1
Experimental design
Anne Segonds-Pichon
v2019-11
• **Universal principles**
  - The same-ish questions should always be asked
    - What is the question?
    - What measurements will be made?
    - What factors could influence these measurements?
  - But the answers/solutions will differ between areas

• **Examples:**
  - **Experimental design** will be affected by the question
    - but also by practical feasibility, factors that may affect causal interpretation ...
    - e.g. number of treatments, litter size, number plants per bench ...
  - **Sample size** will be affected by ethics, money, model ...
    - e.g. mouse/plant vs. cell, clinical trials vs. lab experiment ...
  - **Data exploration** will be affected by sample size, access to raw data ...
    - e.g. >20,000 genes vs. weight of a small sample of mice
Vocabulary, tradition and software

- People use different words to describe the same data/graphs …
- There are different traditions in different labs, areas of science …
- Different software mean different approaches: R, SPSS, GraphPad, Stata, Minitab …
- Examples:
  - Variable names: qualitative data = attribute
  - Scatterplots in GraphPad Prism = stripchart in R
  - 2 treatment groups in an experiment = 2 arms of a clinical trial
  - Replicate = repeat = sample
  - QQ plots in SPSS versus D’Agostino-Pearson test …
  - Sample sizes

- Very different biological questions, very different designs, sophisticated scientific approach or very simple
  - Similar statistical approach
  - Example:
    - **Data:** Gene expression values from The Cancer Genome Atlas for samples from tumour and normal tissue, **question:** which genes are showing a significant difference? **t-test**
    - **Data:** weight from WT and KO mice, **question:** difference between genotypes? **t-test**
Experimental Design

Statistical Analysis

Type of Design

Common Sense

Technical vs. Biological
• Translate the hypothesis into statistical questions
  • Think about the statistical analyses before you collect any data

• What data will I collect?

• How will it be recorded/produced?

• Will I have access to the raw data?

• I have been told to do this test/use that template, is that right?

• Do I know enough stats to analyse my data?
  • If not: ask for help!
• **Example:**

  • **Hypothesis:** exercise has an effect on neuronal density in the hippocampus.

  • **Experiment:** 2 groups of mice on 2 different levels of activity:
    • No running or running for 30 minutes per day
    • After 3 weeks: mice are euthanized and histological brain sections are prepared
      • Neuronal density by counting the number of neurons per slide

  • **Stats:** one factor: activity and one outcome: number of neurons
**Experiment**: exercise has an effect on neuronal density in the hippocampus

Difference between the groups

- Correlation?
  - Parametric
    - Pearson Correlation
  - Nonparametric
    - Spearman Rank Correlation
- Categories?
  - Parametric
    - Chi Square test
- How many factors?
  - One
- Two or more
  - 2 way ANOVA, General Linear (Mixed) Model, etc.
  - Same or different subjects?
    - Same
      - Parametric
        - Paired T-test
      - Nonparametric
        - Wilcoxon signed rank test
    - Different
      - Parametric
        - F-test / ANOVA
      - Nonparametric
        - Manh Whitney U

2 different groups of mice

Neurons counts: Normality Homogeneity of variance

That’s the one!
Experimental Design

- Statistical Analysis
- Type of Design
- Common Sense
- Technical vs. Biological
• **Experimental unit**: cell, tissue sample, leaf, mouse, plant, litter ...
  • Neuronal density experiment: **experimental unit**: mouse

• **Factor**:
  • Fixed factor: factor of interest, predictor, grouping factor, arm in controlled trial, independent variable ...
    • e.g. : treatment, gender, genotype ...
    • Neuronal density experiment: **fixed factor**: running
  
  • Random factor: factor we need to account for, blocking factor, nuisance factor ...
    • e.g. : experiment, batch, plate, lanes ...
    • Neuronal density experiment: **uh oh**

• **Key concepts**:
  • Blinding: not always possible, single and double-blinding
  • Randomisation
Experimental Design

**Type of design**

**Completely random**

CRD

Simplest: experimental units randomly allocated to groups
e.g.: treatment ...

**Complete Randomised block**

CRBD

Accounting for random factors, nuisance variables
e.g.: batch effect, experimental effect, day-to-day variation ...

