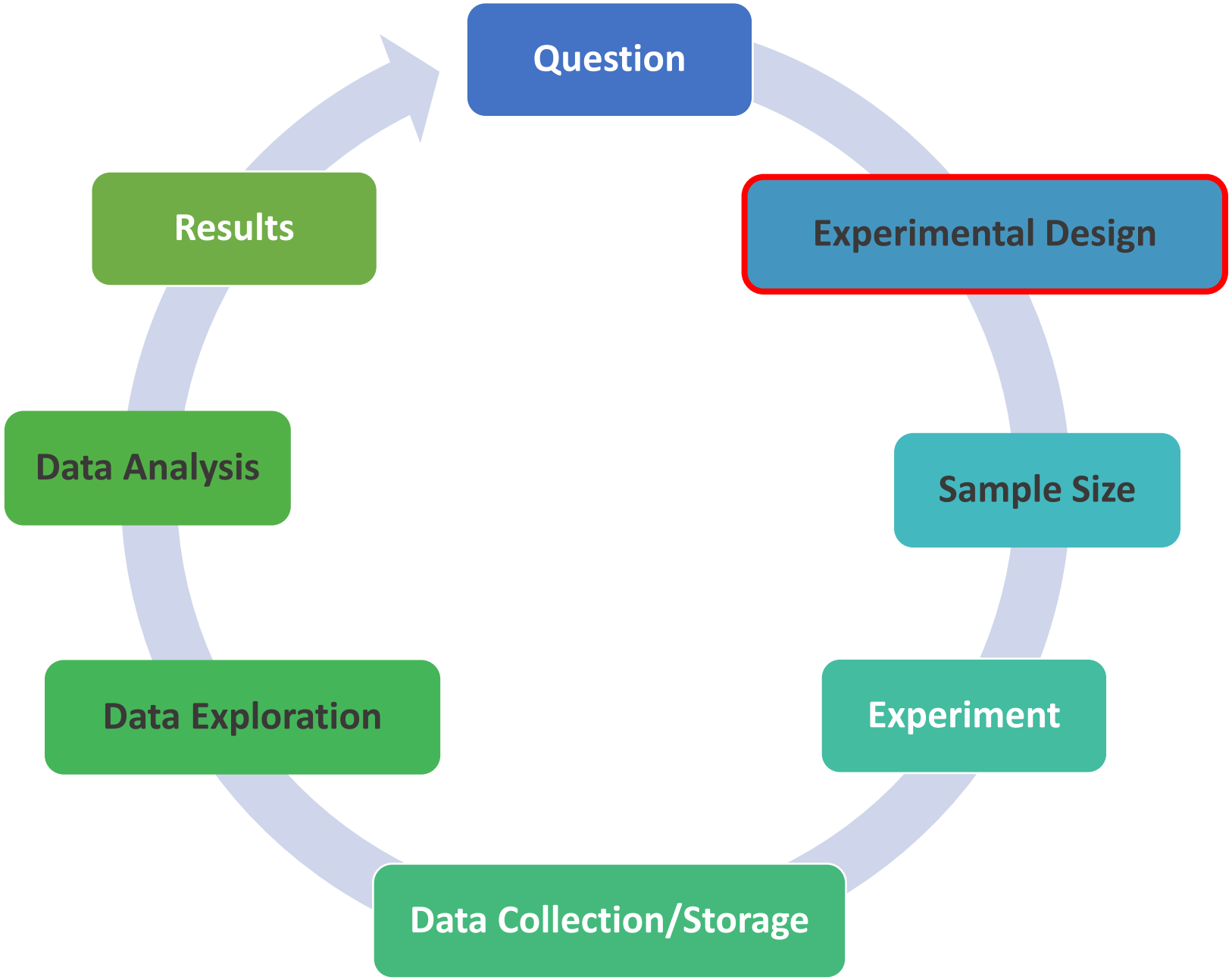
Babraham Institute

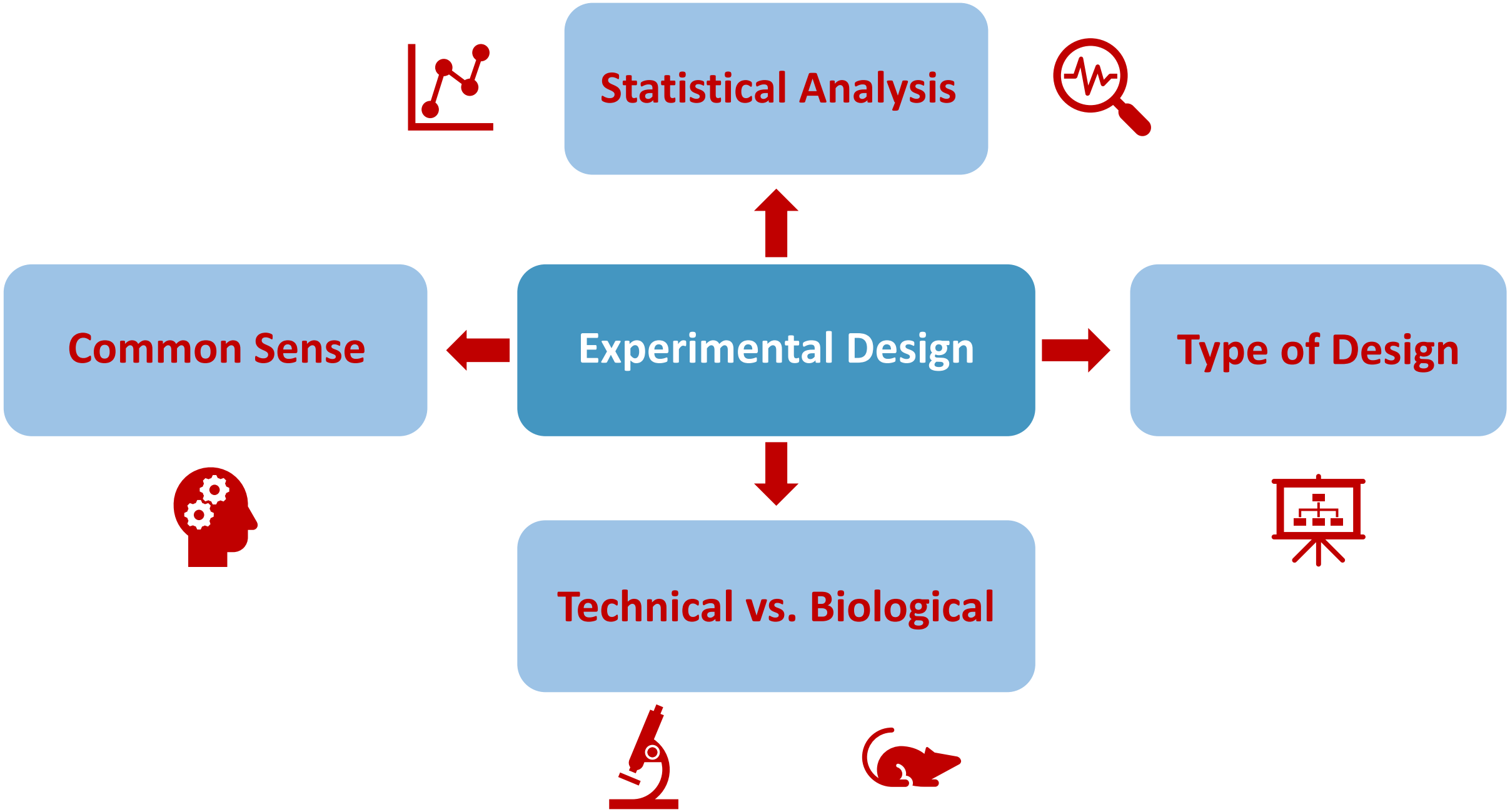
# Experimental design

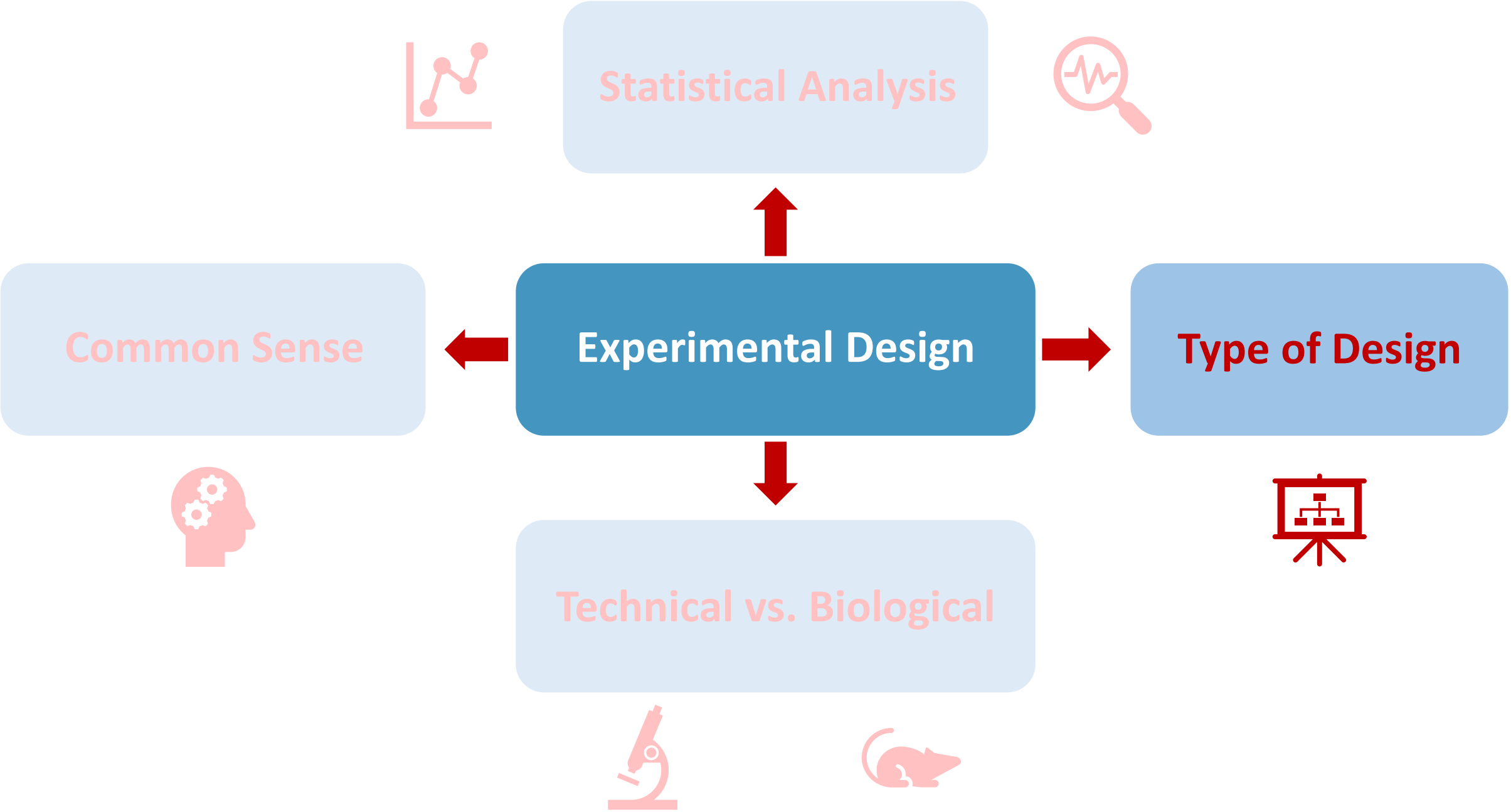
Hayley Carr & Anne Segonds-Pichon

v2025-01

Babraham Bioinformatics


Question
Experimental Design
Sample Size
Experiment
Data Collection/Storage
Data Exploration
Data Analysis
Results


Statistical Analysis
Common Sense
Experimental Design
Type of Design
Technical vs. Biological


Statistical Analysis
Common Sense
Experimental Design
Type of Design
Technical vs. Biological

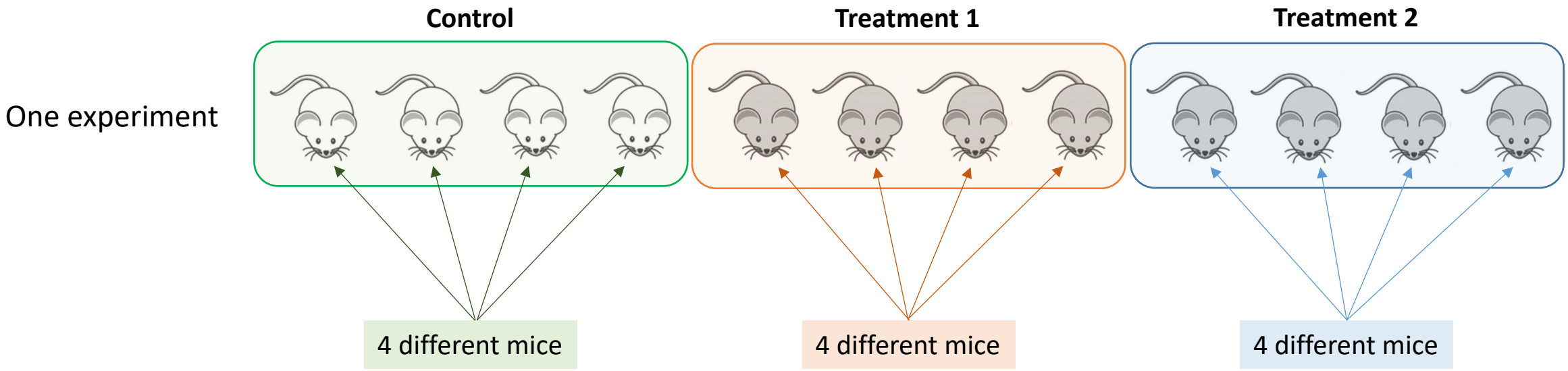
# Experimental design

## Independent versus matched design

# Experimental design

## Independent design

- 2 or more groups in an experiment with **independent** subjects
- **Example**: 3 groups with n=4 in the control group and n=4 in each treated group

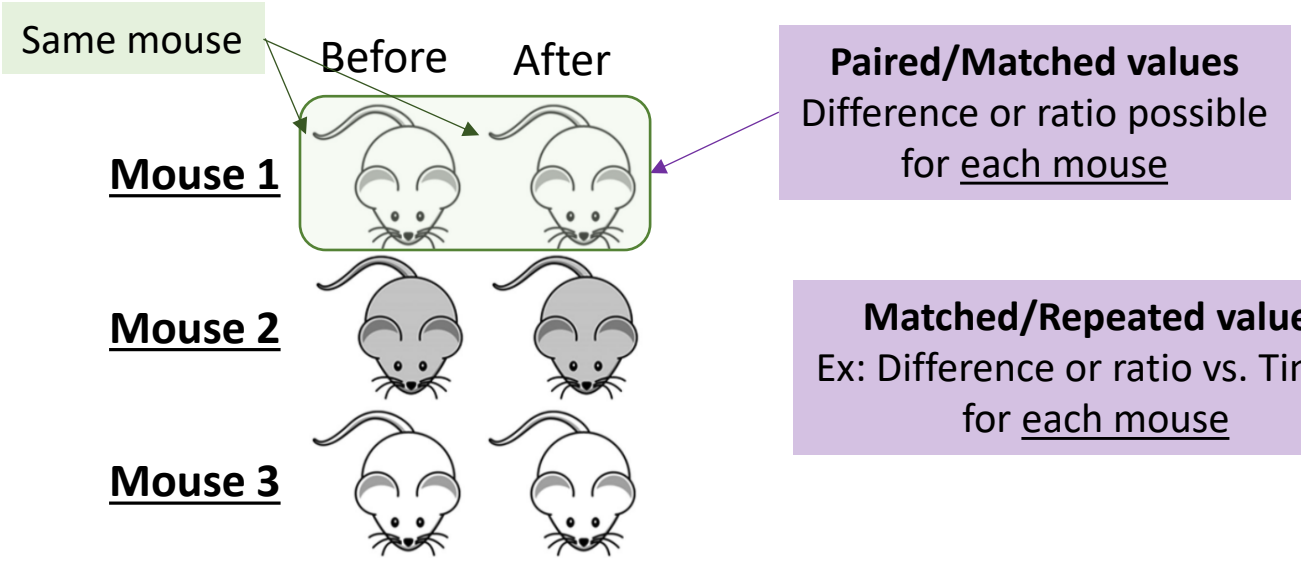


# Experimental design

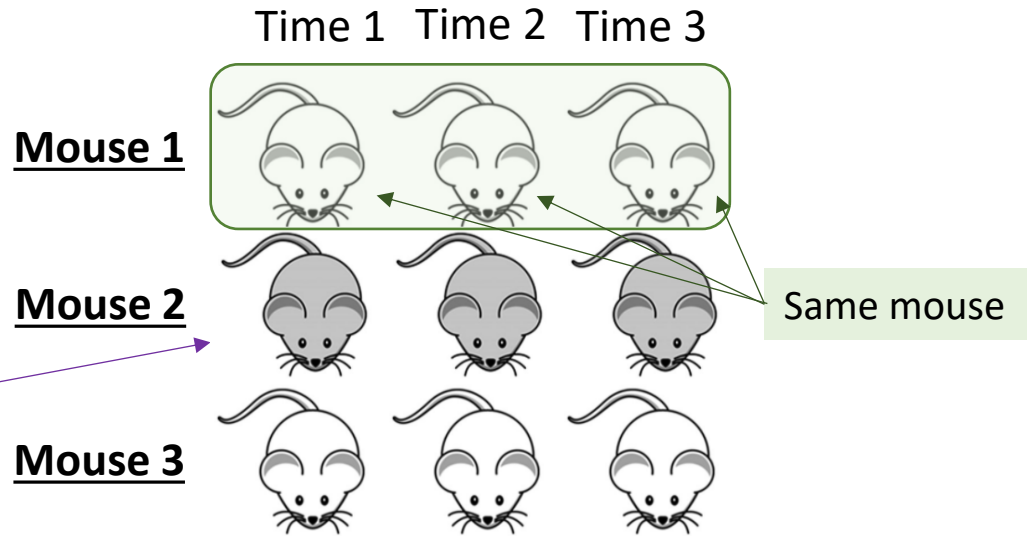
## Matched design

- Also called repeated = dependent = paired (2 groups)
  - **Design 1**: ≥2 measures per animal/subject/petri dish

**Example 1**: before/after treatment measures




**Example 2**: 3 time points



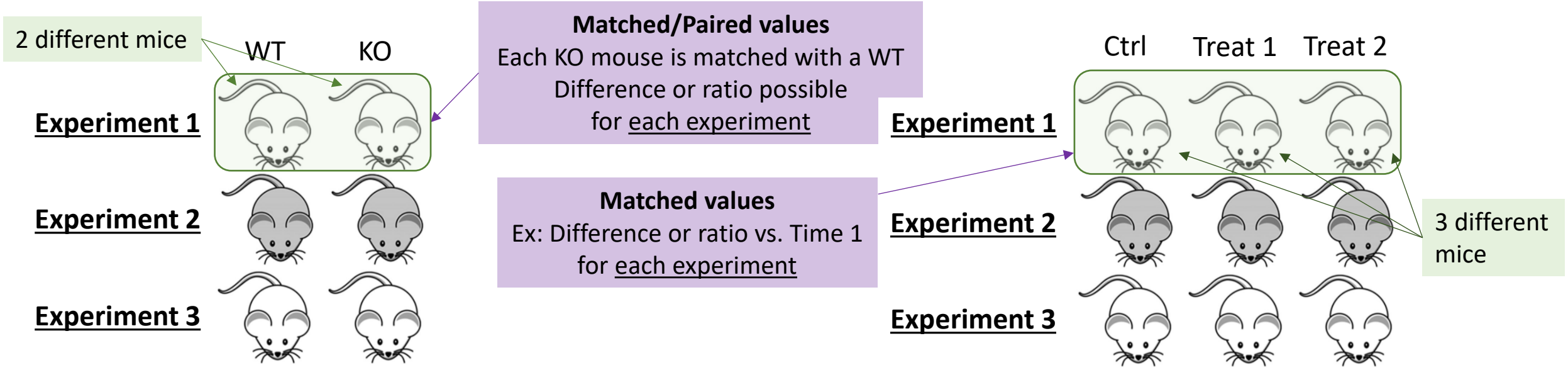
# Experimental design

## Matched design

- **Design 2**: experiment repeated independently

**Example 1**: 3 independent experiments
2 mice within each: WT and KO

**Example 2**: 3 independent experiments
3 mice within each: control and 2 treatments



# Experimental design

## Other design considerations: bias

*The lack of bias-reducing measures such as randomisation and blinding can contribute to as much as 30-45% inflation of effect sizes*

- Simple **randomisation** or randomisation within blocks
- Example nuisance variables for blocking:
  - Time or day of experiment
  - Litter, cage, etc.
  - Person carrying out experiment
  - Sex, age, body weight, etc.
  - Another related measure (e.g. starting cell numbers, level of cytokine, or similar)
- Use random number generator, flip a coin, roll a dice

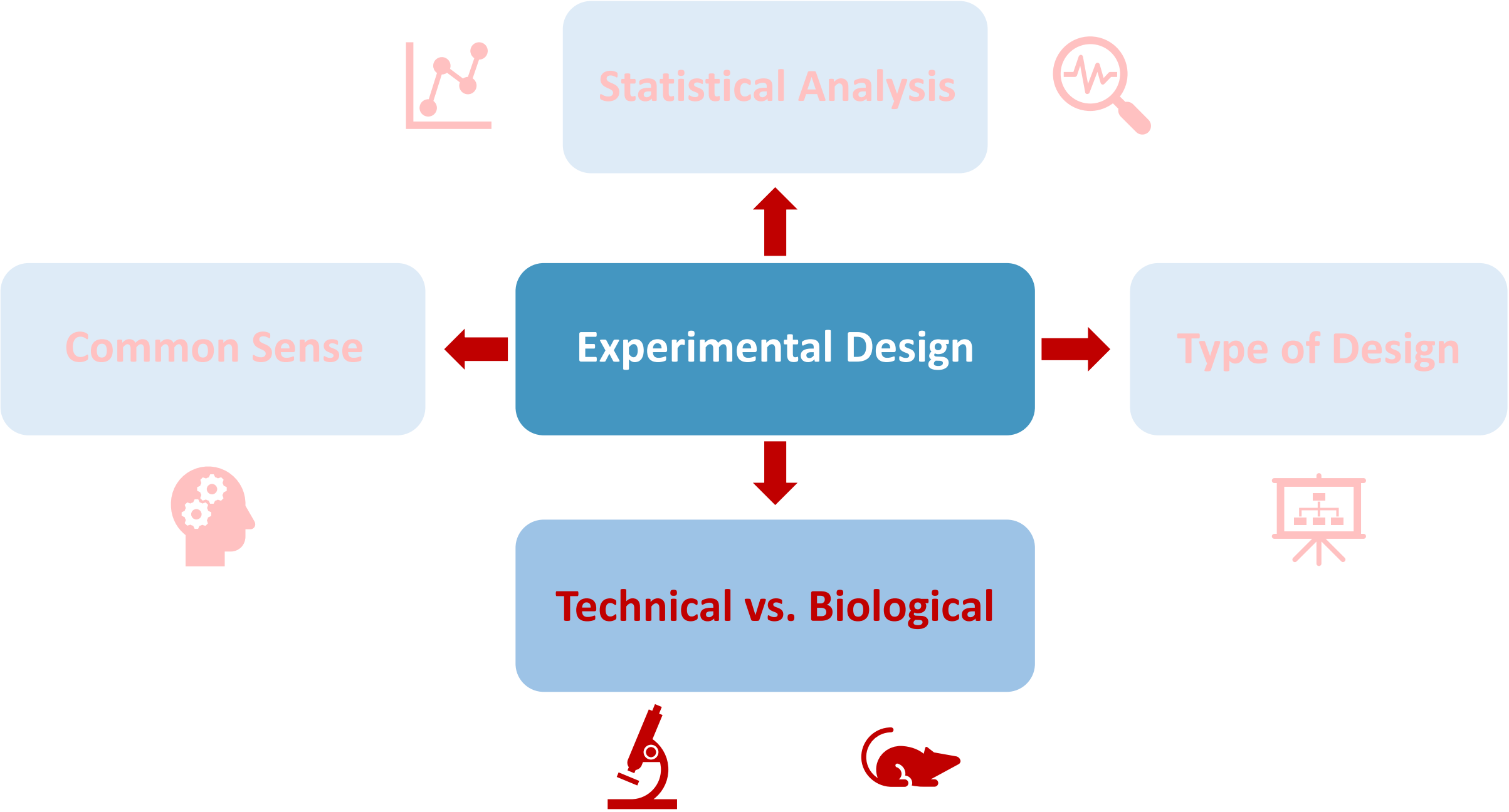
**Blinding** should be thought about:

- When allocating groups
- When doing the experiment
- When measuring outcomes
- When doing the analysis

**The ARRIVE guidelines 2.0**

1. Study design
2. Sample size
3. Inclusion and exclusion criteria
4. **Randomisation**
5. **Blinding/Masking**
6. Outcome measures
7. Statistical methods
8. Experimental animals
9. Results

https://arriveguidelines.org/arrive-guidelines

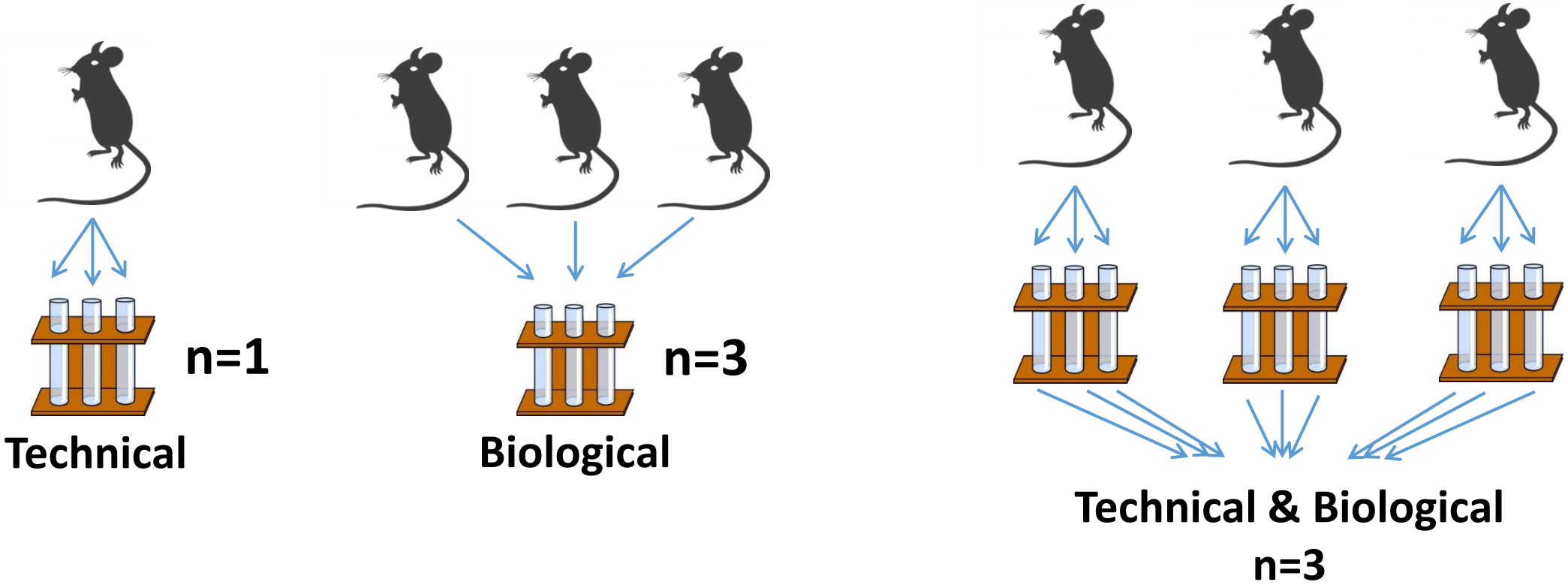

Statistical Analysis
Common Sense
Experimental Design
Type of Design
Technical vs. Biological