**Split-plot**

Also nested design, repeated measures
e.g.: several measures per animal, several treatments per plot, pups in a litter...
**Experimental Design** → **Type of design**

**Completely random**

**CRD**

**Complete Randomised block**

**CRBD**

<table>
<thead>
<tr>
<th>Control</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse 1</td>
<td>Mouse 6</td>
</tr>
<tr>
<td>Mouse 2</td>
<td>Mouse 7</td>
</tr>
<tr>
<td>Mouse 3</td>
<td>Mouse 8</td>
</tr>
<tr>
<td>Mouse 4</td>
<td>Mouse 9</td>
</tr>
<tr>
<td>Mouse 5</td>
<td>Mouse 10</td>
</tr>
</tbody>
</table>

**Good design:**
GenADA multi-site collaborative study 2010
Alzheimer’s study on 875 patients

**Bad design**
Differences between Control, Treatment 1 and Treatment 2 are confounded by day and plate.

**Controls and Cases**
Plate effects by case/control

http://blog.goldenhelix.com/?p=322
**Experimental Design**

**Type of design**

**Complete Randomised block**

- **RNA-Seq experiments**: multiplexing allows for randomization
  - Multiplexing: barcodes attached to fragments
  - Barcodes: distinct between libraries (samples)
  - **Important**: identify the sources of noise (nuisance variable)
    - Library preparation: big day-to-day variability
      - **Batch effect**
    - Big variability between runs
    - **Lane effect**

---

**Auer and Doerge, 2010**
**Incomplete Randomised block**

- **RNA-Seq experiments:**
  - Incomplete block design:
    - All treatments/samples are not present in each block
  - **Balanced Incomplete Block Design (BIBD):**
    - where all pairs of treatments/samples occur together within a block an equal number of times

- Statistical analysis:
  - account for missing values
  - e.g.: a model fits blocks then samples

---

**Type of design**

*Six samples*

- S1, S2, S3, S4, S5, S6

*Five samples per lanes*

- Lane 1: S1, S3, S4, S5, S6
- Lane 2: S2, S3, S4, S5, S6
- Lane 3: S1, S2, S4, S5, S6
- Lane 4: S1, S2, S3, S5, S6
- Lane 5: S1, S2, S3, S4, S5
- Lane 6: S1, S2, S3, S4, S5
Split-plot

- from agriculture: fields are **split** into **plots** and subplots.

**Example:** *in vivo* effect of a drug on gene expression on 2 tissues.

**Experimental Design**

- **Type of design**

**Split plot**

- Mouse/Plant = plot = random factor
- Tissue = subplot (split-plots) = nested factor

**Housing unit/Bench** = blocking factor = nuisance factor

Krzywinski and Altman, 2015
More complex design:

- **Split-plot + Completely Random Design**: commonly used for repeated measures designs.
• Other designs: crossover, sequential ....

**Factorial Design**: more an arrangement of factors than a design

• When considering more than one factor

• Back to our neuronal density experiment: exercise has an effect on neuronal density in the hippocampus

![Running | Not running
n mice | n mice](image)

**Completely random**

• Not enough: we want to account for:
  • **Sex**: factor of interest: **factorial design** (2 factors: running and sex)
  • **Experimental variability**: random factor: **blocking factor** (one experiment = one block)
  • **Several histological slides**: **nested variable**
- Neuronal density experiment: Complete Randomised block design + Split-plot

### Experimental Design

<table>
<thead>
<tr>
<th>Random factor</th>
<th>Experiment</th>
<th>Blocks</th>
<th>Fixed factor</th>
<th>Interest</th>
<th>Fixed factor</th>
<th>Interest</th>
<th>Fixed factor</th>
<th>Interest</th>
<th>Fixed factor</th>
<th>Biological</th>
<th>Nested variable</th>
<th>Biological</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Histological slides</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S1</td>
<td>$2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S3</td>
<td>$4$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S5</td>
<td>$6$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S7</td>
<td>$8$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S9</td>
<td>$10$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S11</td>
<td>$12$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S13</td>
<td>$14$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S15</td>
<td>$16$</td>
</tr>
</tbody>
</table>

If samples from 2 different tissues

- Rule of thumb: Block what you can, randomize what you cannot
  - **Blocking** is used to remove the effects of a few of the most important nuisance variables (known/controllable)
  - **Randomisation** is then used to reduce the contaminating effects of the remaining nuisance variables (unknown/uncontrollable, lurking).
- Drawing the experimental design can help!
Two factors of interest per experiment:
Activity and Sex

**Experiment**: exercise has an effect on neuronal density in the hippocampus

That’s the one!
• Statistical tests are tools used to quantify our level of confidence in what we see.
Statistical Analysis

• Statistical tests are tools
  • How do we choose the right tool?
    