# Technical/biological replicates

## Not always easy

# Technical versus biological replicates

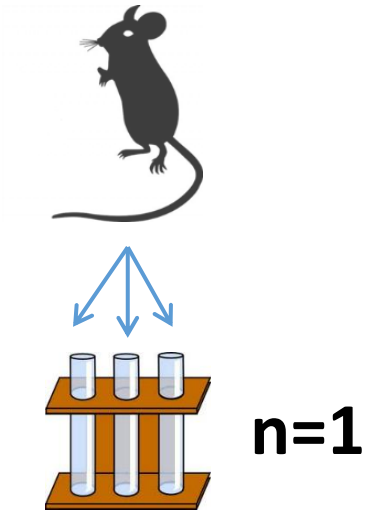
- **Technical**: repeated measures of the same sample → variability in the protocol
- **Biological**: measures of biologically distinct samples → biological variation
- **Average** of technical replicates = 1 biological replicate → ↓measurement error



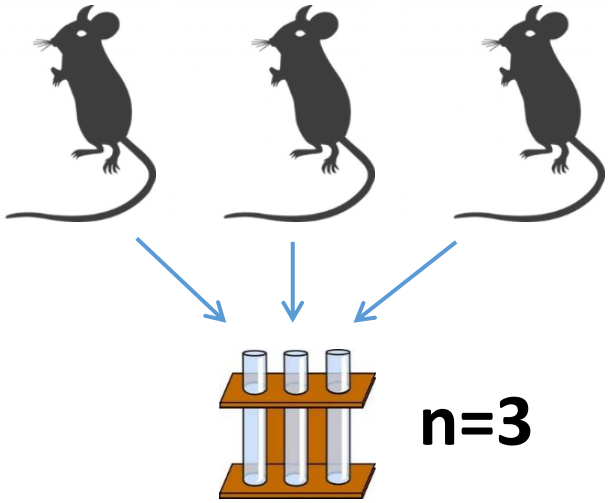
# Technical versus biological replicates

## Not always easy to tell the difference

- Definition of technical and biological depends on the model
- Mouse, human, plant, or other complex organism
  - One value per individual organism = biological replicate




**Technical**

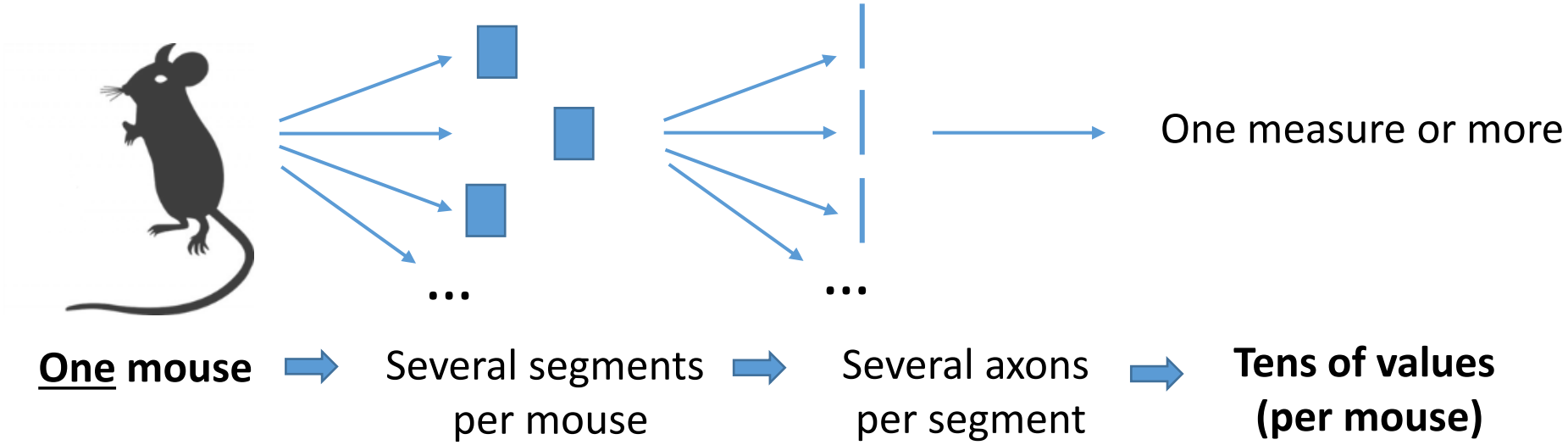



**Biological**

# Technical versus biological replicates

## Not always easy to tell the difference

- The model: mouse or other complex organism
  - >1 value per individual, e.g. axon degeneration

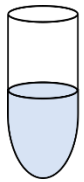



- **What to do**? **Not one good answer**
  - In this case: mouse = experiment unit, nerve segments = biological replicates, axons = technical replicates
  - But how generalisable to a wider population is this?

# Technical versus biological replicates

## Not always easy to tell the difference

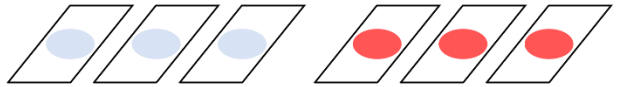
- Cells, worms, etc. = many ‘individuals’:
  - What is ‘n’ in cell culture experiments?



Vial of frozen cells

Control Treatment

Dishes, flasks, wells, …
Cells in culture
**Point of Treatment**

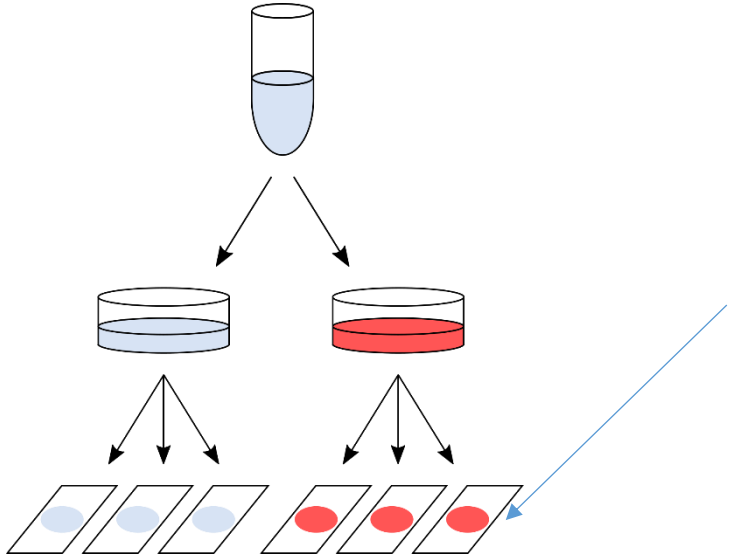


Glass slides, microarrays,
lanes in gel, wells in plate,
…
**Point of Measurements**

# Technical versus biological replicates

## Not always easy to tell the difference

- **Design 1**:

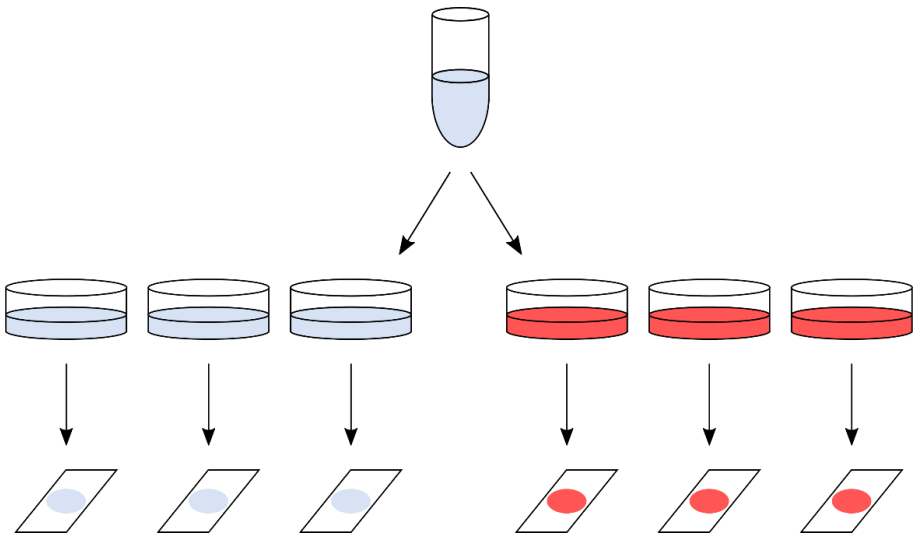



- After quantification: 6 values
  - Sample size: **n = 1**
    - no independence between slides
    - variability = pipetting/measurement error

# Technical versus biological replicates

## Not always easy to tell the difference

- **Design 2**:



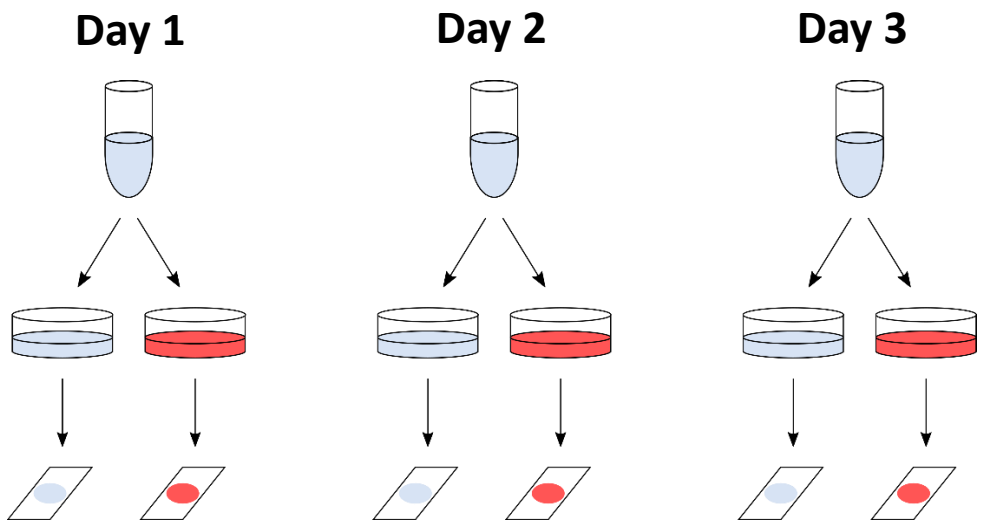
Everything processed on the same day

- After quantification: 6 values
  - Sample size: **n = 1**
    - no independence between plates
    - variability = bit better as sample split higher up in the hierarchy

# Technical versus biological replicates

## Not always easy to tell the difference

- **Design 3**: Often, as good as it can get

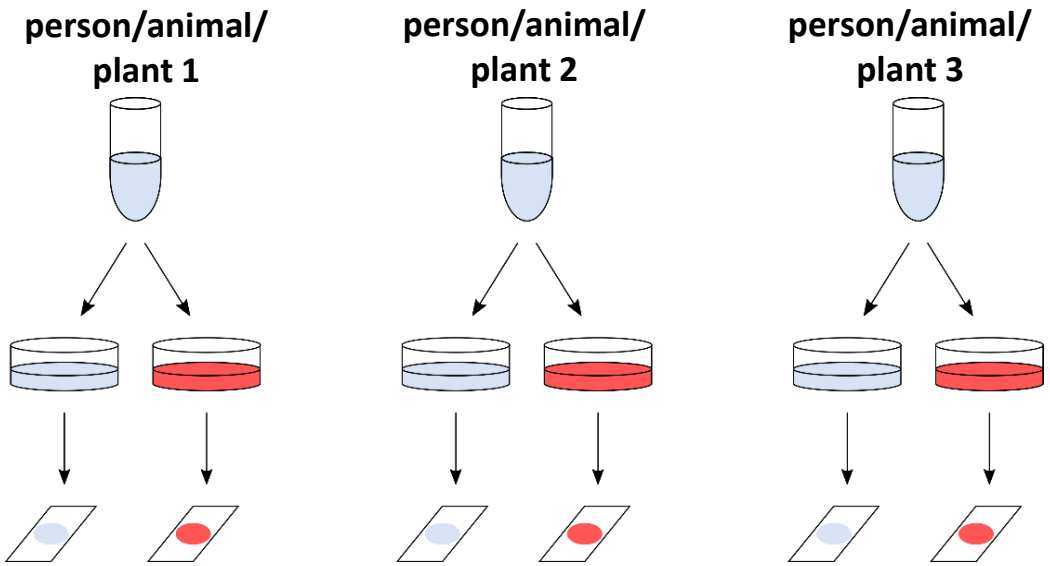



- After quantification: 6 values
  - Sample size: **n = 3**
    - **Whole procedure** repeated **3 separate times**
      - 3 days are (mostly) independent
    - Technical variability but at the highest hierarchical level
    - 2 glass slides = paired observations

# Technical versus biological replicates

## Not always easy to tell the difference

- <u>Design 4</u>: The ideal design



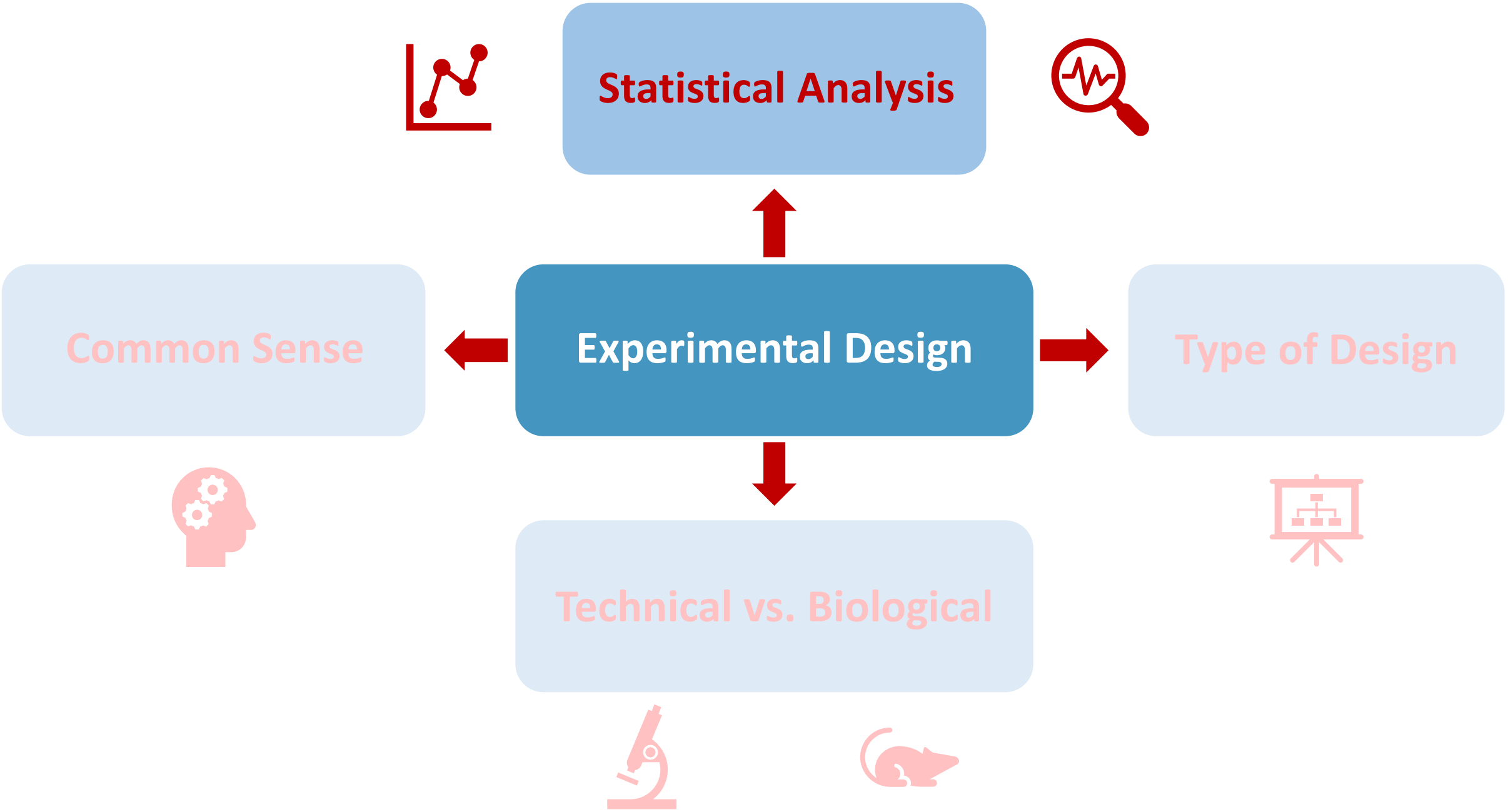

- After quantification: 6 values
  - Sample size: **n = 3**
    - Real biological replicates

# Technical versus biological replicates

- Identify technical and biological replicates
- Make the replicates as independent as possible
- Consider wider factors, e.g. rarity of samples, cost and accuracy of measurements

- Never mix technical and biological replicates
- Do not generalise your results beyond what you are able to show

- How ‘good’ your biological replicates are determines how generalisable your results are
  - ↑ confidence if true biological replicates
  - ↓ confidence if single cell line


Statistical Analysis
Common Sense
Experimental Design
Type of Design
Technical vs. Biological

# Experimental Design

## Statistical analysis

- Think about the statistical analyses **before** you collect any data
  - **Translate the hypothesis into statistical questions**

What data will I collect?

How will it be recorded/produced?

Do I know enough stats to analyse my data?

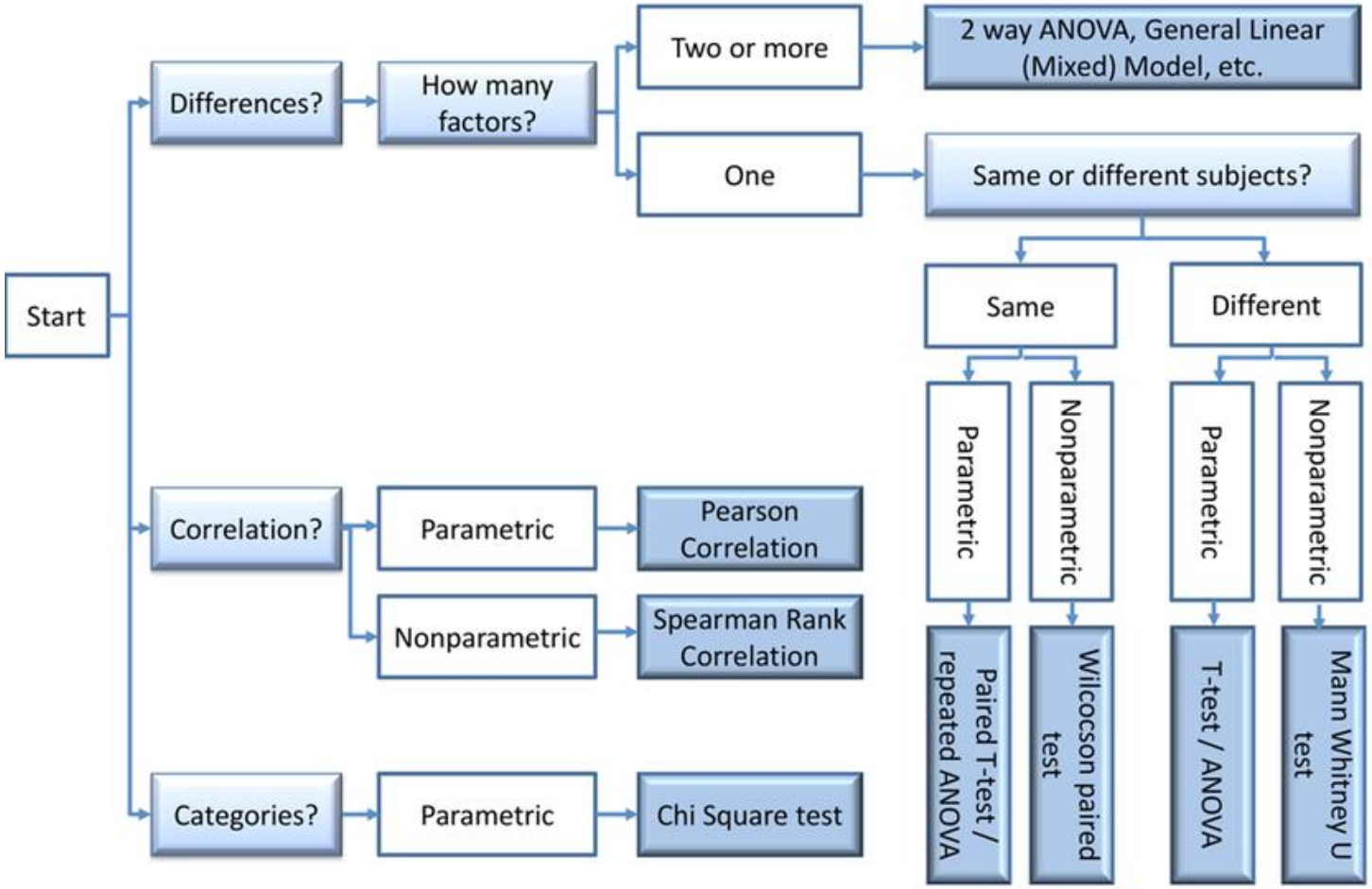
Will I have access to the raw data?

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

If not: ask for help!

# Experimental Design

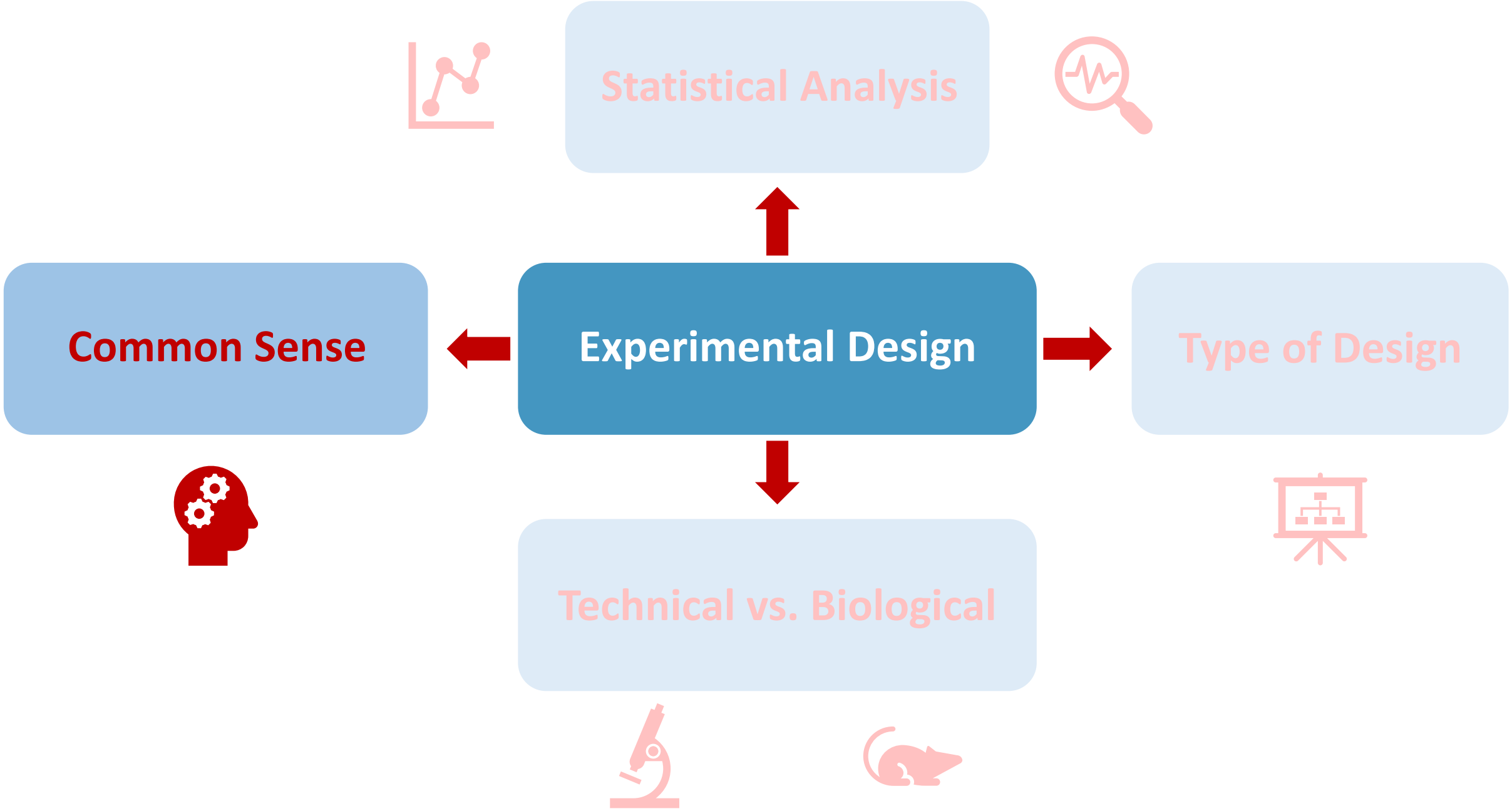
## Statistical analysis



# Experimental Design

## Exploratory data analysis (EDA)

- Purpose of EDA = discovery
- Less confidence in results so follow up with confirmatory tests
- Confirmatory approaches (hypothesis testing) provide stronger statistical evidence
- EDA ≠ “p-hacking” but could be if reported as if confirmatory
- **Be clear about approach taken** – harm comes from misrepresenting processes


Statistical Analysis
Common Sense
Experimental Design
Type of Design
Technical vs. Biological

# Experimental Design

## Common sense

- Design your experiment to be analysable
- Imagine how your results will look
- Imagine what could go wrong at each step
- Accept limitations and account for them (be prepared for follow up experiments, if required)

- The gathering of results or carrying out of a procedure is not the end goal
- Don’t get fixated on being able to perform a cool technique or experimental protocol
- Don’t overcomplicate
- Don’t get overwhelmed (ask for help)

Will these results address your hypothesis?

Babraham Institute

# Power Analysis

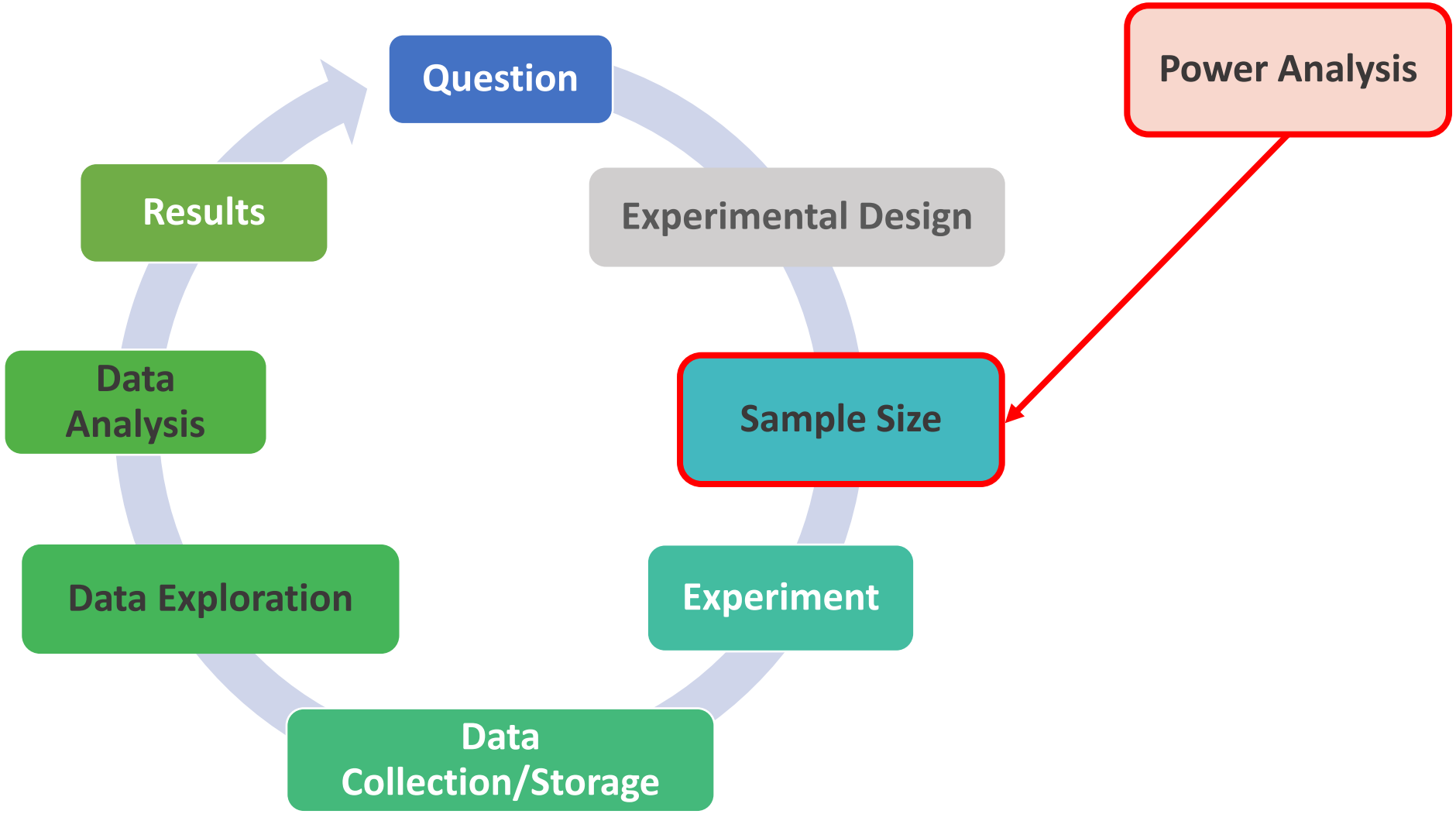
Hayley Carr & Anne Segonds-Pichon

v2024-05

Babraham Bioinformatics

# Power analysis

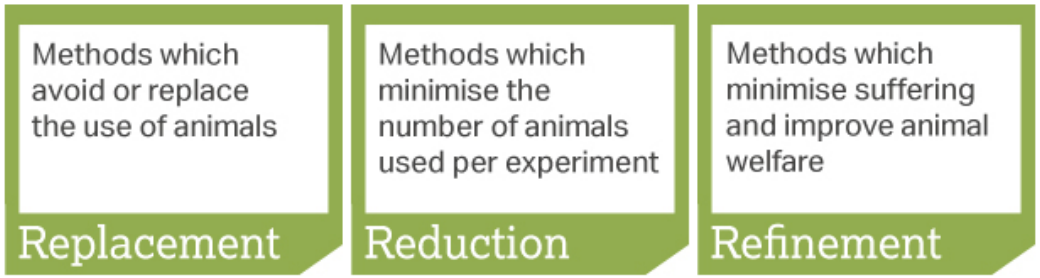
- **Power analysis** is about estimating the **appropriate sample size**.



# Sample size

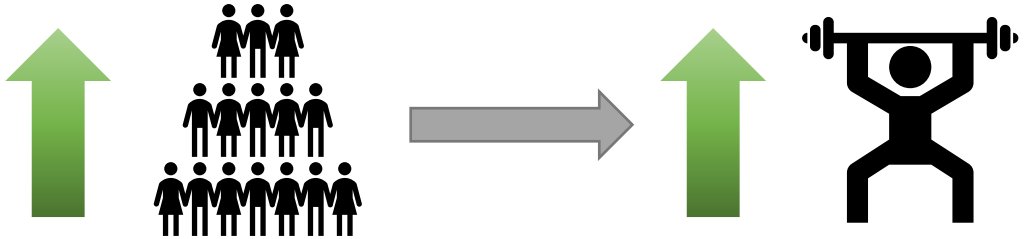
- **Too big**: waste of resources
- **Too small**: may miss the effect (p>0.05) + waste of resources
- **Grants**: justification of sample size
- **Publications:** reviewers ask for power calculation evidence
- **Home office (UK)**: the 3 Rs: Replacement, **Reduction** and Refinement

- To estimate an appropriate **sample size**, we need to do a **power analysis**



# Statistical power

- In a nutshell: the bigger the experiment (bigger sample size), the bigger the power (more likely to pick up a difference)
- **Power** = probability of **detecting an effect**, given that the effect is really there
  - = the probability that a statistical test will reject a false null hypothesis (H0)
    - To really understand power, we first need to understand some statistical concepts…

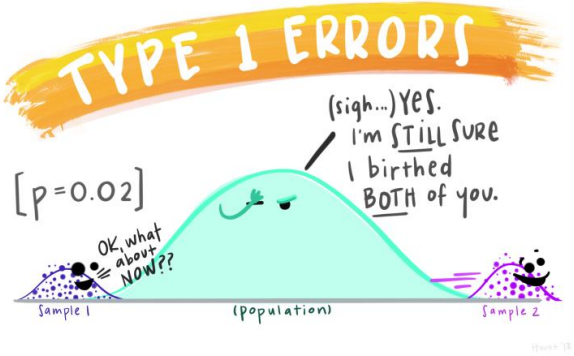


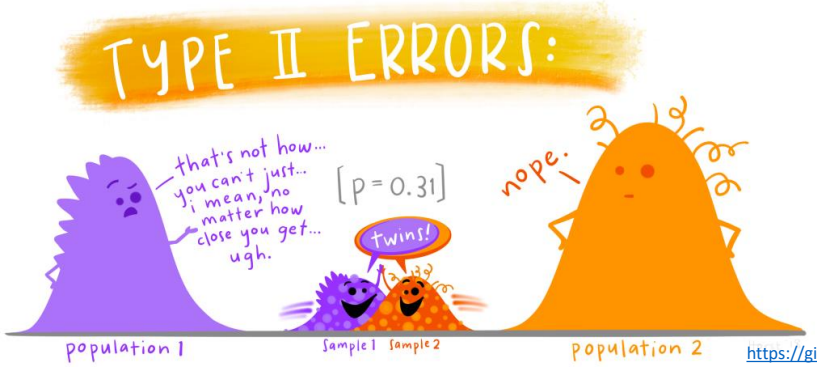
# Hypothesis testing

- The null hypothesis: $H_0$ = no effect
- The aim of a statistical test is to reject or not $H_0$.

| Statistical decision | True state of $H_0$ | |
|---|---|---|
| | **$H_0$ True (no effect)** | **$H_0$ False (effect)** |
| **Reject $H_0$** | **Type I error α**<br>**False Positive** | Correct<br>**True Positive** |
| **Do not reject $H_0$** | Correct<br>**True Negative** | **Type II error β**<br>**False Negative** |

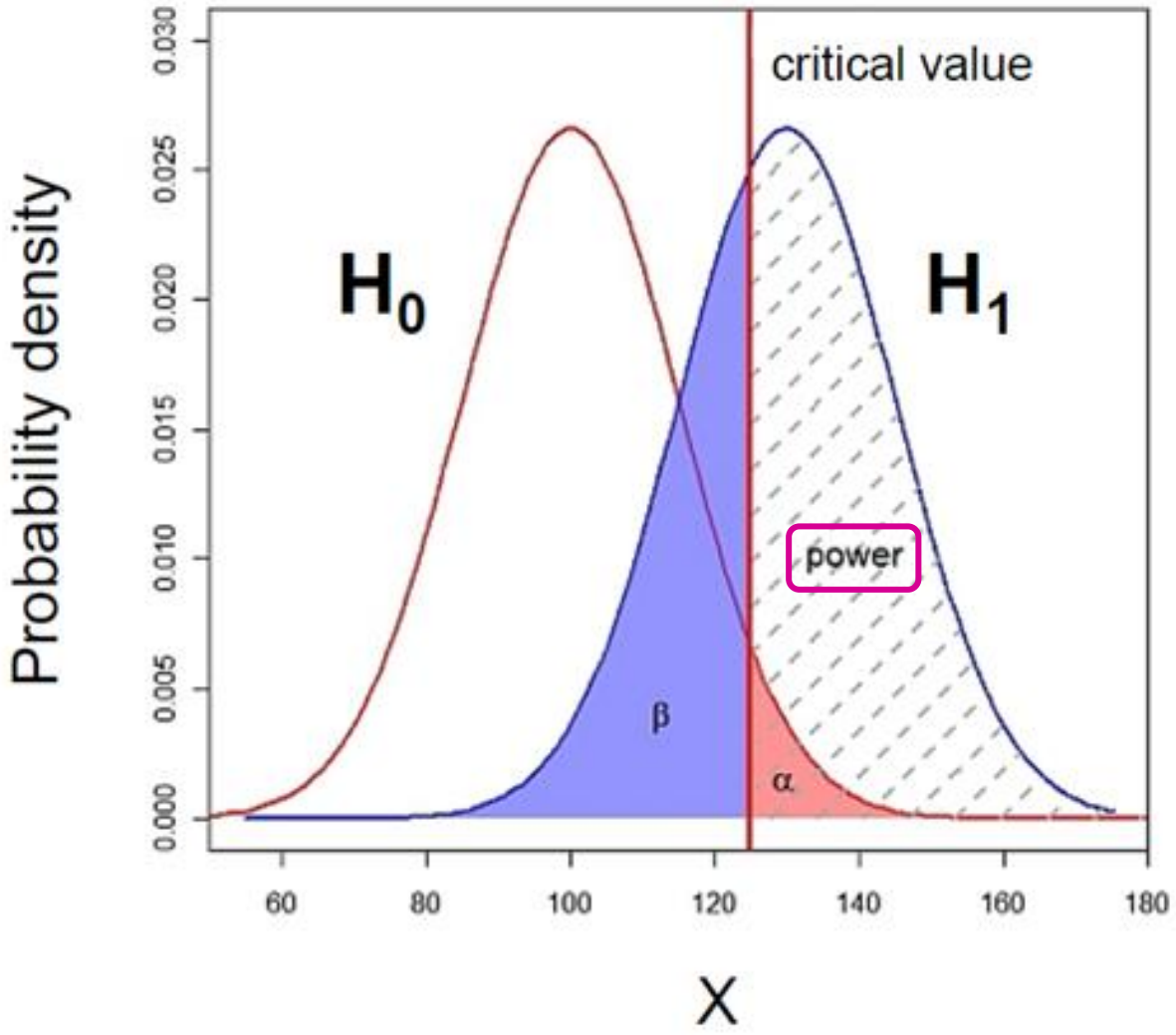
Properly **powered** studies minimise this



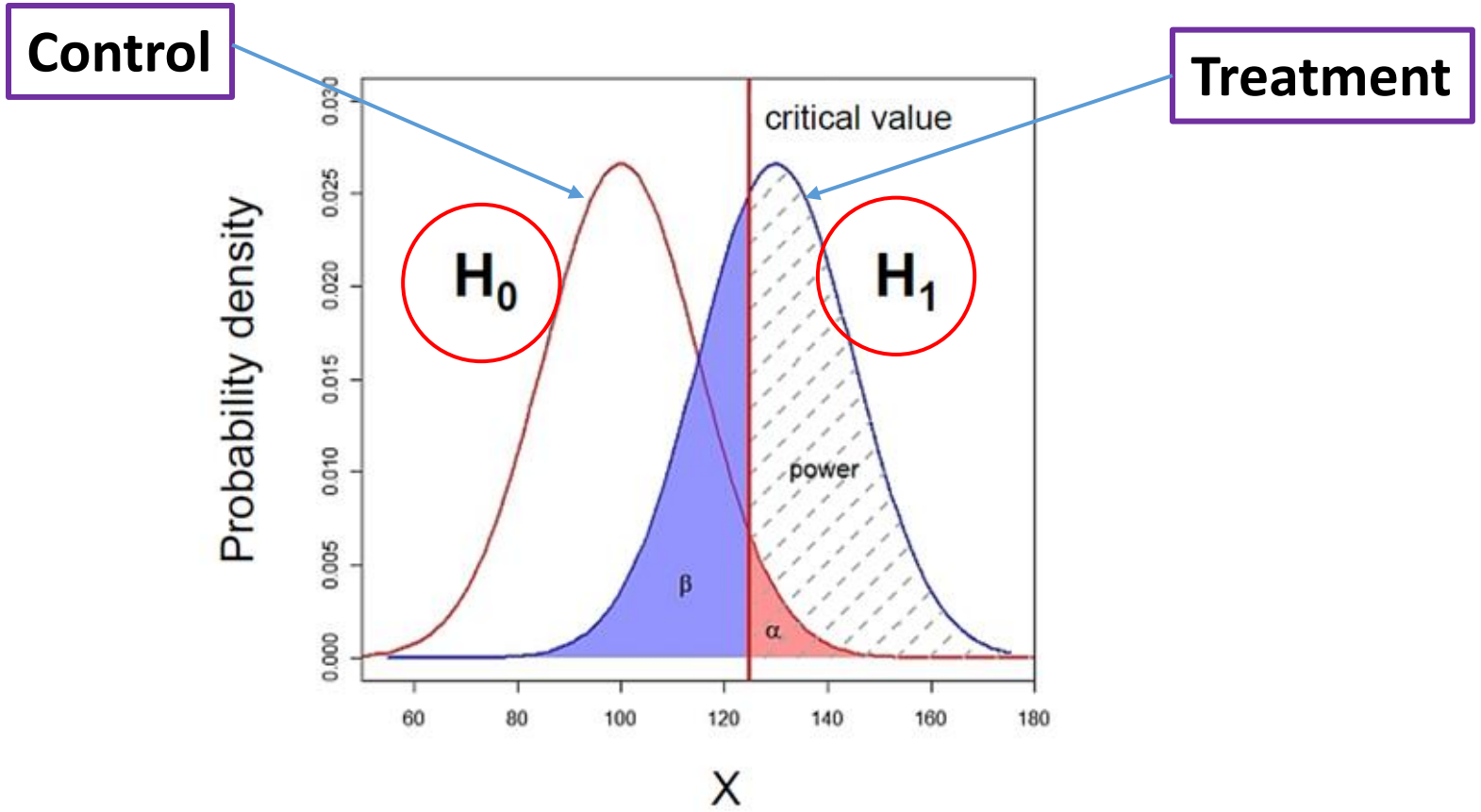




https://github.com/allisonhorst/stats-illustrations#other-stats-artwork

# What does Power look like?

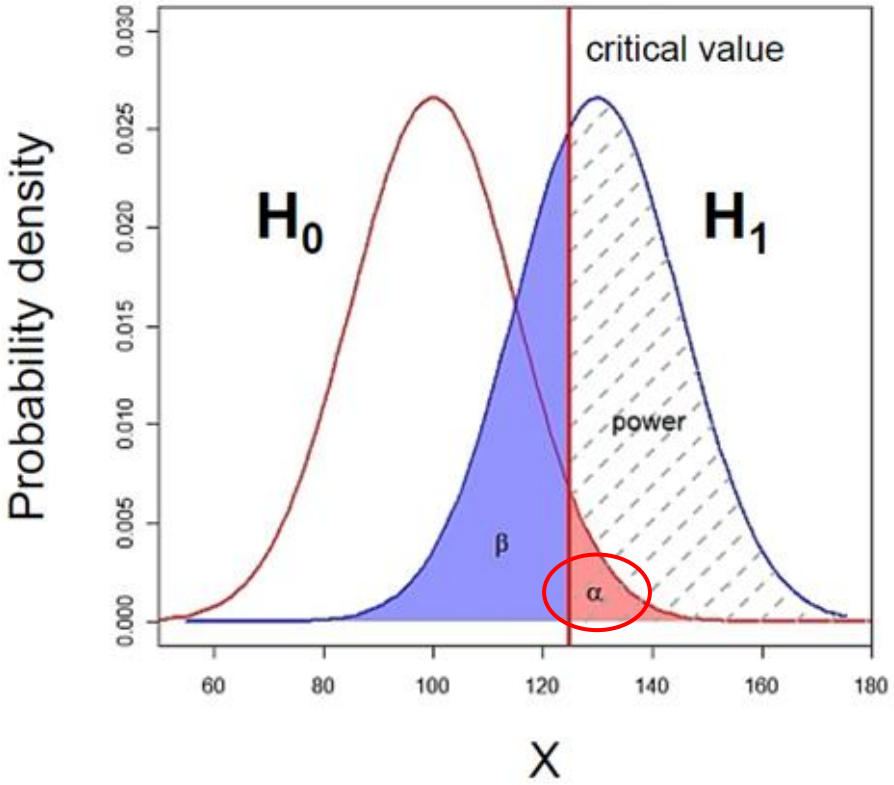


# What does Power look like? Null and alternative hypotheses



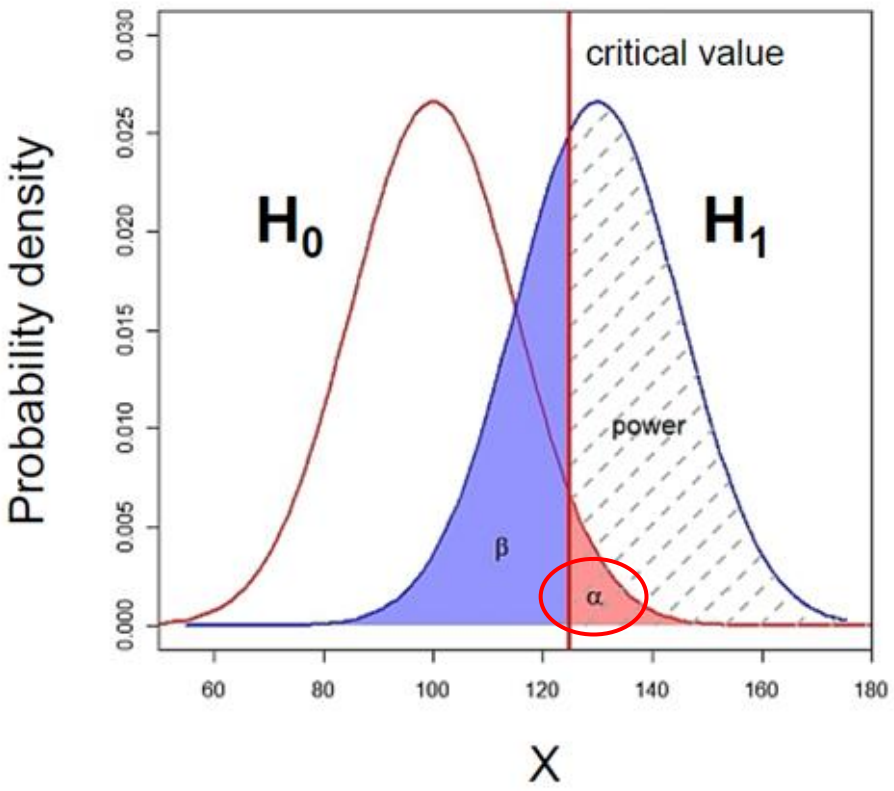

- Probability that the observed result occurs if $H_0$ is true
  - $H_0$ : **Null hypothesis** = absence of effect
  - $H_1$: **Alternative hypothesis** = presence of an effect
    - Statistics is all about rejecting the Null or not.

# What does Power look like? Type I error (α)




- **Type I error** is the failure to reject a <u>true</u> $H_0$
  - <u>Claiming</u> an effect which is not there.
  - **α** : probability of making a Type I error
- **α** : the significance level, usually set at **0.05 or 5%**

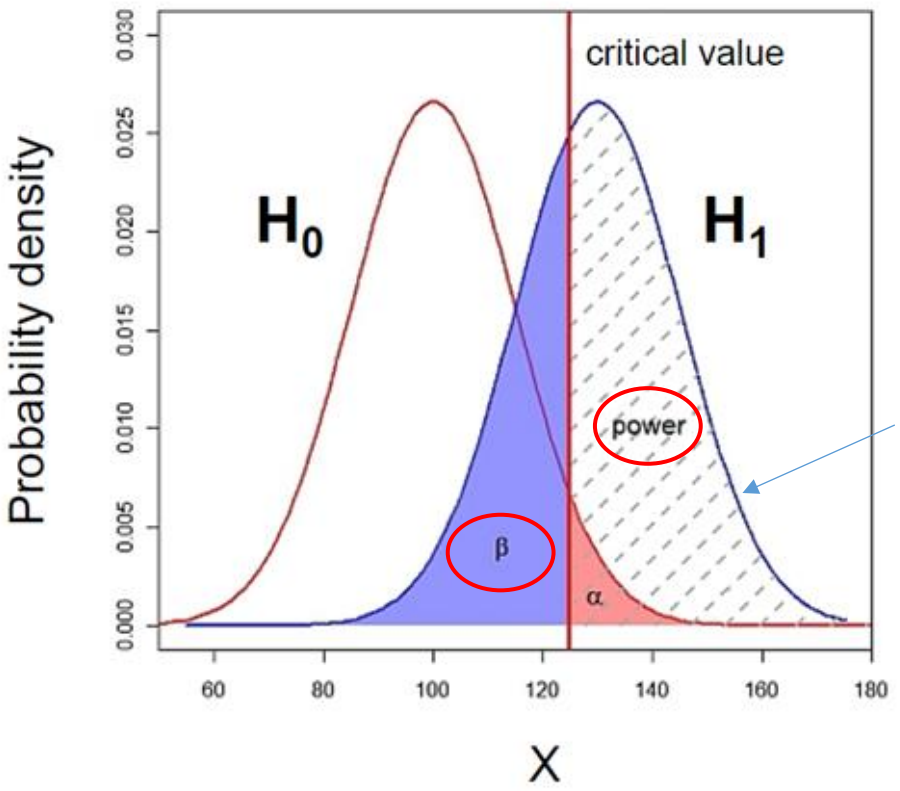
# What does Power look like? Type I error (α) and the p-value




- **p-value**: probability that the observed statistic occurred by chance alone
  - probability that a difference as big as the one observed could be found even if there is no effect.

- **Statistical significance**: comparison between **α (=0.05)** and the **p-value**
  - **p-value < 0.05**: there is a significant difference ☺ (reject $H_0$)
  - **p-value > 0.05**: there is no significant difference ☹ (fail to reject $H_0$)

# What does Power look like? Type II error (β) and Power



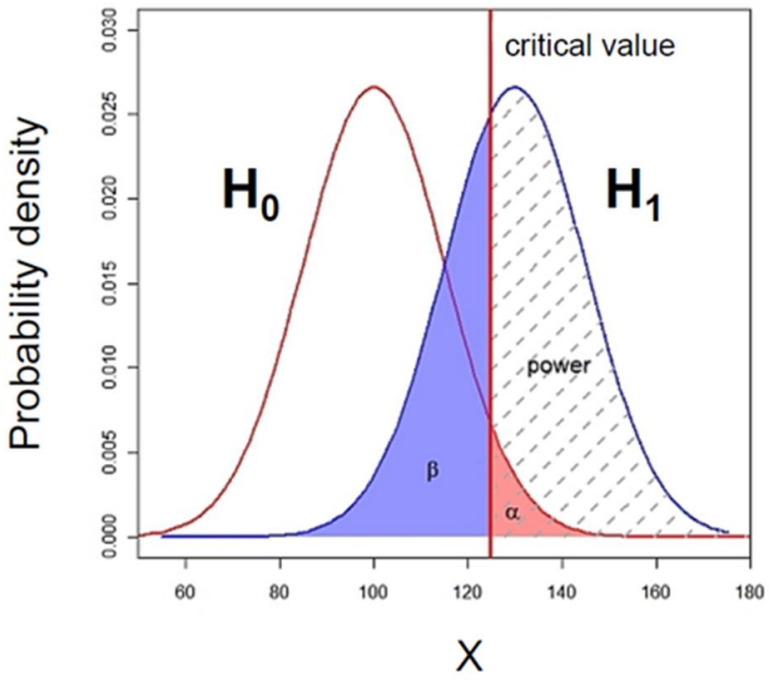

Area = 1
Power = 1 – β

- **Type II error** (**β**) is the failure to reject a false $H_0$
  - Missing an effect which is really there
  - **β** : probability of making a Type II error

- **Power**: Probability of detecting an effect which is really there.
  = Probability of rejecting a false $H_0$
  - Direct relationship between **Power** and **Type II error**: **Power** = 1 – **β**

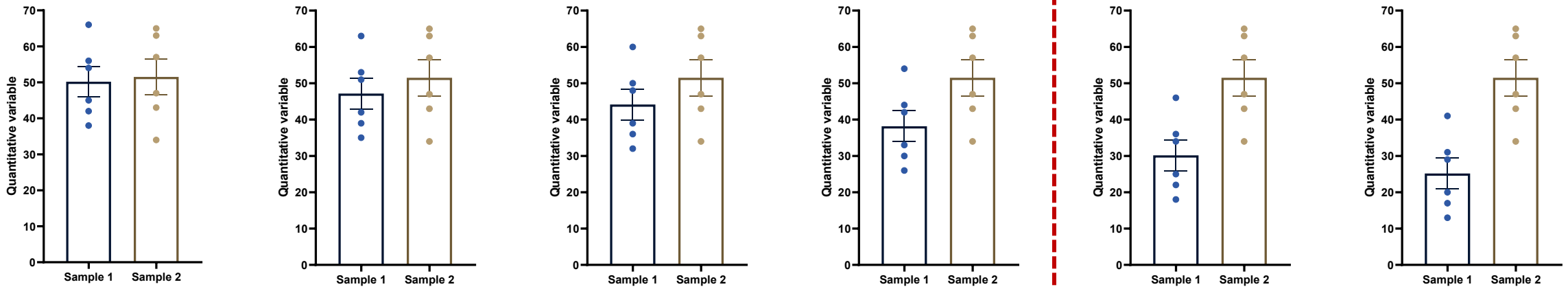
# What does Power look like? Power = 80%

- General convention: 80% but could be more
- Means a true difference will be missed 20% of the time
  - If power = 0.8 then β = 1- power = 0.2 (20%)
- Jacob Cohen (1962):
  - Type I errors are 4x more serious than Type II errors:
    - **0.05 * 4 = 0.2**
- Compromise between power and sample size, e.g. for 2 group comparisons:
  - 90% power = +30% sample size
  - 95% power = +60% sample size