    - Differences?
    - How many factors?
    
    - Two or more
      - 2 way ANOVA, General Linear (Mixed) Model, etc.
    - One
      
      - Same or different subjects?
        
        - Same
          - Correlation
            • Pearson Correlation
        - Different
          
          - Pearson correlation
            • Nonparametric Spearman Rank Correlation
  
  - Correlation?
    - Parametric
      - Chi square test
    - Nonparametric
  
  - Categories?
    - Parametric
      - Paired T-test / repeated ANOVA
    - Nonparametric
      
      - Wilcoxon
        • Mann Whitney U test
  
  
  - Data
    
    • Nature and behaviour of the data:
      • All statistical tests are associated with assumptions
        • e.g. normality and homogeneity of variance
      • If assumptions not met: bad p-values
    
    • Running a statistical test is easy
      • but making sure it’s the right test is not.
  
  • Getting to know the data:
    • Data exploration
    • But also if not one’s data:
      • raw or not raw?
      • If normalised/standardised, how?
      • e.g raw counts (qualitative data) vs. normalised (quantitative)
Definition of technical and biological depends on the model and the question
  • e.g. mouse, cells ...

Question: Why replicates at all?
  • To make proper inference from sample to general population we need biological samples.

Example: difference on weight between grey mice and white mice:
  • cannot conclude anything from one grey mouse and one white mouse randomly selected
    • only 2 biological samples
  • need to repeat the measurements:
    • measure 5 times each mouse: technical replicates
    • measure 5 white and 5 grey mice: biological replicates

Answer: Biological replicates are needed to infer to the general population
Definition of **technical** and **biological** depends on the model and the question.

The model: mouse, plant ... complex organisms in general.
- Easy: one value per individual organism
  - e.g. weight, neutrophils counts ...

**What to do?** Mean of technical replicates = 1 biological replicate
• The model is still: mouse, plant … complex organisms in general.
  • Less easy: more than one value per individual
    • e.g. axon degeneration

• **What to do?** Not one good answer.
  • In this case: mouse = experiment unit (block, split-plot)
    • axons = technical replicates, nerve segments = biological replicates
The model is: worms, cells ...
  • Less and less easy: many ‘individuals’
    • What is ‘n’ in cell culture experiments?

• Cell lines: no biological replication, only technical replication

• To make valid inference: valid design

Vial of frozen cells

Control  Treatment

Dishes, flasks, wells ...
Cells in culture

Point of Treatment

Glass slides
microarrays
lanes in gel
wells in plate

Point of Measurements

Always easy to tell the difference?
• **Design 1:** As bad as it can get

![Diagram of cell count process](image)

• After quantification: 6 values
  • But what is the sample size?
    • $n = 1$
      • no independence between the slides
      • variability = pipetting error
Technical vs. Biological

**Always easy to tell the difference?**

- **Design 2**: Marginally better, but still not good enough

  ![Diagram](image)

  - Everything processed on the same day

- After quantification: 6 values
  - But what is the sample size?
    - **n = 1**
      - no independence between the plates
      - variability = a bit better as sample split higher up in the hierarchy
**Design 3:** Often, as good as it can get

- After quantification: 6 values
- But what is the sample size?
  - $n = 3$
  - Key difference: the whole procedure is repeated 3 separate times
  - Still technical variability but done at the highest hierarchical level
  - Results from 3 days are (mostly) independent
  - Values from 2 glass slides: paired observations
• **Design 4:** The ideal design

![Diagram showing Design 4: The ideal design](image)

• After quantification: 6 values
  • But what is the sample size?
    • \( n = 3 \)
      • Real biological replicates
Technical and biological replicates
What to remember

• Take the time to identify technical and biological replicates

• Try to make the replications as independent as possible

• Never ever mix technical and biological replicates

• The hierarchical structure of the experiment needs to be respected in the statistical analysis (nested, blocks ...).
• Design your experiment to be analysable
• The gathering of results or carrying out of a procedure is not the end goal
  • Think about the analysis of the data and design the experiment accordingly
• Imagine how your results will look
• Ask yourself whether these results will address your hypothesis
• Don’t get fixated on being able to perform a cool technique or experimental protocol.
• Don’t be overwhelmed (or try not to be).
• **Draw your experiment and imagine all that can go wrong at each step**