# The critical value

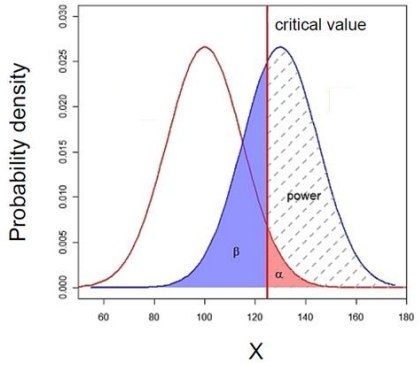
Critical value




Small difference → Big difference

Not significant: p>0.05

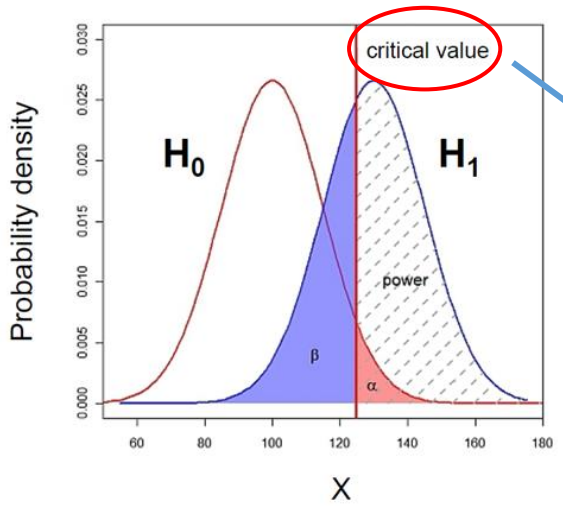
Significant: p<0.05




**Critical value = size of difference + sample size + significance level**

# The critical value: size of difference + sample size + significance

## Example with the *t*-test




Significance level

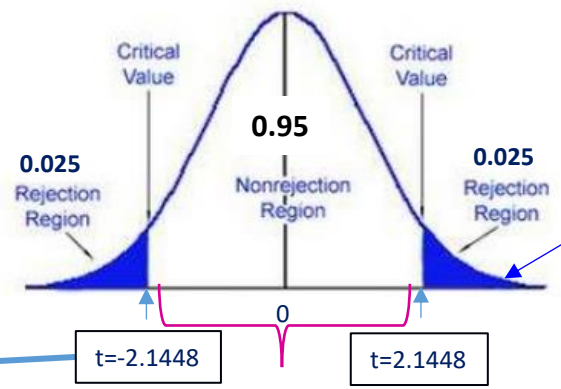
Degrees of freedom

| df | 0.20 | 0.10 | 0.05 | 0.02 | 0.01 | 0.001 |
|---|---|---|---|---|---|---|
| 1 | 3.0777 | 6.3138 | 12.7062 | 31.8205 | 63.6567 | 636.6192 |
| 2 | 1.8856 | 2.9200 | 4.3027 | 6.9646 | 9.9248 | 31.5991 |
| 3 | 1.6377 | 2.3534 | 3.1824 | 4.5407 | 5.8409 | 12.9240 |
| 4 | 1.5332 | 2.1318 | 2.7764 | 3.7469 | 4.6041 | 8.6103 |
| 5 | 1.4759 | 2.0150 | 2.5706 | 3.3649 | 4.0321 | 6.8688 |
| 6 | 1.4398 | 1.9432 | 2.4469 | 3.1427 | 3.7074 | 5.9588 |
| 7 | 1.4149 | 1.8946 | 2.3646 | 2.9980 | 3.4995 | 5.4079 |
| 8 | 1.3968 | 1.8595 | 2.3060 | 2.8965 | 3.3554 | 5.0413 |
| 9 | 1.3830 | 1.8331 | 2.2622 | 2.8214 | 3.2498 | 4.7809 |
| 10 | 1.3722 | 1.8125 | 2.2281 | 2.7638 | 3.1693 | 4.5869 |
| 11 | 1.3634 | 1.7959 | 2.2010 | 2.7181 | 3.1058 | 4.4370 |
| 12 | 1.3562 | 1.7823 | 2.1788 | 2.6810 | 3.0545 | 4.3178 |
| 13 | 1.3502 | 1.7709 | 2.1604 | 2.6503 | 3.0123 | 4.2208 |
| 14 | 1.3450 | 1.7613 | 2.1448 | 2.6245 | 2.9768 | 4.1405 |
| 15 | 1.3406 | 1.7531 | 2.1314 | 2.6025 | 2.9467 | 4.0728 |

Example: 2-tailed t-test with n=15 (df=14)

~ Difference

T Distribution




Diff "big enough"

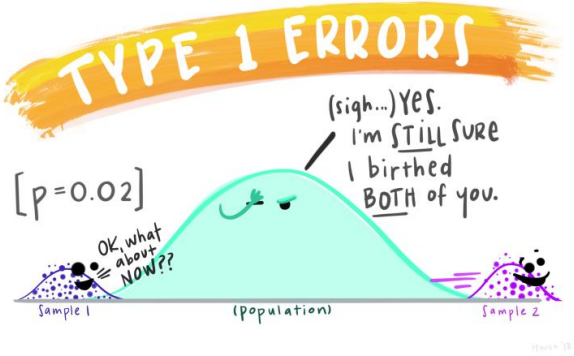
Difference "too small"

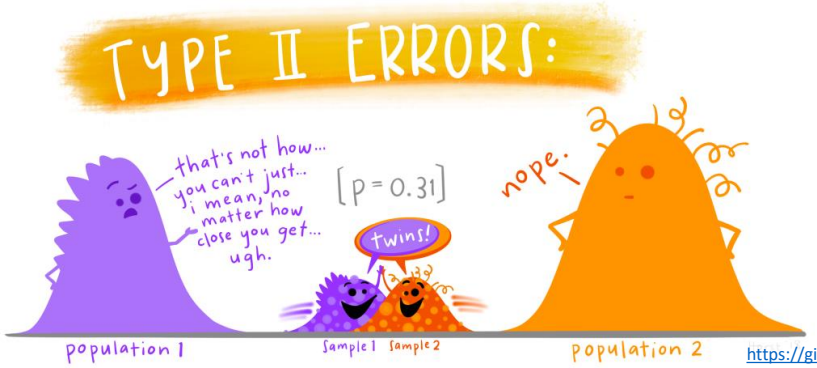
- In **hypothesis testing**:
  - **critical value** is compared to the **test statistic** to determine significance
  - Example of test statistic: t-value
- If **test statistic** > **critical value**: statistical significance and rejection of the null **hypothesis**
  - Example: t-value > critical t-value

# To recap:

- The null hypothesis: $H_0$ = no effect
- The aim of a statistical test is to reject or not $H_0$.

| Statistical decision | True state of $H_0$ | |
|---|---|---|
| | **$H_0$ True (no effect)** | **$H_0$ False (effect)** |
| **Reject $H_0$** | **Type I error α**<br>**False Positive** | Correct<br>**True Positive** |
| **Do not reject $H_0$** | Correct<br>**True Negative** | **Type II error β**<br>**False Negative** |




https://github.com/allisonhorst/stats-illustrations#other-stats-artwork

# Power analysis

**The power analysis depends on the relationship between 6 variables**:

- the significance level (5%)
- the desired power of the experiment (80%)
- the alternative hypothesis (i.e. one or two-sided test)
- the **difference** of biological interest
- the **variability** in the data (standard deviation)

} **Effect size**

- the **sample size**

# Power analysis

**The power analysis depends on the relationship between 6 variables**:

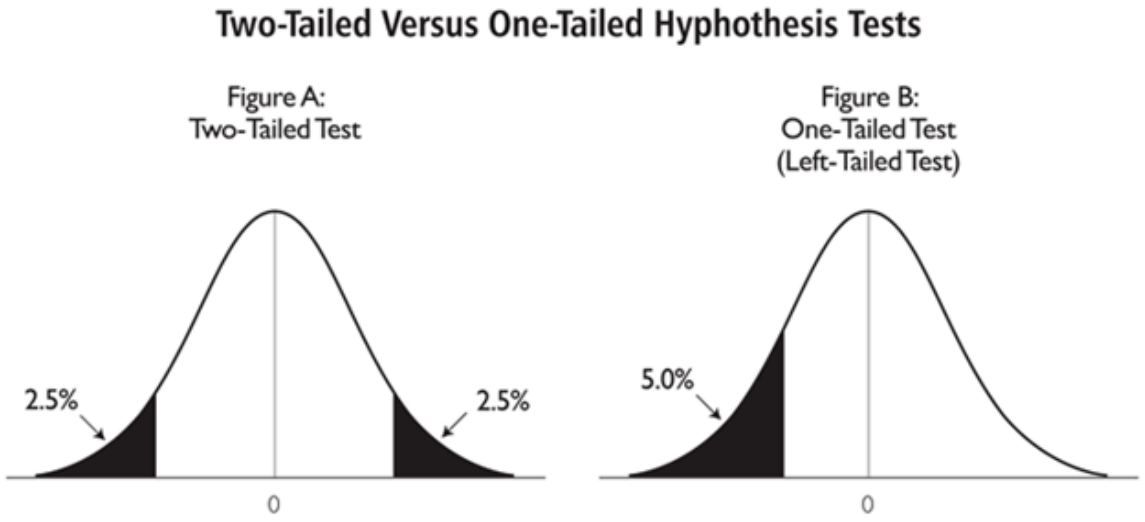
- the significance level (5%)
- the desired power of the experiment (80%)
- the alternative hypothesis (i.e. one or two-sided test)
- the **difference** of biological interest
- the **variability** in the data (standard deviation)

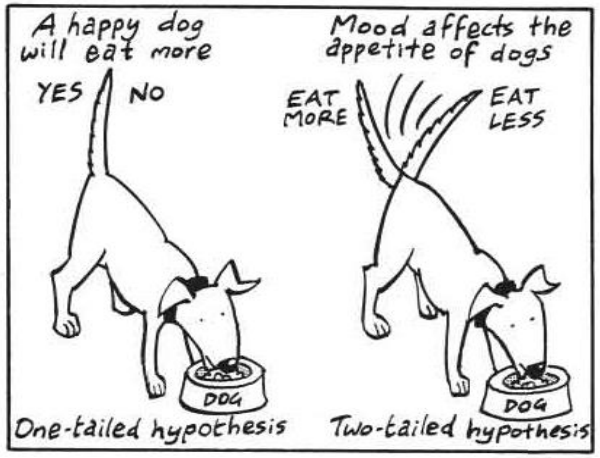
} **Effect size**

- the **sample size**

# The alternative hypothesis: what is it?

- One-tailed or two-tailed test? One-sided or two-sided tests?




- Is the question:
  - Is the there a difference? → Two-tailed
  - Is it larger than or smaller than? → One-tailed

- Can rarely justify the use of a one-tailed test
- Two times easier to reach significance with one-tailed than two-tailed → suspicious reviewer!

# Power analysis

**The power analysis depends on the relationship between 6 variables**:

- the significance level (5%)
- the desired power of the experiment (80%)
- the alternative hypothesis (i.e. one or two-sided test)
- the **difference** of biological interest
- the **variability** in the data (standard deviation)

} **Effect size**

- the **sample size**

# The difference of biological interest

- Determined scientifically (not statistically)
  - Minimum meaningful effect of biological relevance (Minimum Effect of Interest, MEI)
- **How to determine it?**
  - Previous research, pilot study

# The variability

- We need to have an idea of the standard deviation before we start the experiment
- **How to determine it?**
  - Data from previous research on WT, control or baseline

## Effect size

Combination of absolute effect and variability

# The effect size: how is it calculated?

- Depends on the type of difference and the data
  - Easy example: comparison between 2 means

$$\text{Effect Size} = \frac{[\text{Mean of experimental group}] - [\text{Mean of control group}]}{\text{Standard Deviation}}$$

Absolute difference

Variability

- Jacob Cohen defined small, medium and large effects for different tests – but **not recommended**

| Test | Relevant effect size | Effect Size Threshold | | |
|---|---|---|---|---|
| | | Small | Medium | Large |
| t-test for means | d | 0.2 | 0.5 | 0.8 |
| F-test for ANOVA | f | 0.1 | 0.25 | 0.4 |
| t-test for correlation | r | 0.1 | 0.3 | 0.5 |
| Chi-square | w | 0.1 | 0.3 | 0.5 |
| 2 proportions | h | 0.2 | 0.5 | 0.8 |

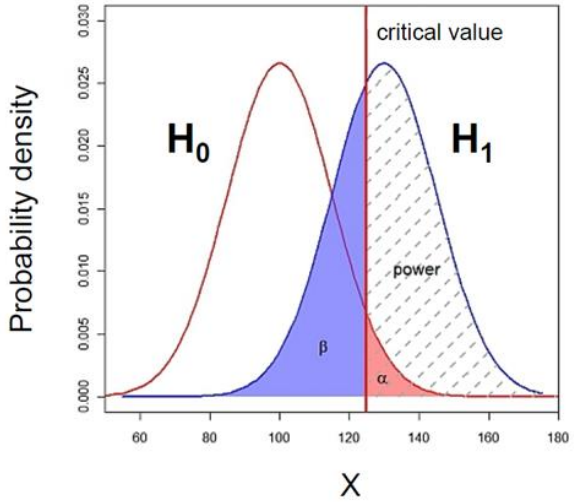
# The effect size: how is it calculated?

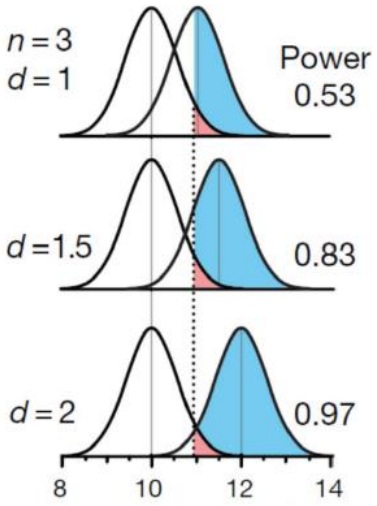
## The absolute difference

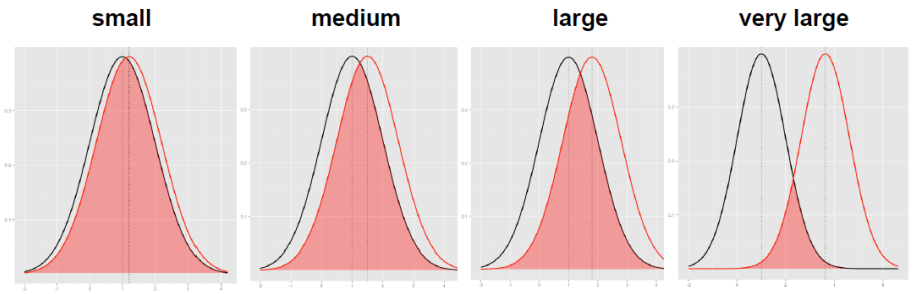
- The bigger the effect (the absolute difference), the bigger the power
  = the bigger the probability of picking up the difference

Absolute difference

$$\text{Effect Size} = \frac{[\text{Mean of experimental group}] - [\text{Mean of control group}]}{\text{Standard Deviation}}$$




http://rpsychologist.com/d3/cohend/

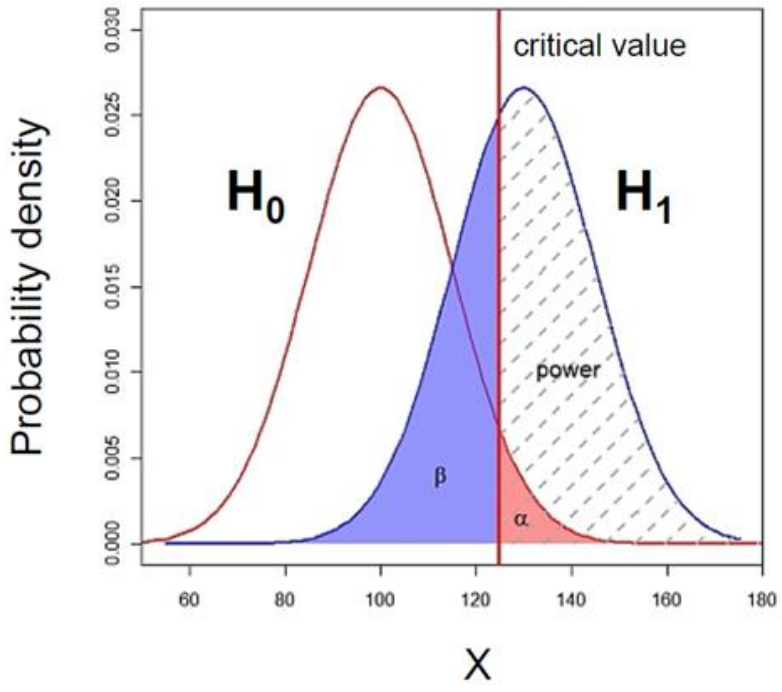
# The effect size: how is it calculated?

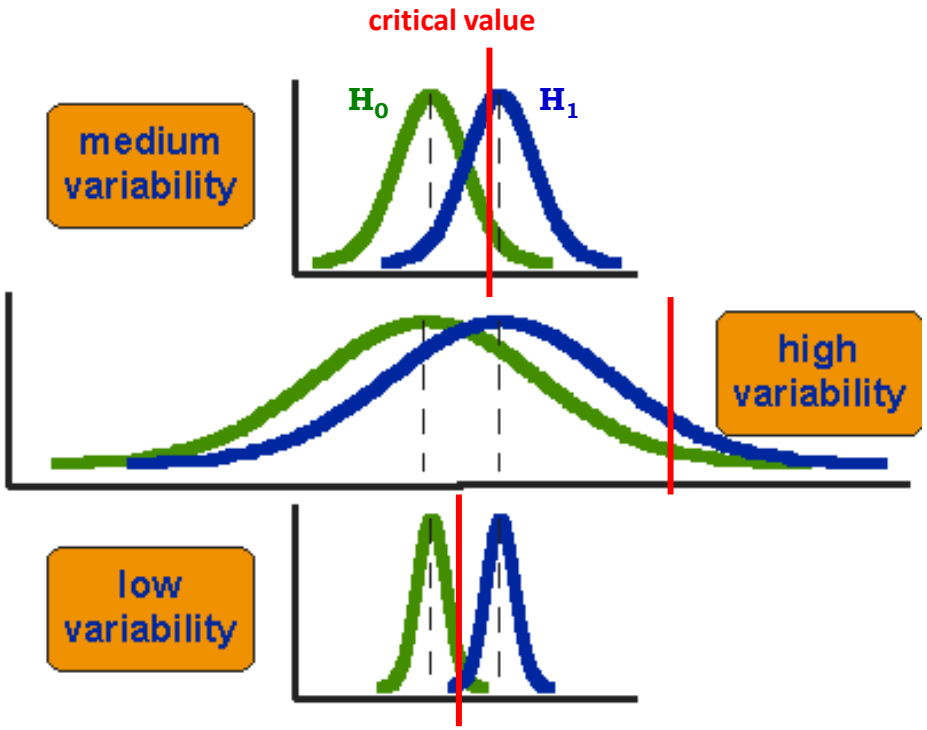
## The standard deviation

- The bigger the variability of the data, the smaller the power

$$\text{Effect Size} = \frac{[\text{Mean of experimental group}] - [\text{Mean of control group}]}{\text{Standard Deviation}}$$

Variability



# Power analysis

**The power analysis depends on the relationship between 6 variables**:

- the significance level (5%)
- the desired power of the experiment (80%)
- the alternative hypothesis (i.e. one or two-sided test)
- the **difference** of biological interest
- the **variability** in the data (standard deviation)

} **Effect size**

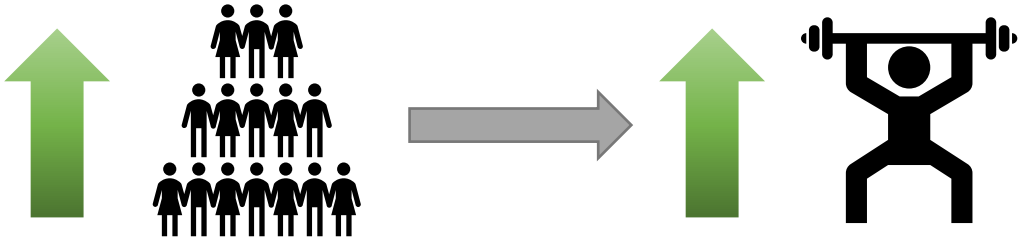
- the **sample size**

# The sample size

- Most of the time, the output of a power calculation

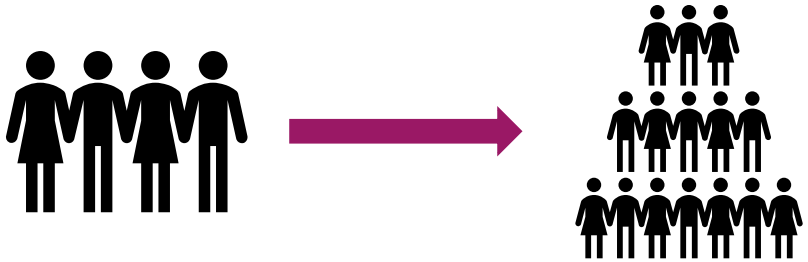
- In reality it is difficult to reduce the variability in data, or the contrast between means
  - most effective way of improving power:
    - **increase the sample size**

- **The bigger the sample, the bigger the power**
  - but how does it actually work?

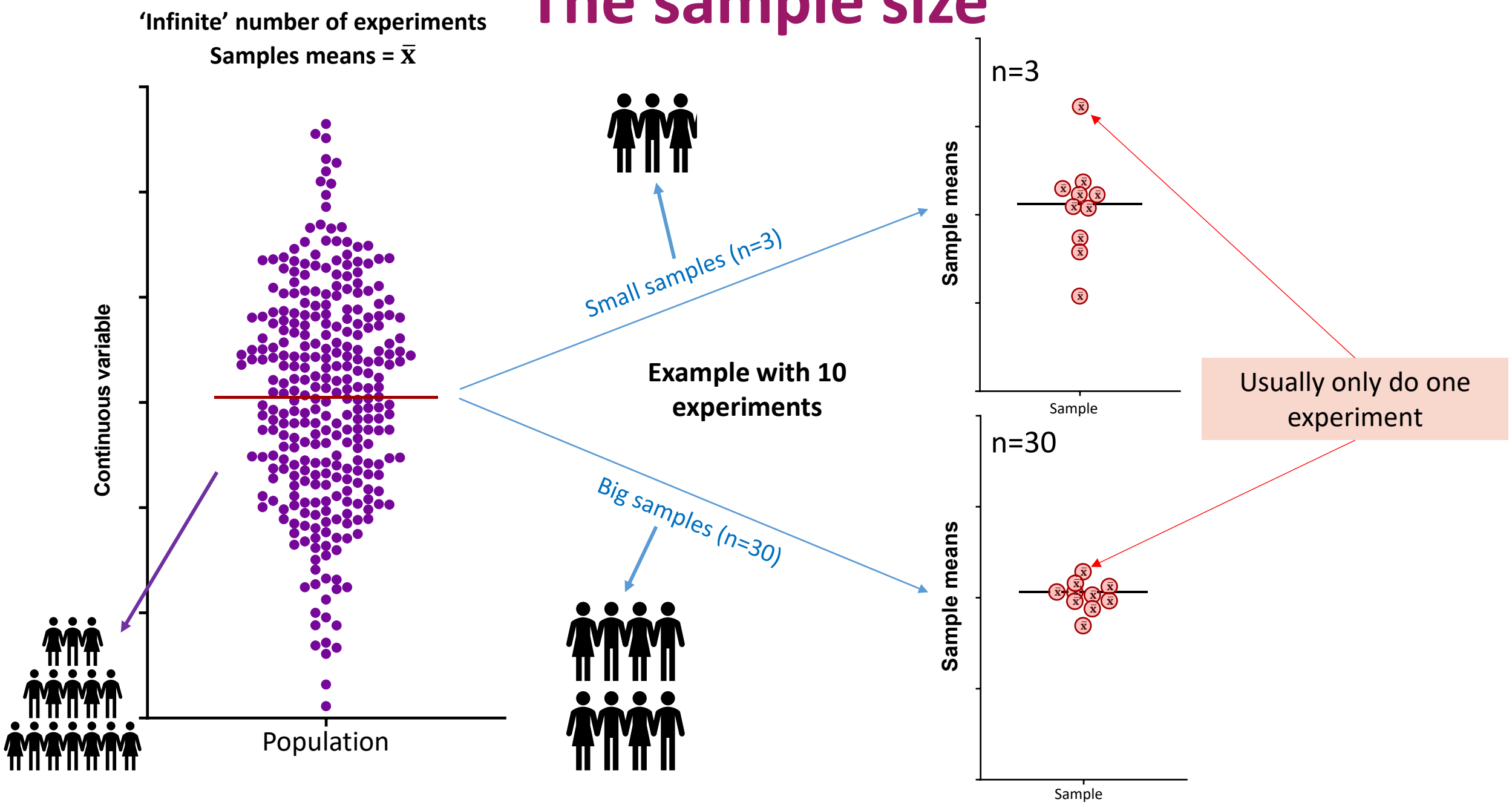


# Samples and population

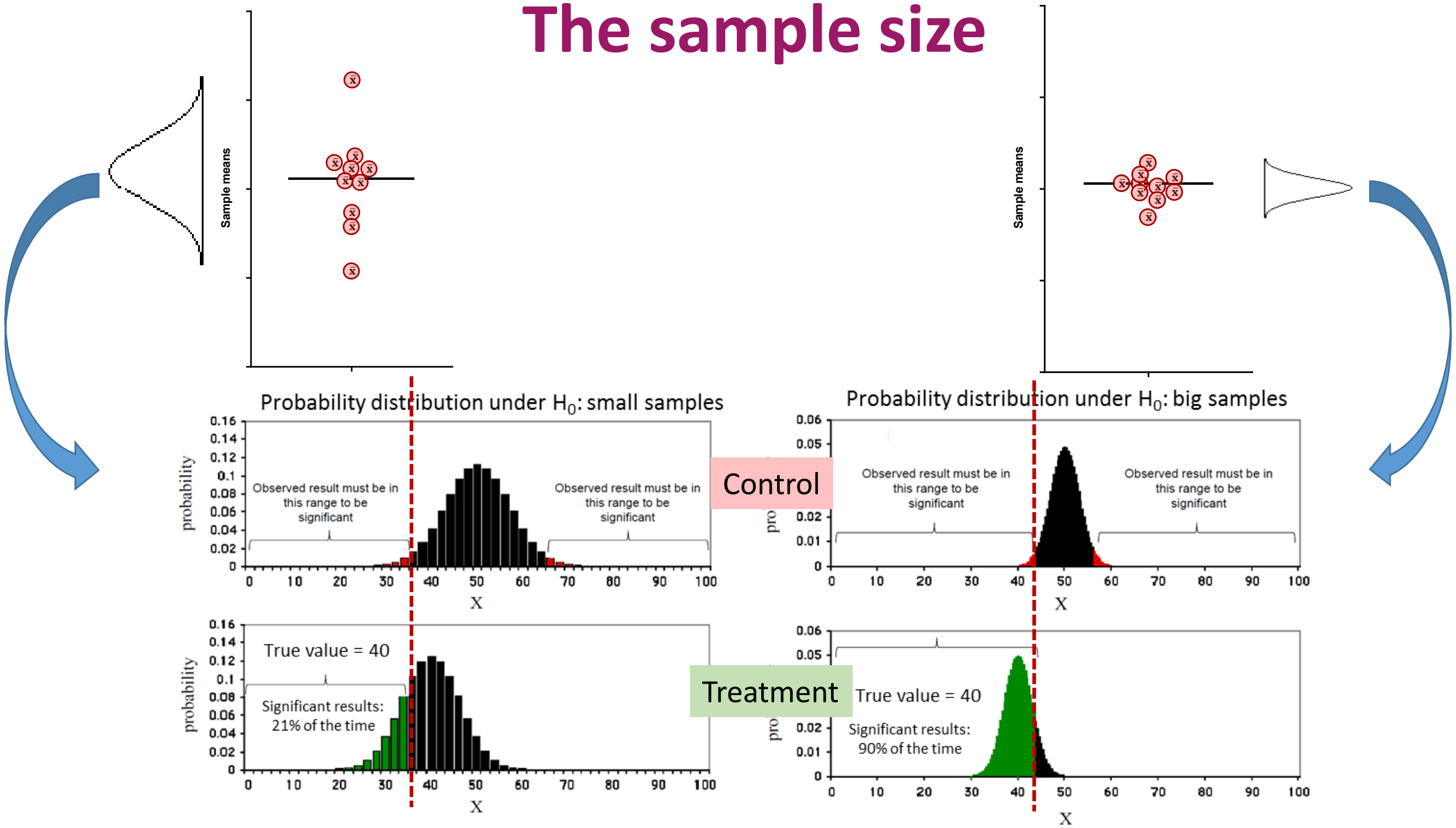
- We want to know about **whole population**
  - *All* people, *all* patients, *all* mice, *all* cells…
- Not possible to measure whole population
- So take a **representative sample**
- Make **inferences** about the population
- Larger samples more likely to be representative of the population



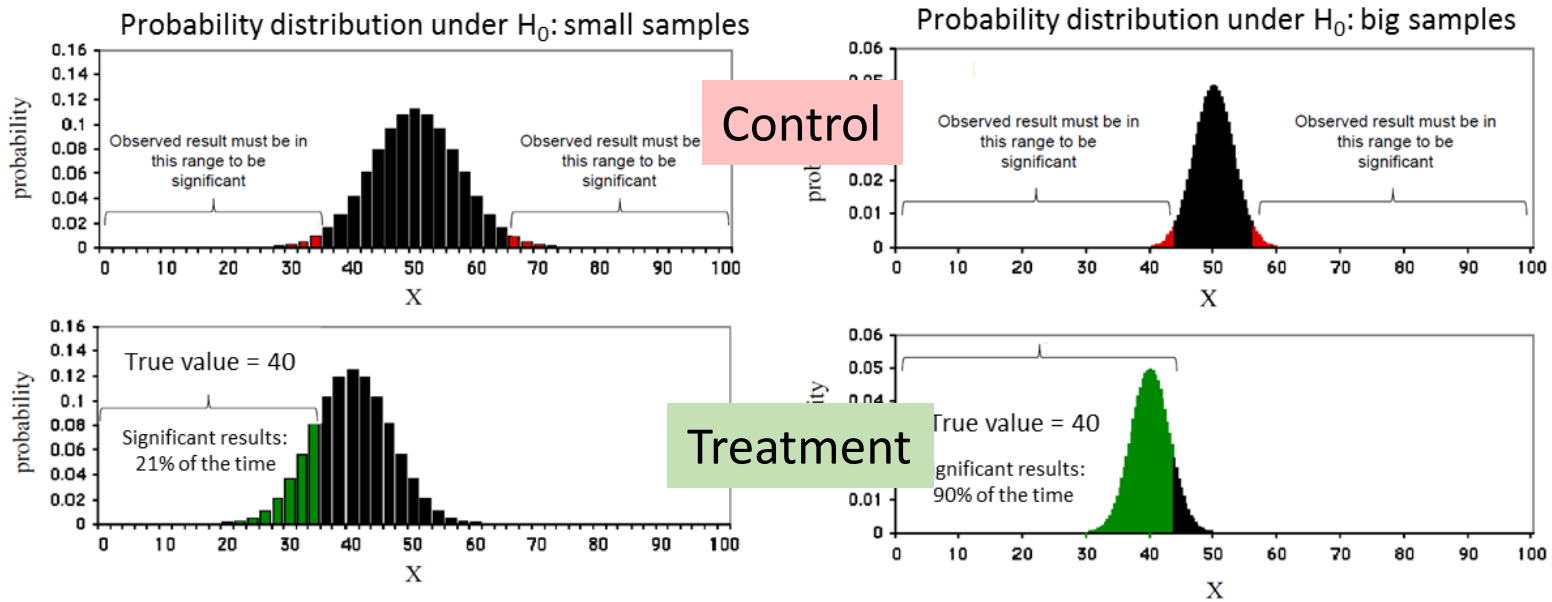
# The sample size

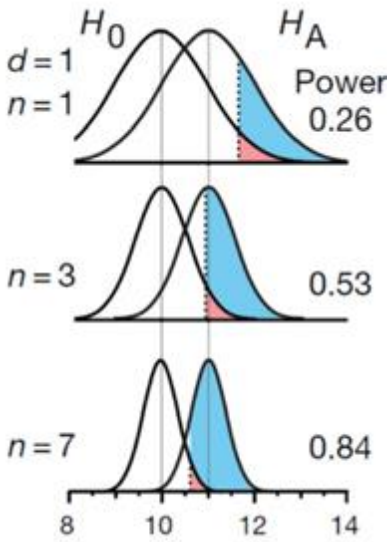


# The sample size



# The sample size



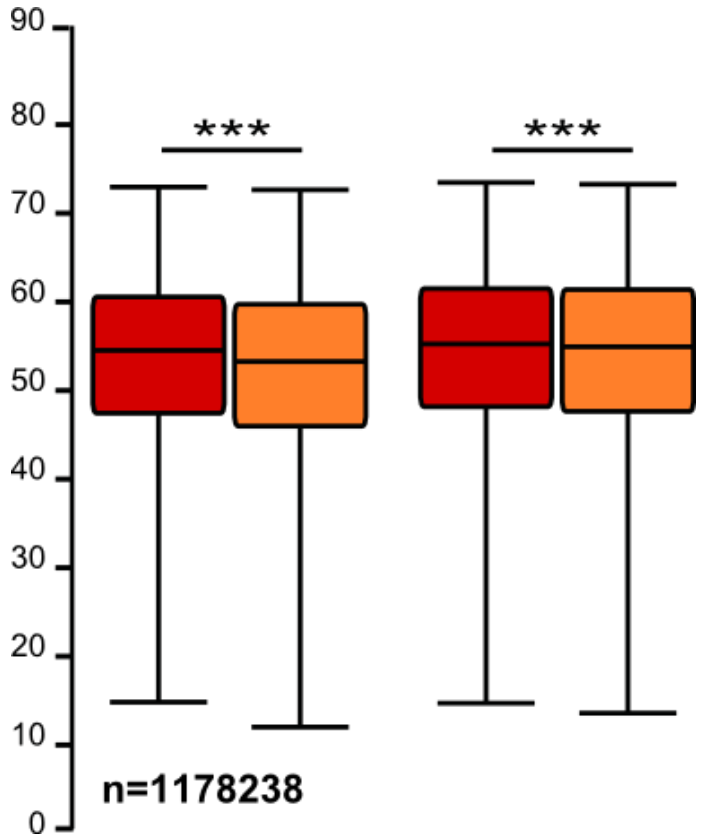



# The sample size: the bigger the better?

- It takes huge samples to detect tiny differences but tiny samples to detect huge differences

- What if the tiny difference is meaningless?
  - Beware of **overpower**
  - Nothing wrong with the stats: it is all about interpretation of the results of the test

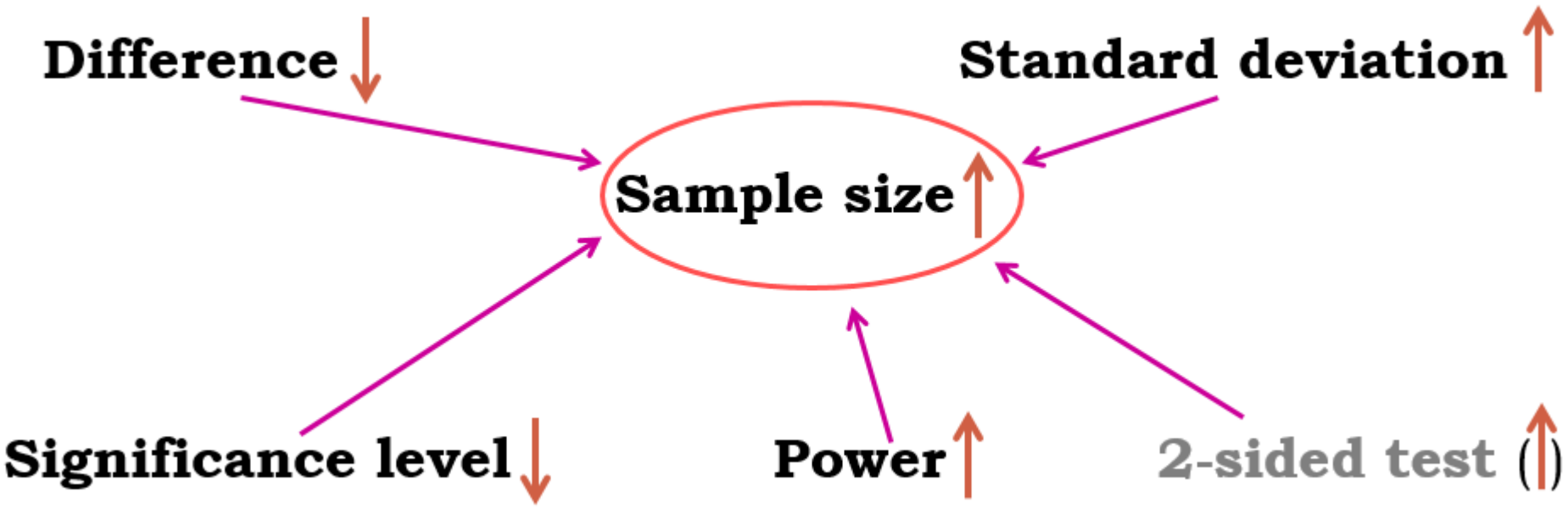
- Remember the important first step of power analysis:
  - **What is the effect size of biological interest?**



# Power analysis

## Typical question

What sample size do I need to have a 80% probability (**power**) to detect this particular effect (**difference** and **standard deviation**) at a 5% **significance level** using a **2-sided test**?



# Power analysis

- Fix any **five of the variables**, a mathematical relationship is used to estimate the **sixth**

Difference of biological interest

+ Variability in the data (standard deviation)

+ Desired power of the experiment (80%)

+ Significance level (5%)

+ Alternative hypothesis (i.e. one or two-sided test)

---

**Appropriate sample size**

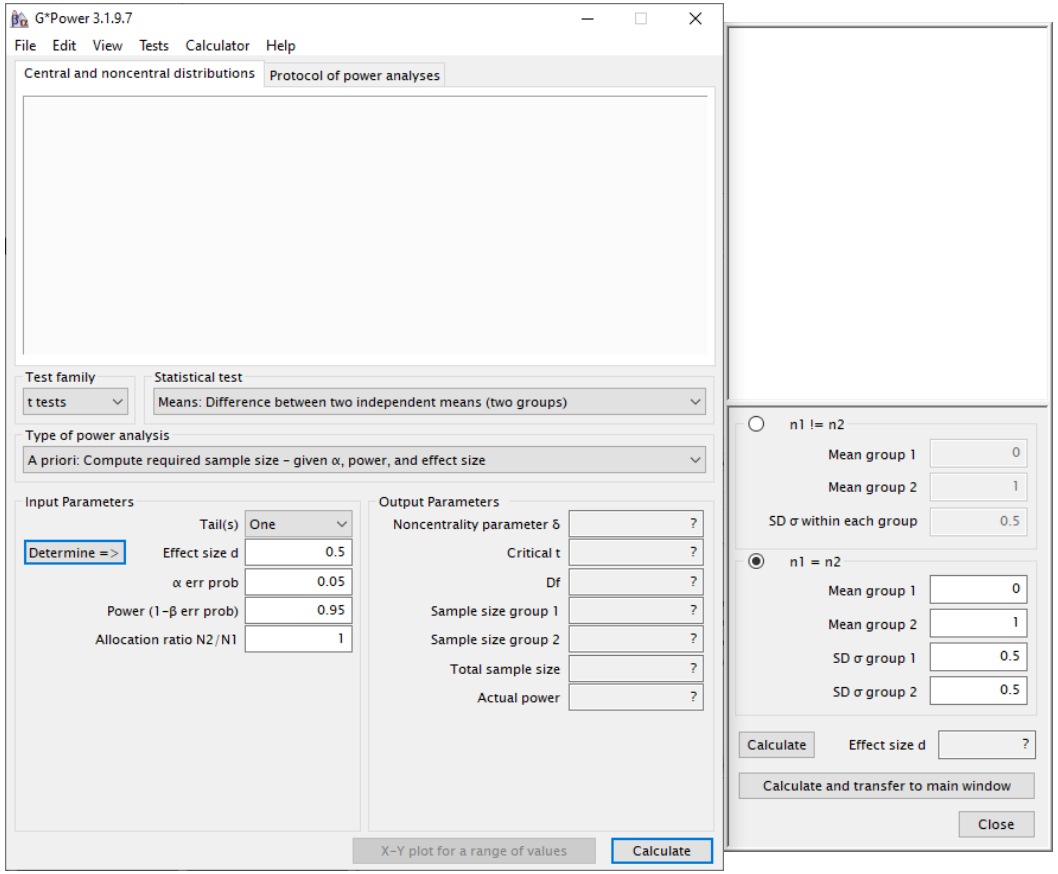
# Power analysis

- **Good news**:

there are packages that can do the power analysis for you, providing you have some prior knowledge of the key parameters!

**difference + standard deviation = effect size**

- R
- **G*Power**



# Power Analysis

## Comparing 2 proportions (Fisher's exact test)

**Exercise:**

- Scientists have come up with a solution that may reduce the number of lions being shot by farmers in Africa:
  - Painting eyes on cows' bottoms.
- Early trials suggest that lions are less likely to attack livestock when they think they're being watched
  - Fewer livestock attacks could help farmers and lions co-exist more peacefully.

- Pilot study over 6 weeks:
  - 3 out of 39 unpainted cows were killed by lions, none of the 23 painted cows from the same herd were killed.

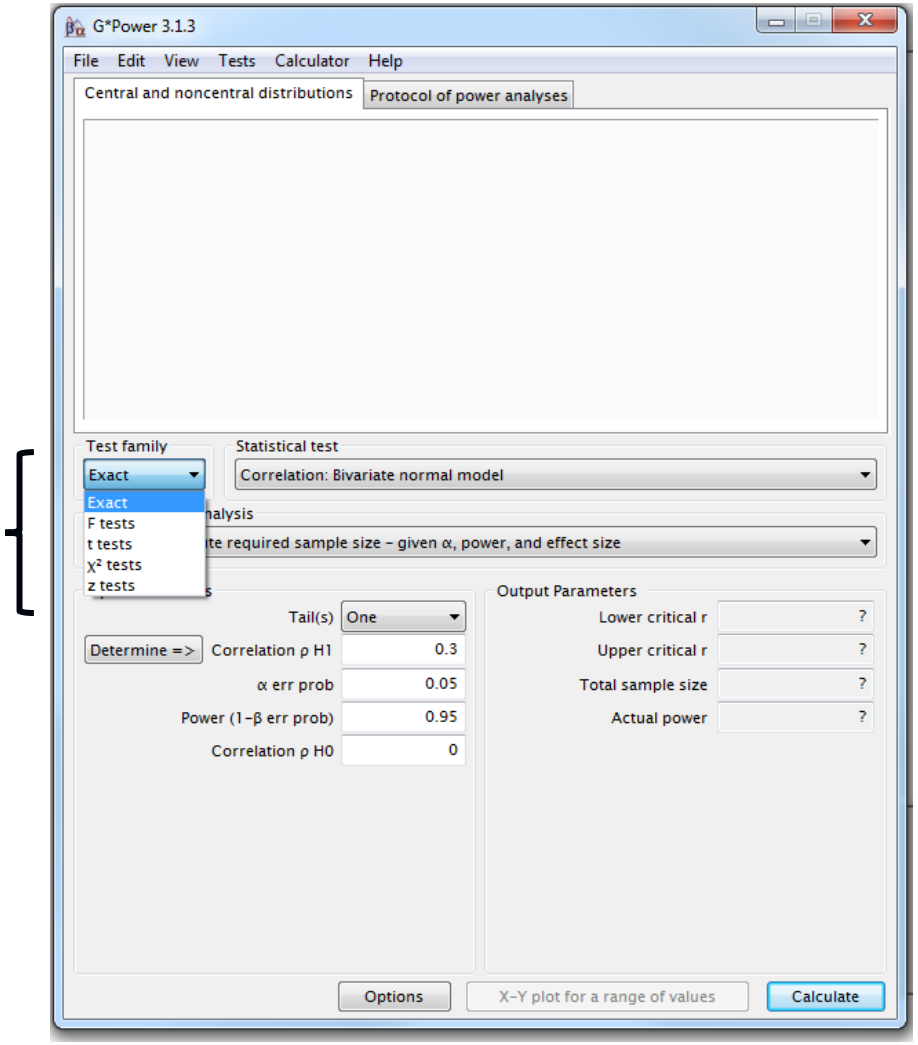
- **Tasks**:
  - Do you think the observed effect is meaningful to the extent that such a 'treatment' should be applied?
    Consider ethics, economics, conservation …
  - Run a power calculation to find out how many cows should be included in the study.

http://www.sciencealert.com/scientists-are-painting-eyes-on-cows-butts-to-stop-lions-getting-shot

# Power Analysis

## Comparing 2 proportions

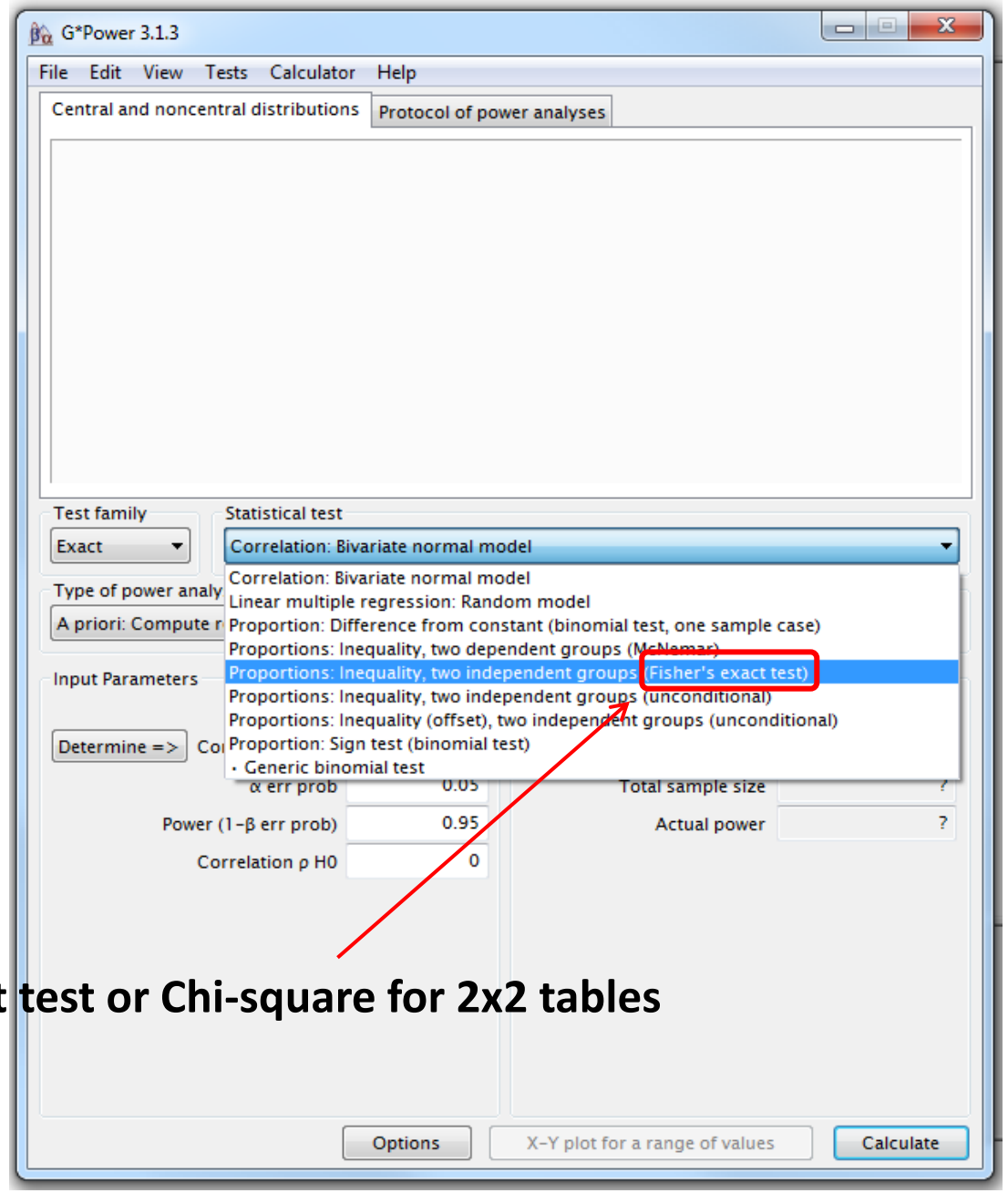
### Four steps to Power




**Step1:** choice of Test family

# G*Power

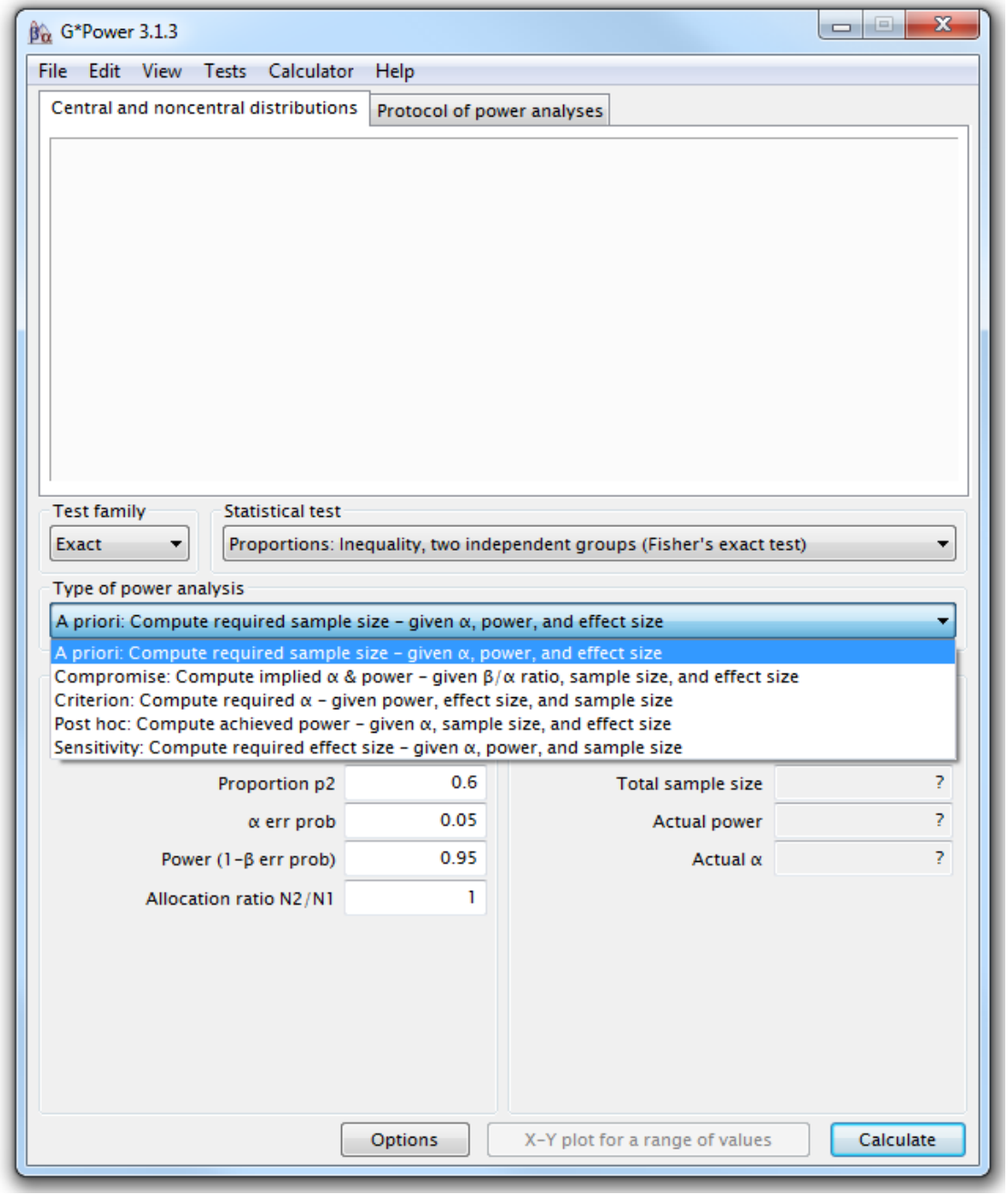
**Step 2 :** choice of Statistical test




**Fisher's exact test or Chi-square for 2x2 tables**

# G*Power

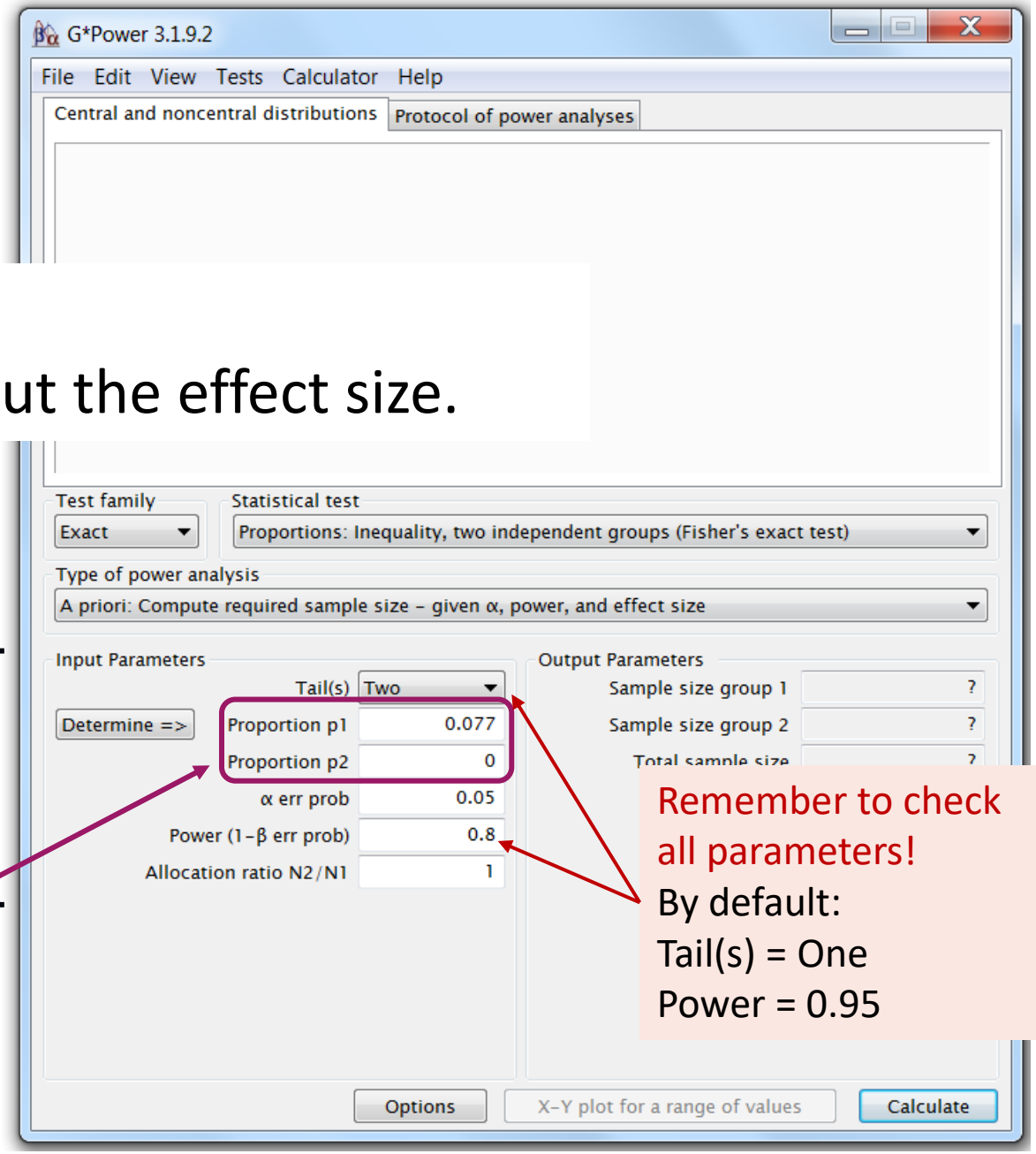
**Step 3:** Type of power analysis



# G*Power

**Step 4**: Choice of Parameters

Tricky bit: need information about the effect size.




Remember to check all parameters!

By default:

Tail(s) = One

Power = 0.95

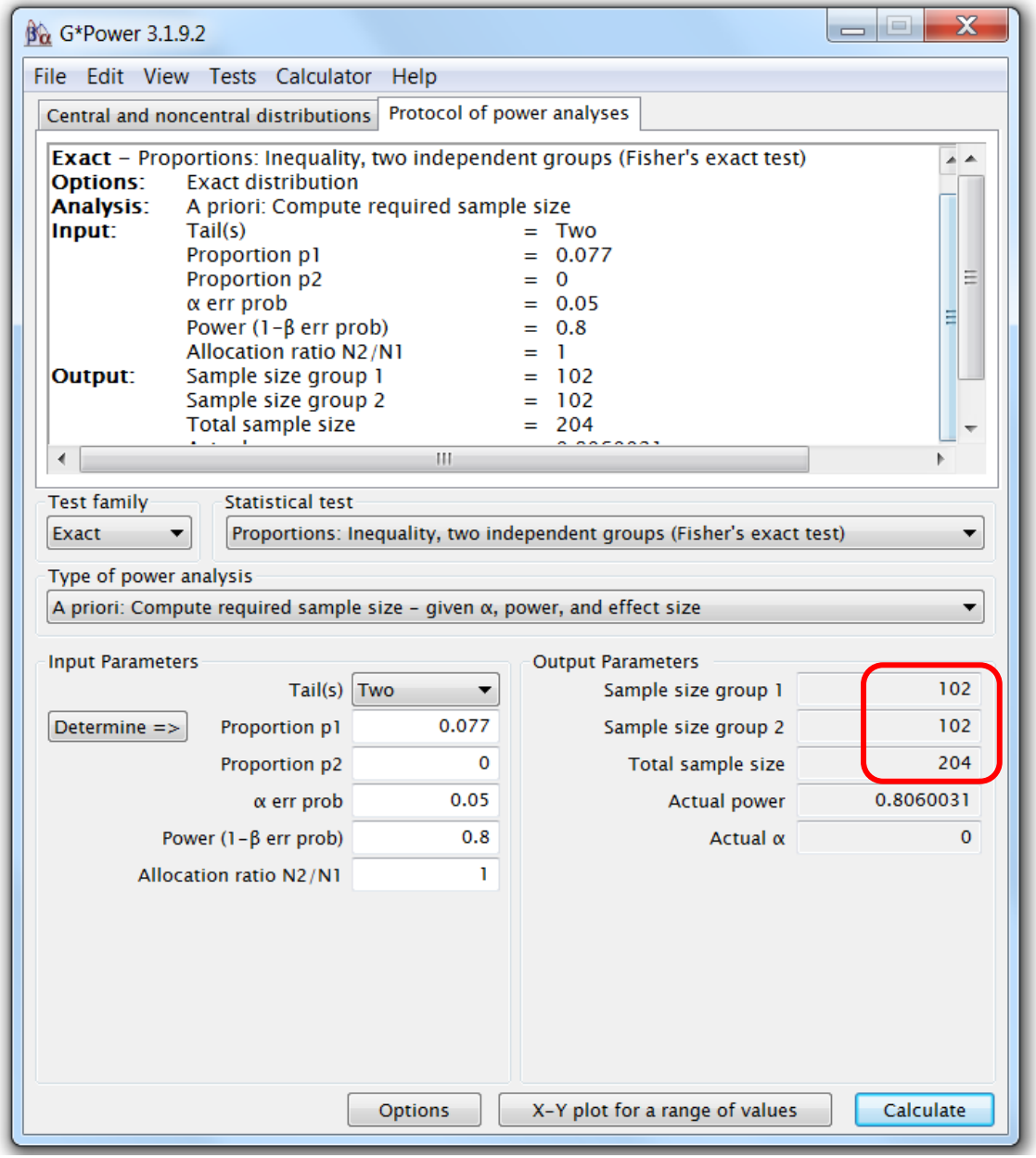
<u>Example case</u>:

0 cows killed in the painted group

3 out 39 in unpainted

# G*Power

Providing the preliminary results are to be trusted, to be able to pick up such a difference between the 2 groups, with a **power of 80%** and a **significance level of 5%**, we will need at least **102 cows in each group**.



Article | Open Access | Published: 07 August 2020

# Artificial eyespots on cattle reduce predation by large carnivores

Cameron Radford, John Weldon McNutt, Tracey Rogers, Ben Maslen & Neil Jordan

*Communications Biology* **3**, Article number: 430 (2020) | Cite this article

**49k** Accesses | **1** Citations | **2040** Altmetric | Metrics

## Abstract

Eyespots evolved independently in many taxa as anti-predator signals. There remains debate regarding whether eyespots function as diversion targets, predator mimics, conspicuous startling signals, deceptive detection, or a combination. Although eye patterns and gaze modify human behaviour, anti-predator eyespots do not occur naturally in contemporary mammals. Here we show that eyespots painted on cattle rumps were associated with reduced attacks by ambush carnivores (lions and leopards). Cattle painted with eyespots were significantly more likely to survive than were cross-marked and unmarked cattle, despite all treatment groups being similarly exposed to predation risk. While higher survival of eyespot-painted cattle supports the detection hypothesis, increased survival of cross-marked cattle suggests an effect of novel and conspicuous marks more generally. To our knowledge, this is the first time eyespots have been shown to deter large mammalian predators. Applying artificial marks to high-value livestock may therefore represent a cost-effective tool to reduce livestock predation.

**a** artificial eyespots (bicolour as pictured, or white/yellow inner only, or black outer only, for maximum contrast depending on cattle coat colour). **b** cross-marked procedural control (black or white depending on coat colour for contrast). **c** unmarked control. Images provided by C.R.

https://www.nature.com/articles/s42003-020-01156-0.pdf

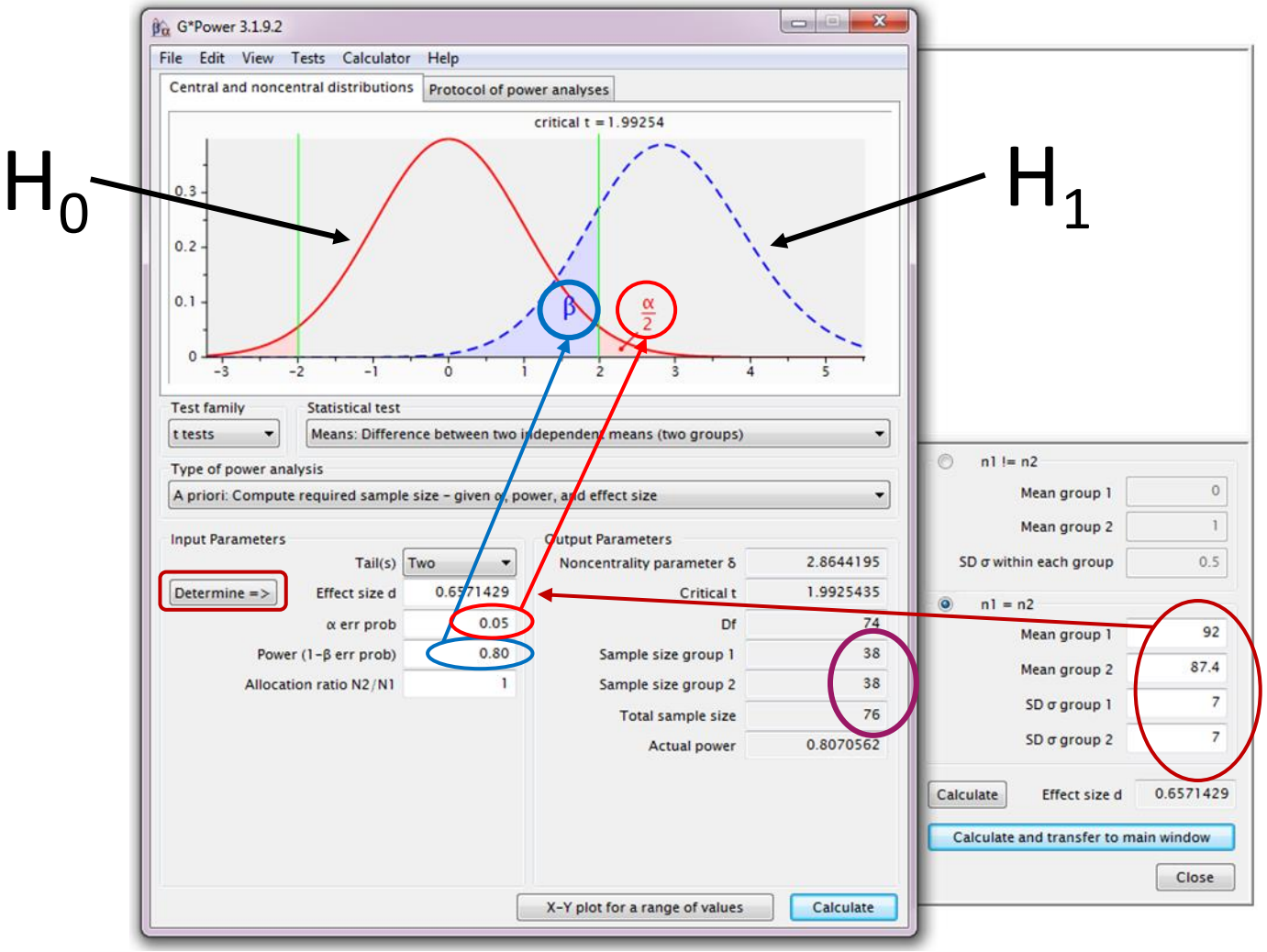
# Power Analysis

## Comparing 2 means (t-test)

**Exercise:**

- We want to know whether male and female coyotes differ in size
- No data from a pilot study but we have found some information in the literature:
  - In a study run in similar conditions as in the one we intend to run, **male coyotes** were found to measure: **92cm +/- 7cm (SD)**
- We expect a **5% difference** between the sexes
  **= smallest biologically meaningful difference**
- **Task**:
  - Run a power calculation to find out how many coyotes should be included in the study

# Power Analysis



# G*Power

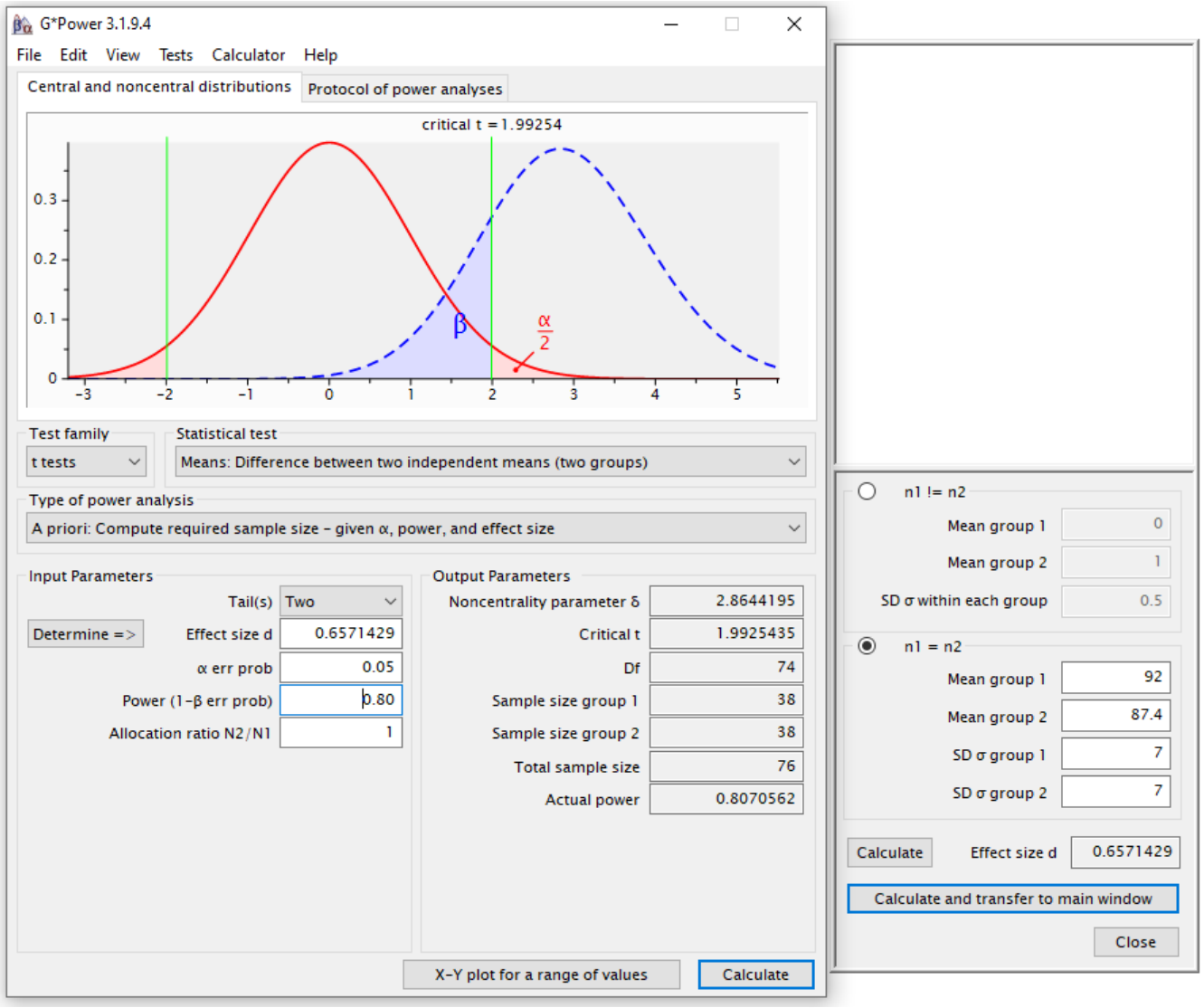
## Independent t-test

*A priori* Power analysis

**Example case:**

From a similar study, male coyotes were found to measure:
**92cm+/- 7cm (SD)**

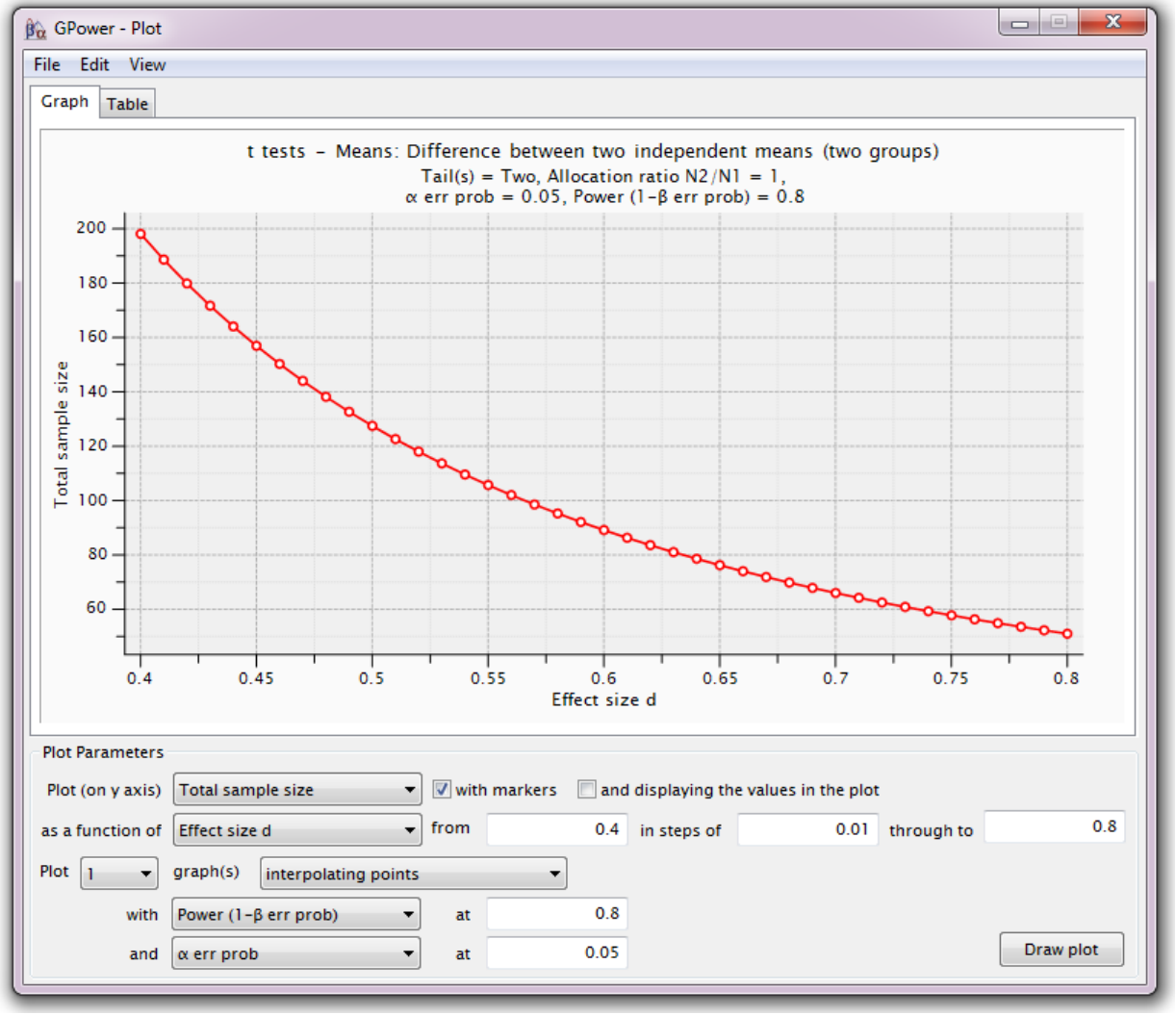
You expect a **5% difference** between the sexes with similar variability in the female sample




You need a sample size of **n=76 (2*38)**

# Power Analysis

For a range of sample sizes:



# Unequal sample sizes

- No simple trade-off – if need 2 groups of 30 → 20 and 40 = decreased power
  - **Unbalanced design = bigger total sample**
- **Solution 1** *(old school)*:

  Step 1: power calculation for equal sample size

  Step 2: adjustment

**Where**:
- N = total sample size in unbalanced design
- n = sample size per group in balanced design
- $n_1$ = group 1 size
- $n_2$ = group 2 size
- k = ratio $n_2/n_1$

$$N = \frac{2n(1+k)^2}{4k}$$

$$n_1 = \frac{N}{(1+k)}$$

$$n_2 = \frac{kN}{(1+k)}$$

Cow example:
- Balanced design: **n = 102**
- If unpainted group is 2 times bigger than painted (k=0.5):

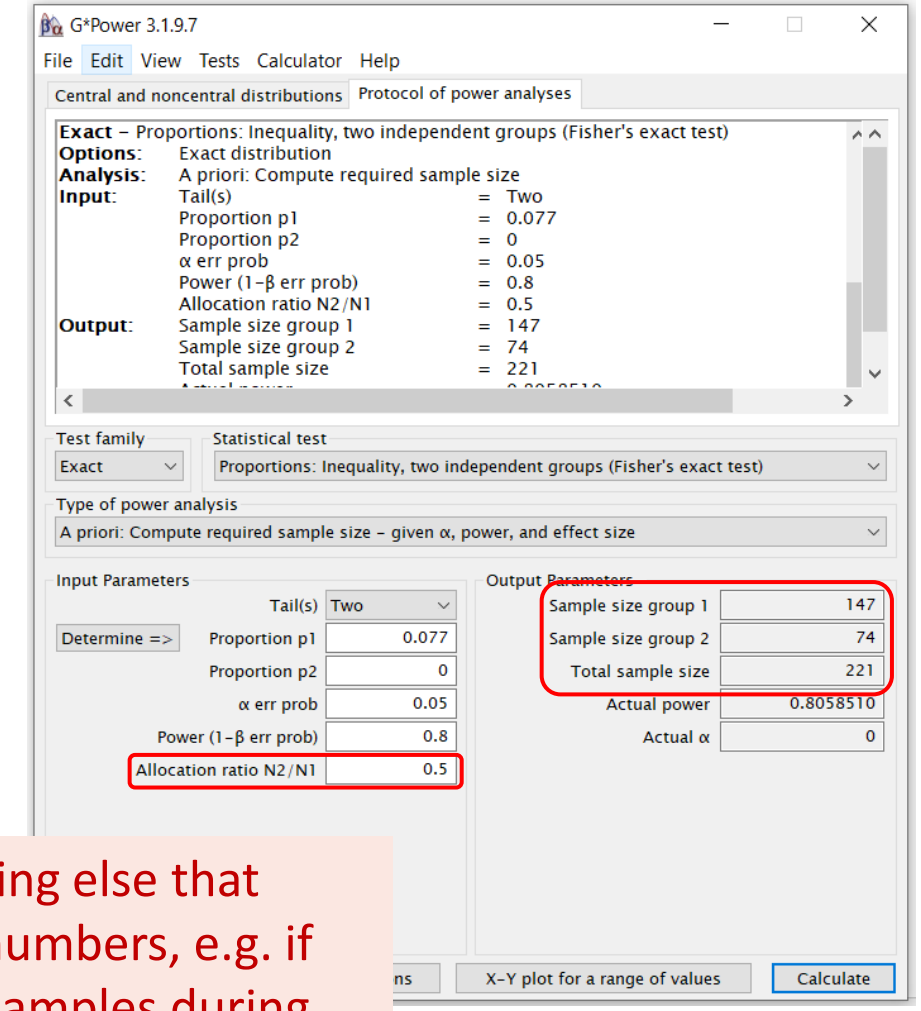
$$N = \frac{2 \times 102 \times (1+0.5)^2}{4 \times 0.5} = 229.5 \approx 230$$

Unpainted butts **($n_1$)=153** Painted butts **($n_2$)=77**

- Balanced design: **n = 2*102 = 204**
- Unbalanced design: **n= 153+77 = 230**

# Unequal sample sizes

- **Solution 2:**
  Use G*Power




Also consider anything else that might impact final numbers, e.g. if likely to lose some samples during experiment

Different methods give slightly different sample sizes:

- Using adjustment
  - Painted ($n_1$) = 77
  - Unpainted ($n_2$) = 153
  - **Total sample = 230**
- Using G*Power:
  - Painted ($n_2$) = 74
  - Unpainted ($n_1$) = 147
  - **Total sample = 221**

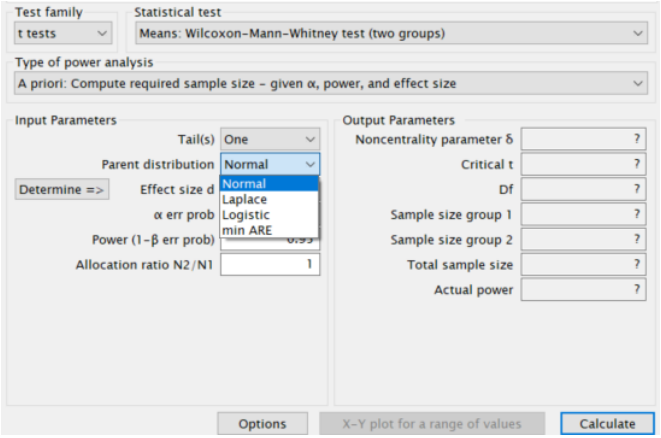
# Non-parametric tests

- Do not assume data come from a Gaussian/normal distribution
  - Based on ranking values from low to high
  - Almost always less powerful

- Proper power calculations need to specify which kind of distribution we are dealing with – not easy

- Non-parametric usually do not require more than 15% additional subjects compared to parametric

- Crude rule of thumb:
  - **Compute sample size required for a parametric test and add 15%**



# What happens if we ignore the power of a test?

- **Misinterpretation** of the results

- Never ever interpret p-values without context!
  - **Significant p-value (<0.05)**: but what is the difference?
    - >= smallest meaningful difference: **exciting effect**
    - < smallest meaningful difference: essentially a false positive/type 1 error
      - Too big sample, overpowered – difference has **no biological relevance**

  - **Not significant p-value (>0.05)**: but how big was the sample?
    - Big enough = enough power: **no effect**
    - Not big enough = underpowered: potentially a false negative/type 2 error
      - Too small sample – **potentially miss a meaningful difference**

# Exercise 1

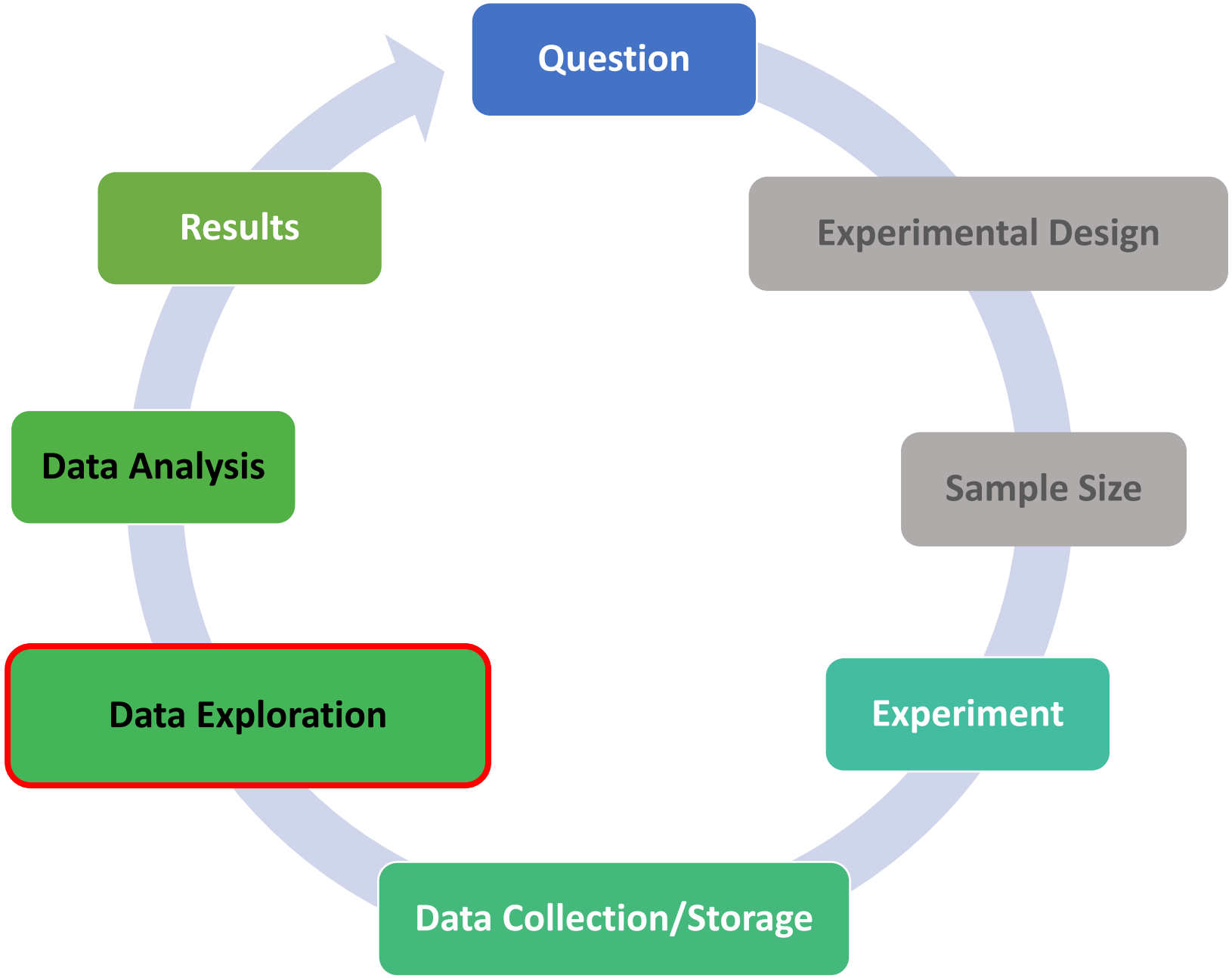
Babraham Institute

# Descriptive Stats and Data Exploration

Hayley Carr & Anne Segonds-Pichon

v2024-05

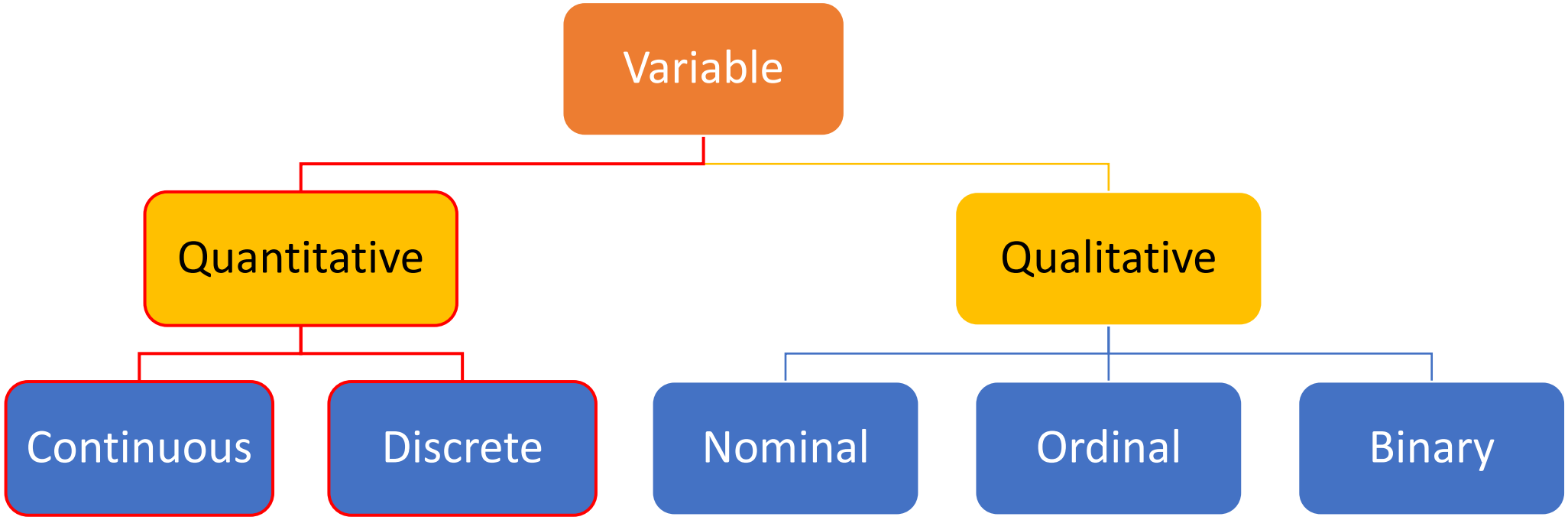
Babraham Bioinformatics


Question
Experimental Design
Sample Size
Experiment
Data Collection/Storage
Data Exploration
Data Analysis
Results

# Variable

- Quantitative
  - Continuous
  - Discrete
- Qualitative
  - Nominal
  - Ordinal
  - Binary

https://github.com/allisonhorst/stats-illustrations#other-stats-artwork



https://github.com/allisonhorst/stats-illustrations#other-stats-artwork

# Quantitative data

- They take numerical values (units of measurement)

- Discrete: obtained by counting
  - Example: number of students in a class
  - values vary by finite specific steps
- Continuous: obtained by measuring
  - Example: height of students in a class
  - any values – can have decimal places

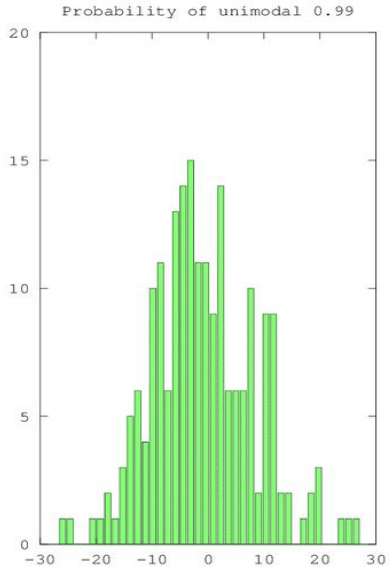
https://github.com/allisonhorst/stats-illustrations#other-stats-artwork

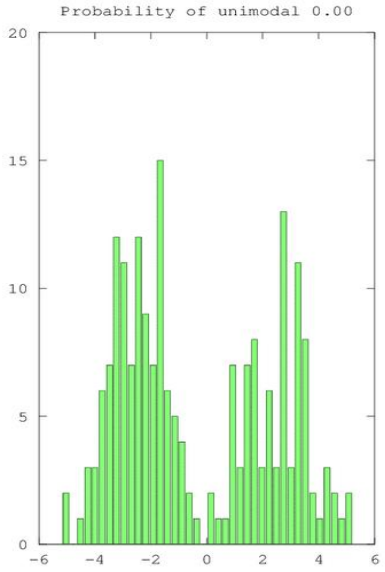
- They can be described by a series of parameters:
  - Mean, variance, standard deviation, standard error and confidence interval

# Measures of central tendency

## Mode and Median

- **Mode:** most commonly occurring value in a distribution



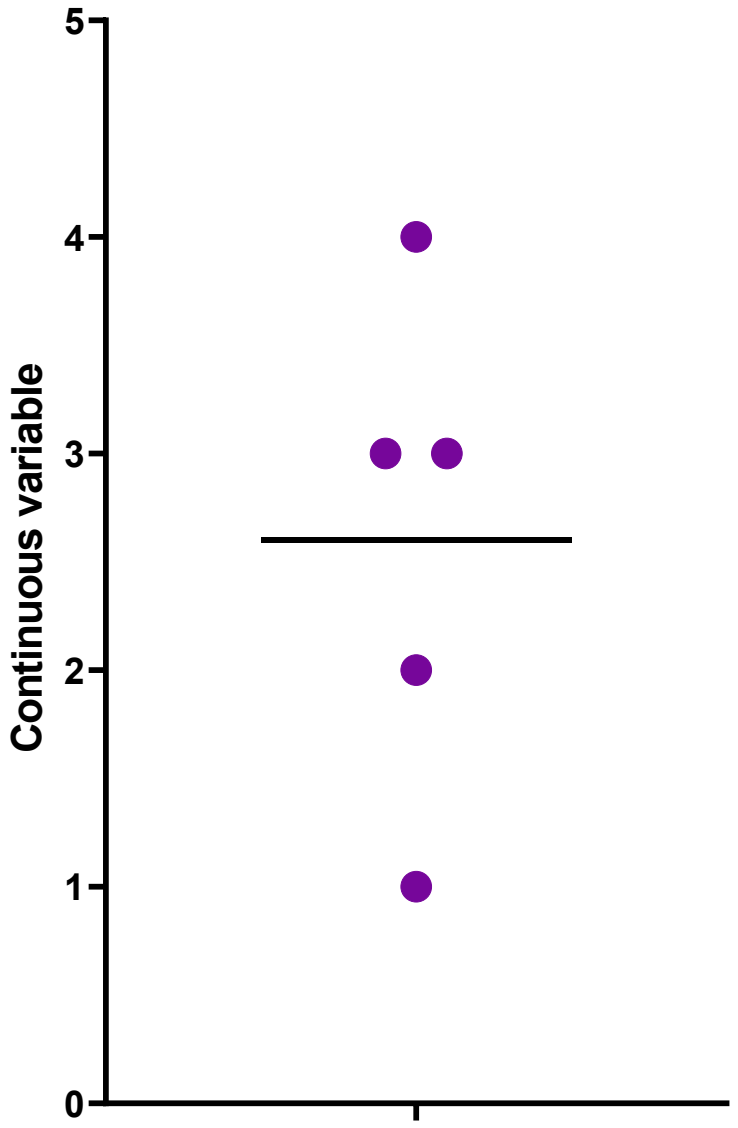

- **Median:** value exactly in the middle of an ordered set of numbers

Example 1: 18 27 34 52 54 59 61 68 78 82 85 87 91 93 100, Median = 68
Example 2: 18 27 27 34 52 52 59 61 68 68 85 85 85 90, Median = 60

# Measures of central tendency

## Mean

- Definition: average of all values in a column
- Example: mean of: 1, 2, 3, 3 and 4 = (1+2+3+3+4)/5 = 2.6
- The mean is a model because it summaries the data
- How do we know that it is an accurate model?
  - Difference between the real data and the model created



# Measures of dispersion

- Calculate the magnitude of the differences between each data and the mean

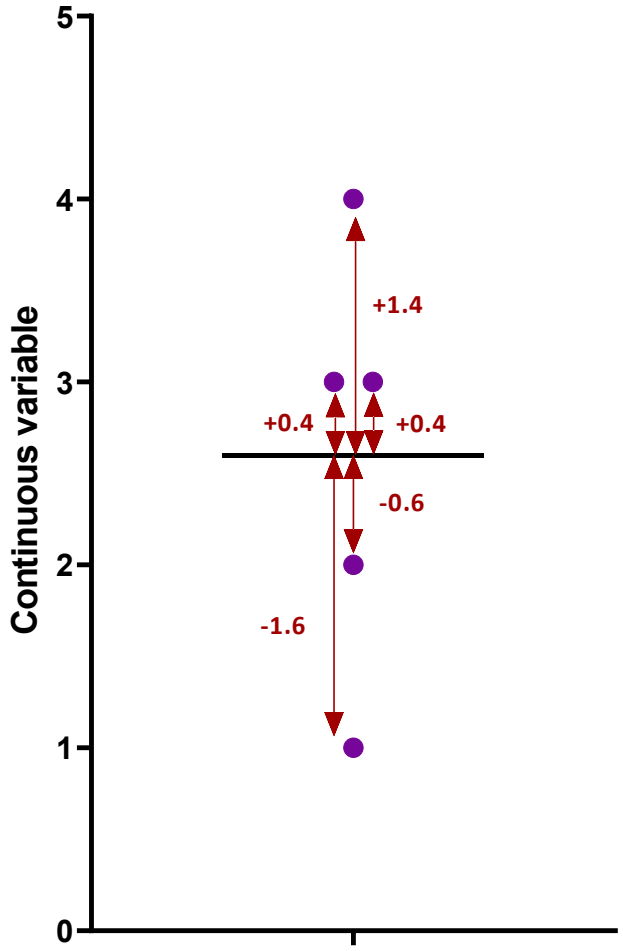
- Total error = sum of differences

  $= \Sigma(x_i - \overline{x})$ = -1.6 - 0.6 + 0.4 + 0.4 + 1.4 = 0

  No errors: positive and negative cancel each other out

- To solve that problem we square the errors

  → Sum of squared errors (SS)



# Sum of Squared errors (SS)

- Sum of squared errors = $(SS) = \Sigma(x_i - \bar{x})^2$

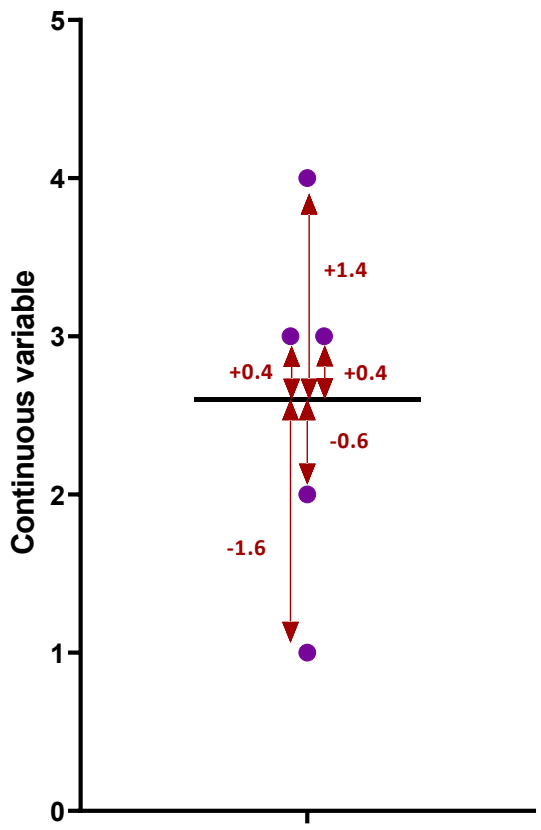
  $= (-1.6)^2 + (-0.6)^2 + (0.4)^2 + (0.4)^2 + (1.4)^2$

  $= 5.20$

- Good measure of the accuracy of the model

- Depends on amount of data: larger sample → larger SS
  - Account for number of observations (N) by dividing SS by N-1 (degrees of freedom)

    → the **variance** ($S^2$) = SS/N-1



# Degrees of freedom

$$variance\ (s^2) = \frac{SS}{N-1} = \frac{\Sigma\,(x_i - \boxed{\bar{x}})^2}{\boxed{N-1}}$$

- To calculate the variance, we need the mean
- If we know the mean, we do not need all the values in the sample to calculate the variance
- Example: Sample: n = 5, Mean ($\bar{x}$) = 2.6
  - 2.6 x 5 – (1+2+3+3) = 4

$$\frac{\boxed{1+2+3+3}+4}{5} = 2.6$$

- Once we know the mean, we only need to know the first 4 numbers (N-1) and we can calculate the last number

# Degrees of freedom

$$variance\ (s^2) = \frac{SS}{N-1} = \frac{\Sigma\,(x_i - \boxed{\bar{x}})^2}{\boxed{N-1}}$$

- The last ($n^{th}$) value in the sample is no longer independent, **is not free.**

**n – 1 degrees of freedom**

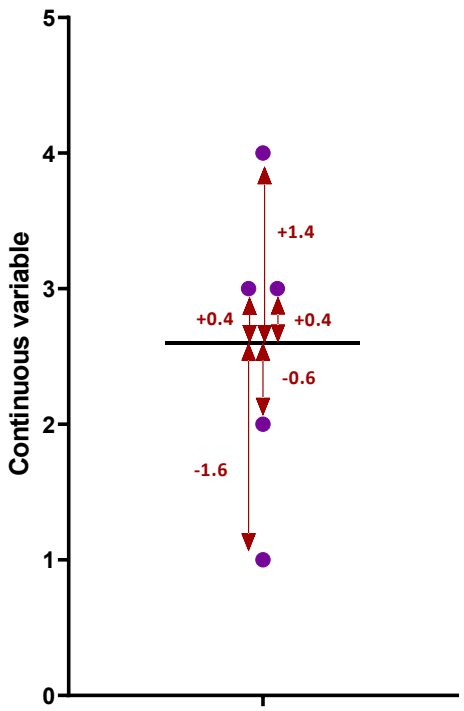
- Because we know the mean, the variance does not depend on all of the values of the sample, only on n-1 of the values

# Variance and standard deviation

- $variance\ (s^2) = \frac{SS}{N-1} = \frac{\Sigma\,(x_i - \overline{x})^2}{N-1} = \frac{5.20}{4} = 1.3$

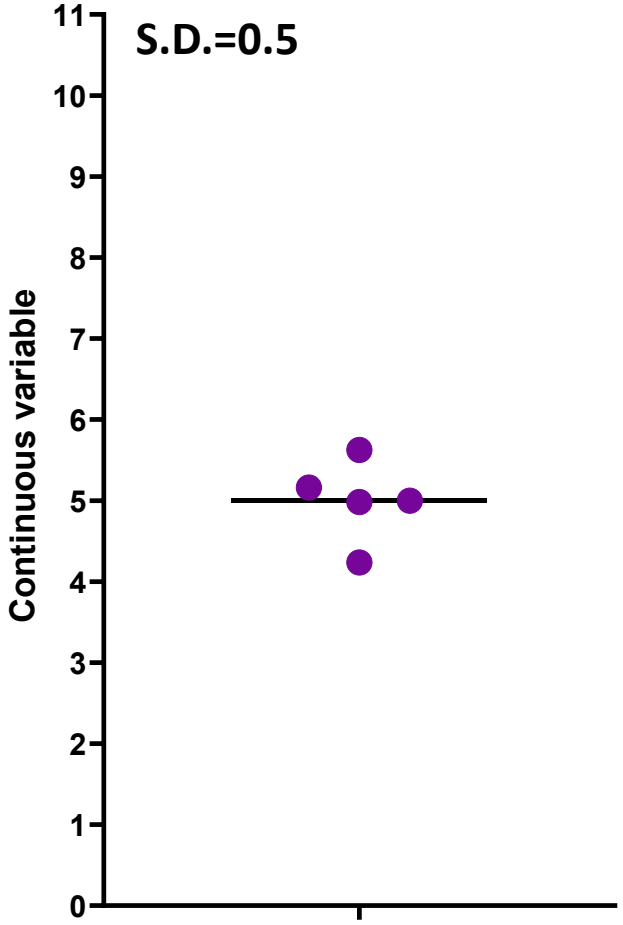
- Problem with variance: in squared units
  - Take the square root to get the same unit as the original measure
    → the **standard deviation**

$$S.D. = \sqrt{(SS/N-1)} = \sqrt{(s^2)} = s = \sqrt{1.3} = 1.14$$




- SD = a measure of how well the mean represents the data.

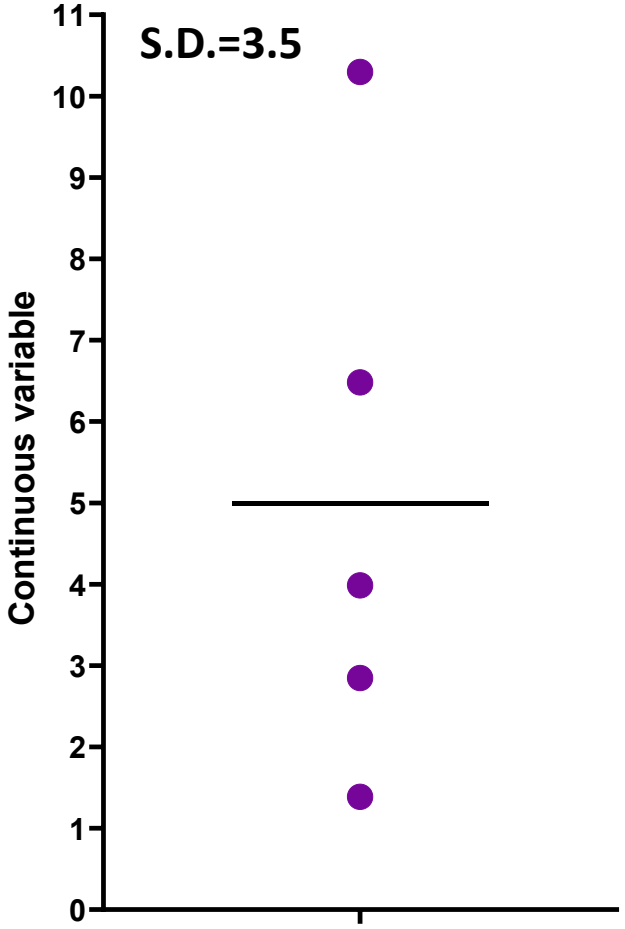
# Standard deviation




Small S.D.

data close to the mean:

good fit of the data

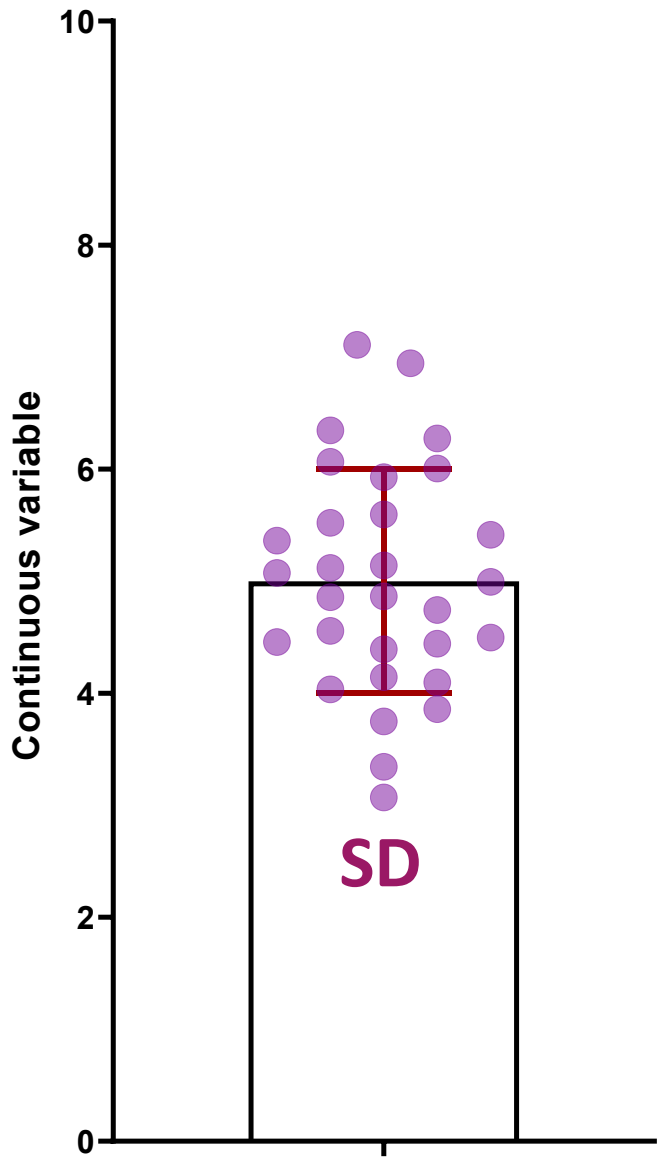


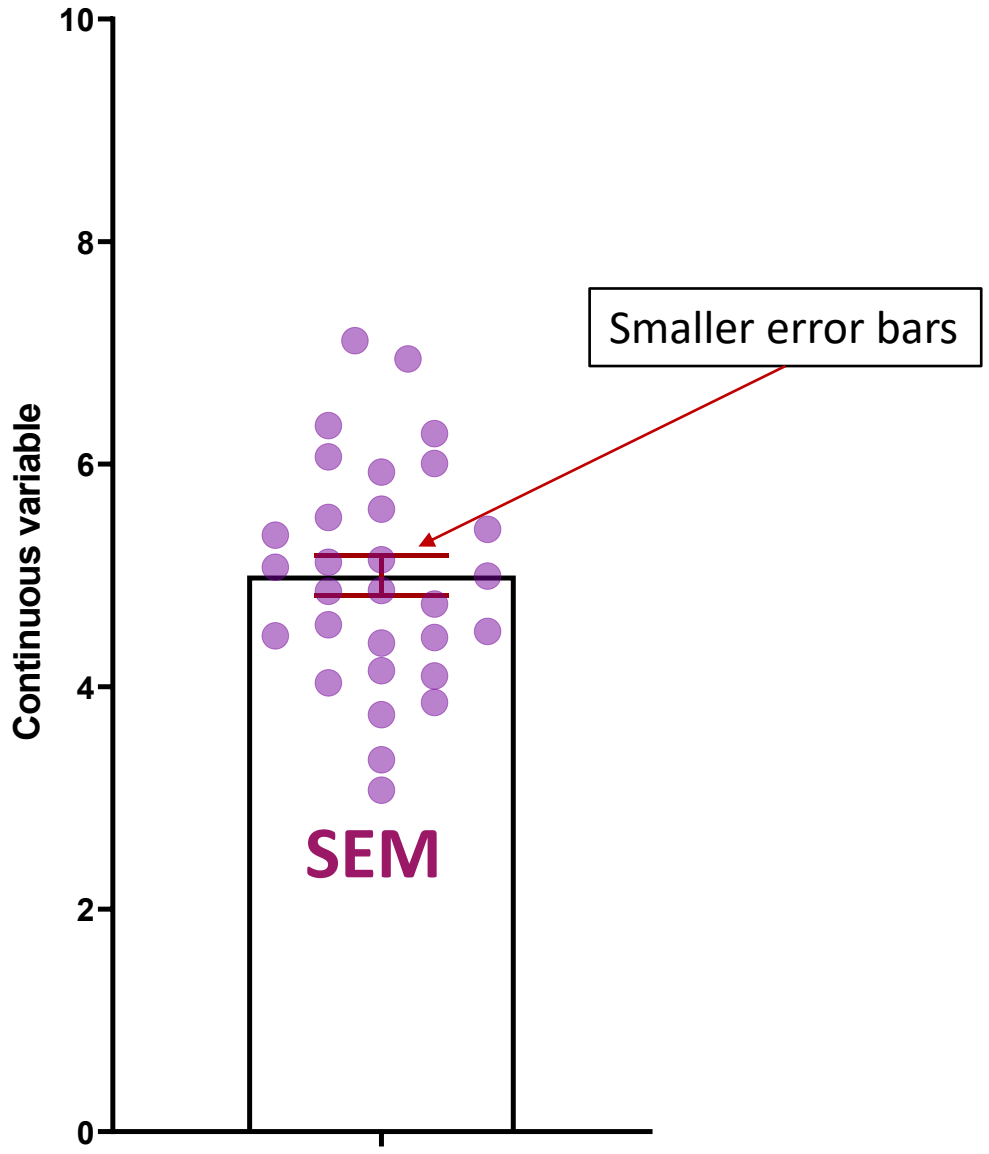

Large S.D.

data distant from the mean:

not an accurate representation

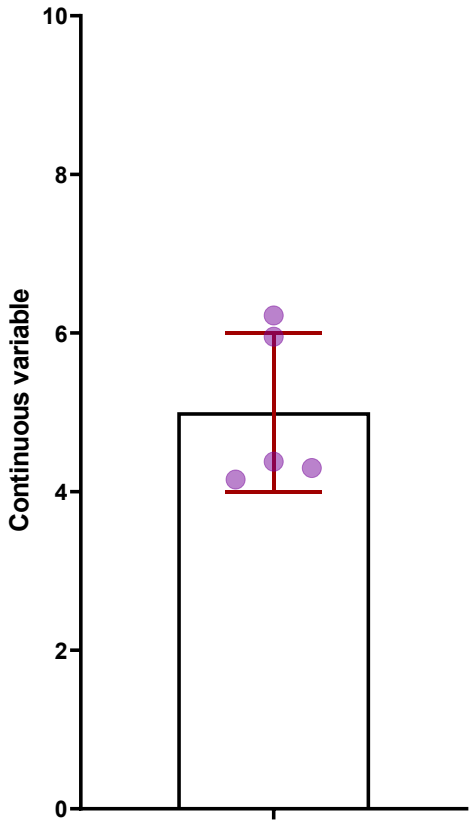
# Standard Deviation (SD) or Standard Error Mean (SEM)?

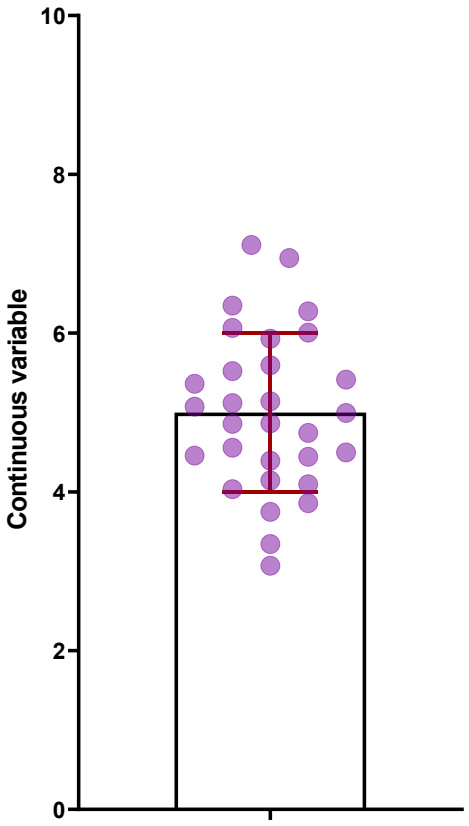


# Standard Deviation

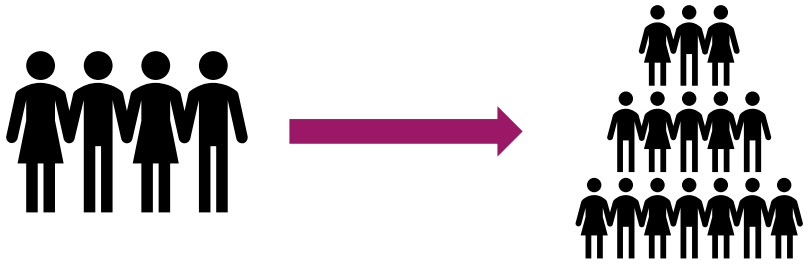
- The SD quantifies how much the values vary from one another
  → **scatter or spread**
- Does not change predictably as you acquire more data



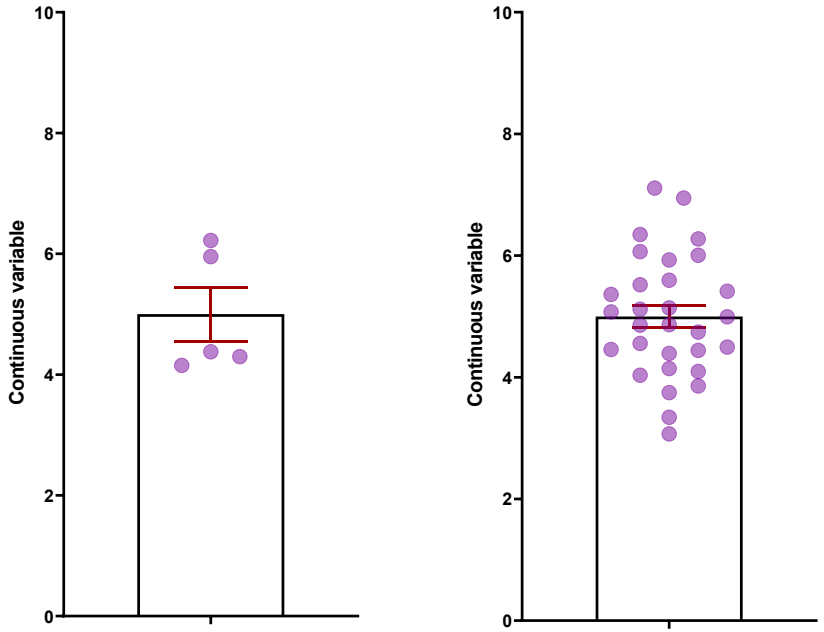



# Standard Error of the Mean

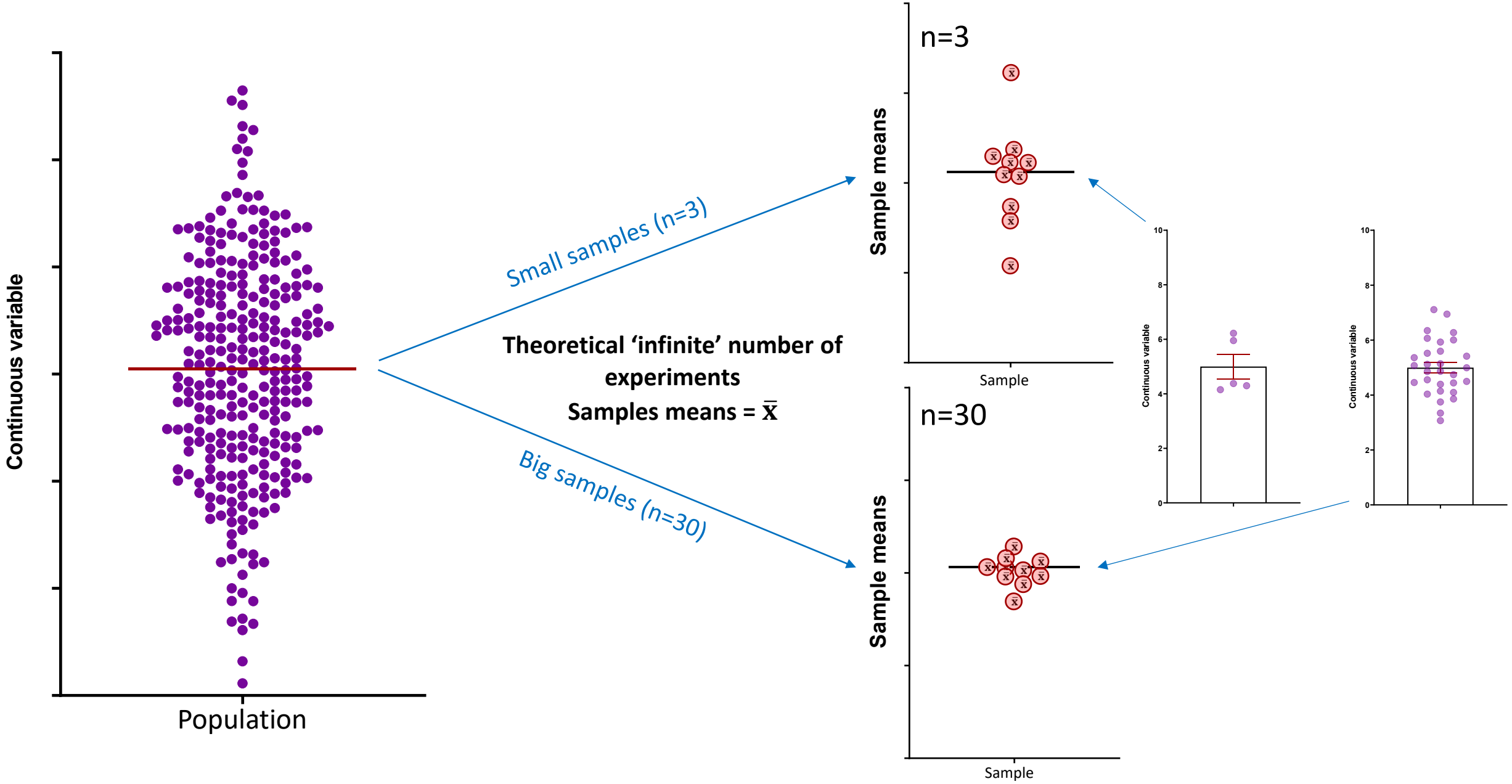
$$\text{SEM} = \frac{\text{SD}}{\sqrt{N}}$$



- The SEM quantifies how accurately we know the **true mean** of the population
  - Takes into account: SD + sample size
  - Make **inferences** about the population
- The SEM gets smaller as the sample gets larger
  - Mean of a large sample likely closer to true mean than mean of a small sample



# The SEM and the sample size



# SD or SEM ?

Whichever you choose make sure to report it accurately!

- If we want to show the variation among values:

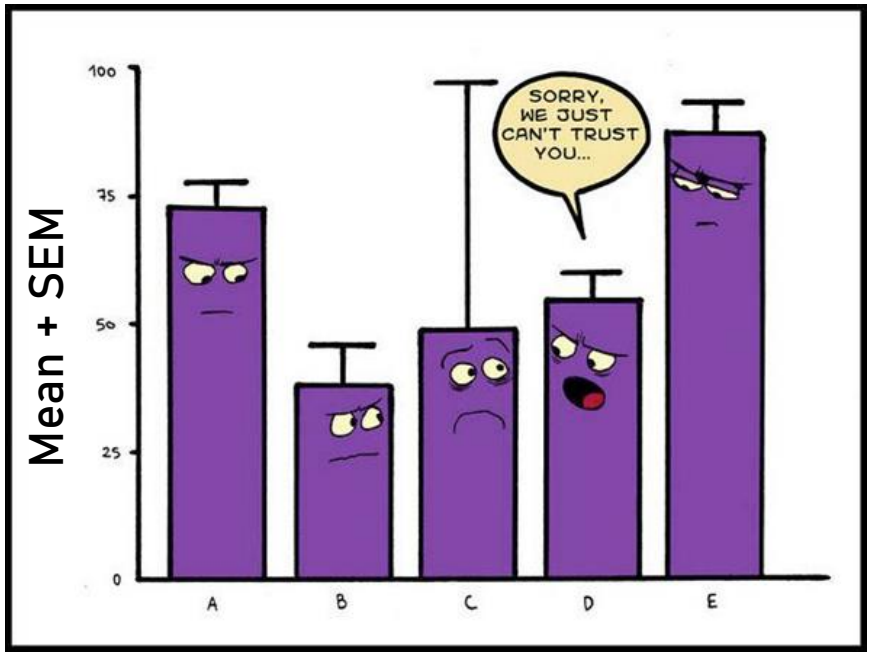
  → Report the **SD**

- If we want to show how precisely we have determined the population mean:

  → Report the **SEM**

- Preferably show all data points and the SEM

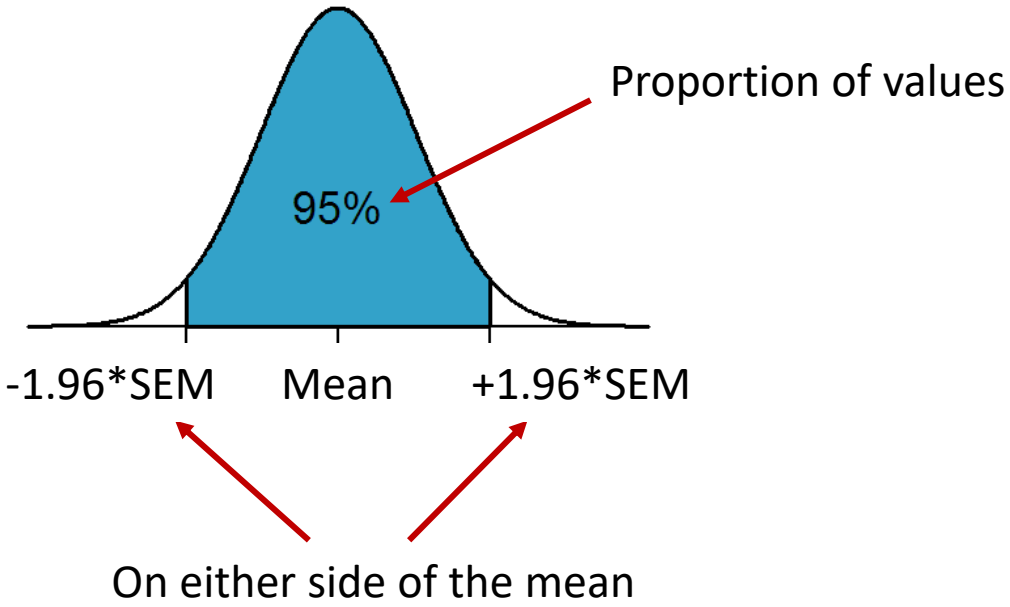
  → Both variation and precision

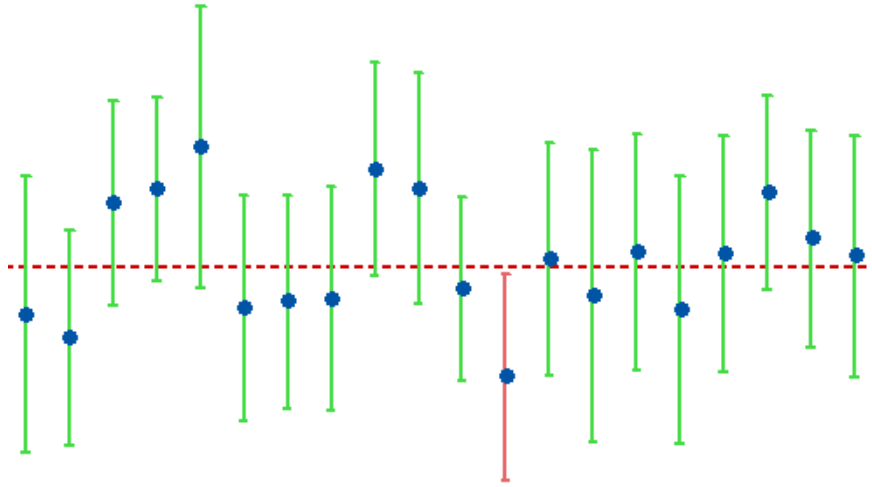



https://lymielynn.medium.com/a-little-closer-to-cooks-distance-e8cc923a3250

# Confidence interval

- Range of values that we can be 95% confident contains the true mean of the population
  - Limits of 95% CI: [Mean - 1.96*SEM; Mean + 1.96*SEM] (SEM = SD/√N)
- On average 19/20 experiments include the population mean

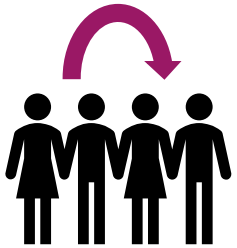


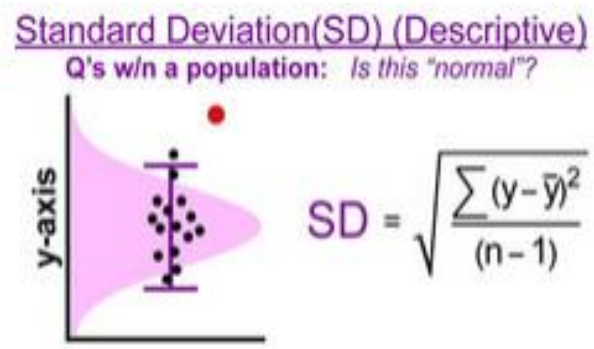
https://statisticsbyjim.com/hypothesis-testing/confidence-interval/

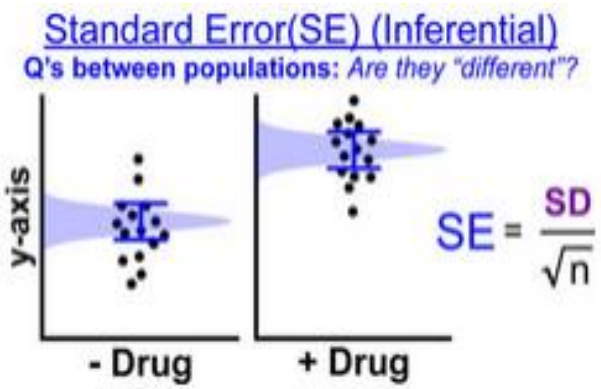
# To recap

- The Standard Deviation is **descriptive**
  - Just about the sample

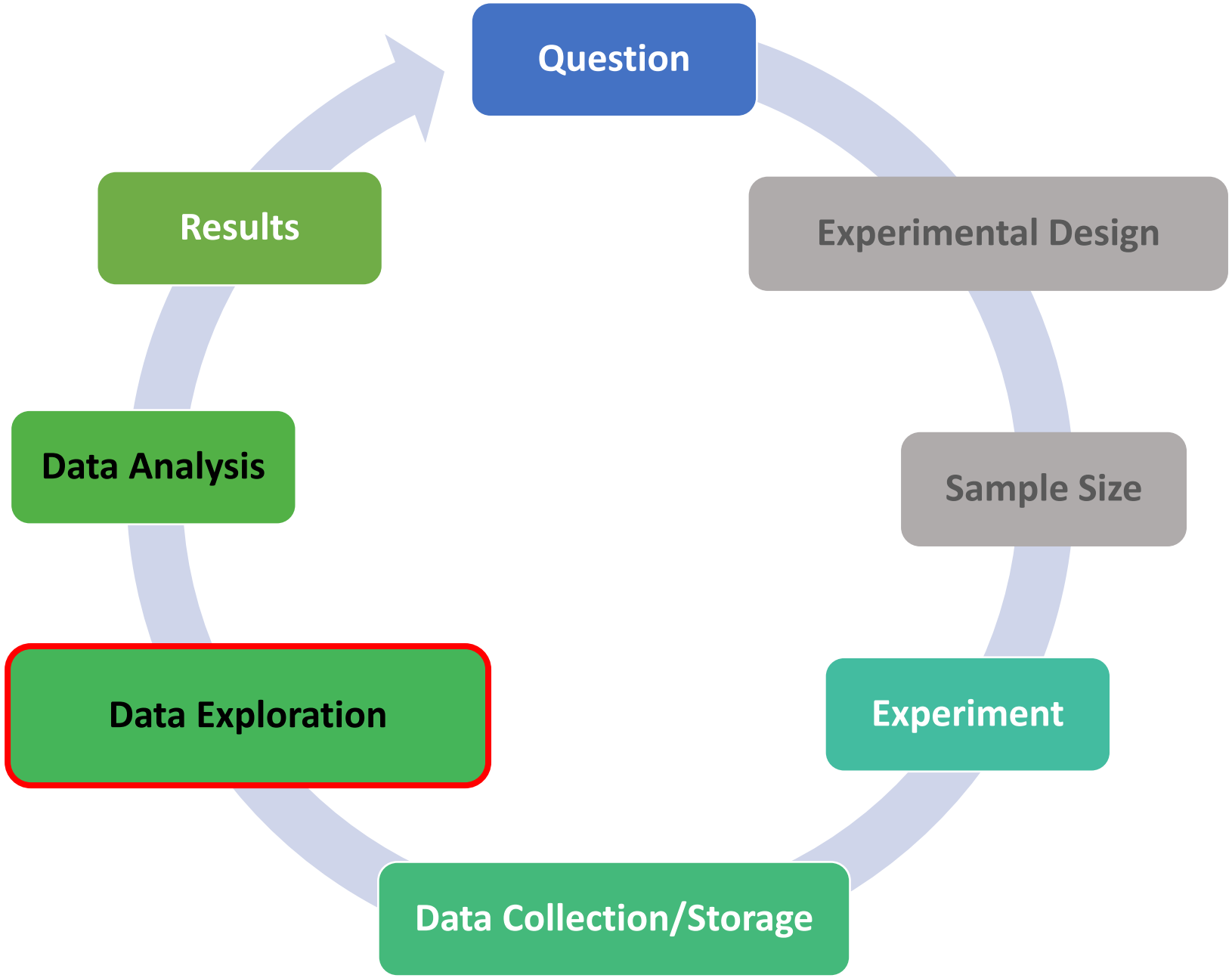


- The Standard Error and the Confidence Interval are **inferential**
  - Sample → General Population

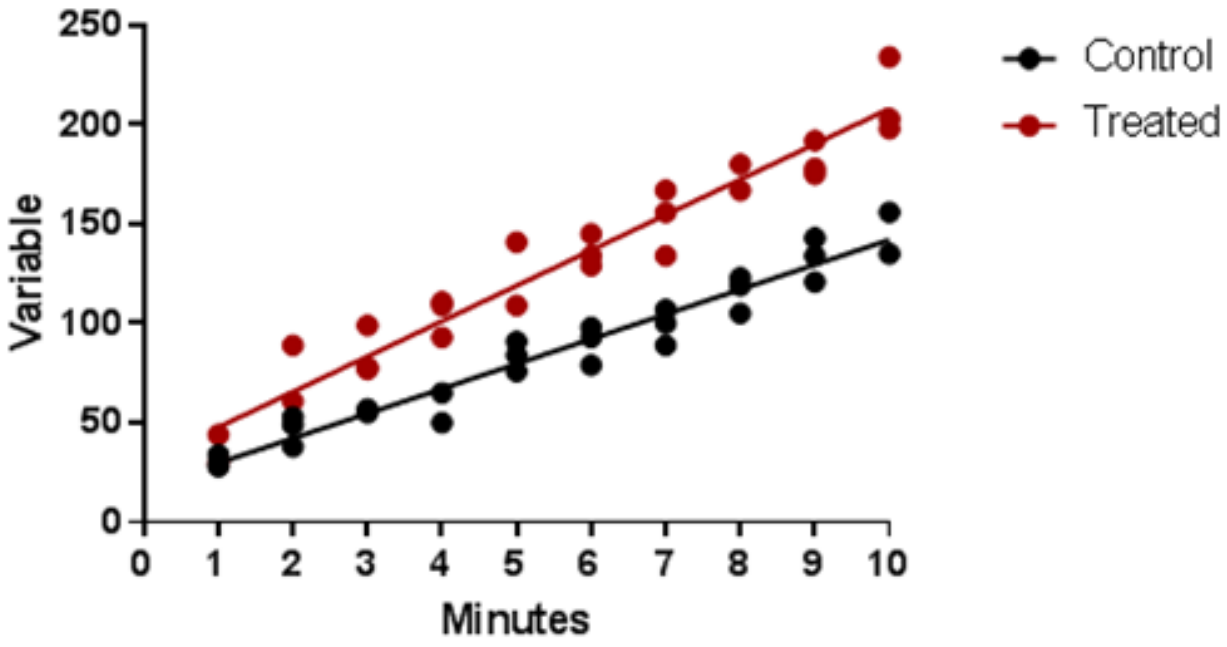


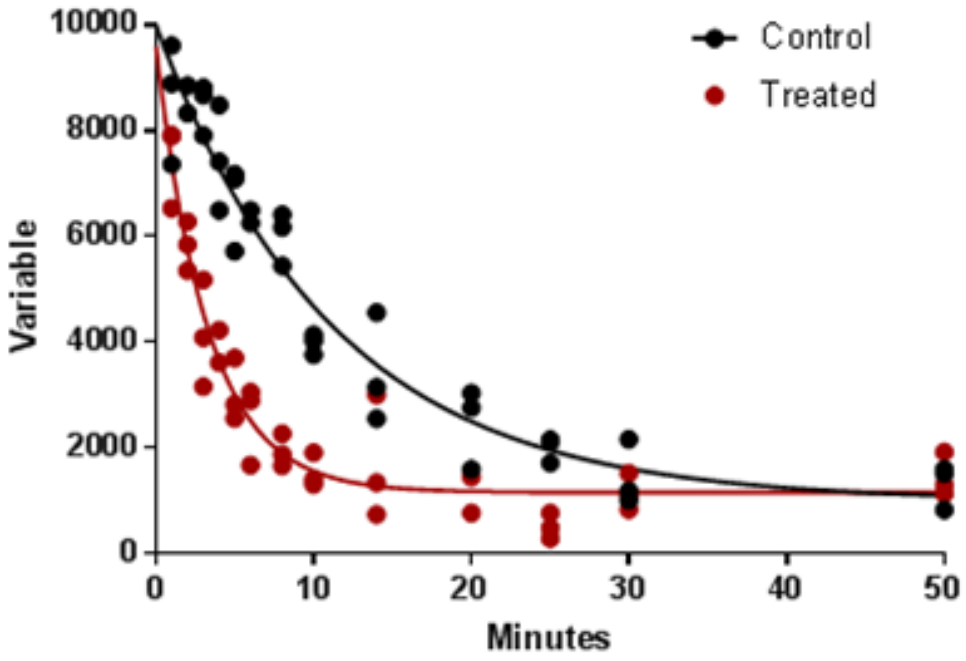




Statistical tests are also **inferential**

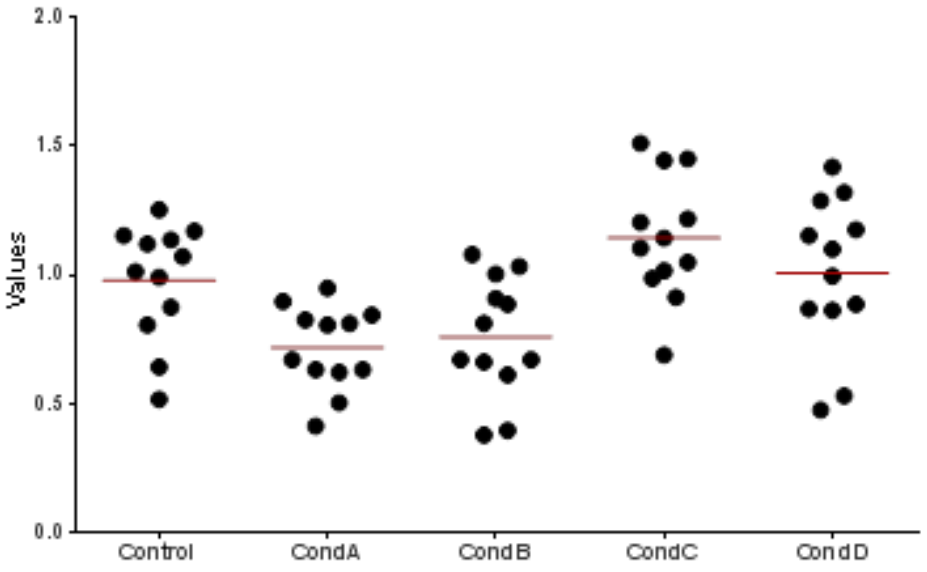

Question
Experimental Design
Sample Size
Experiment
Data Collection/Storage
Data Exploration
Data Analysis
Results

# Quantitative data: Scatterplot

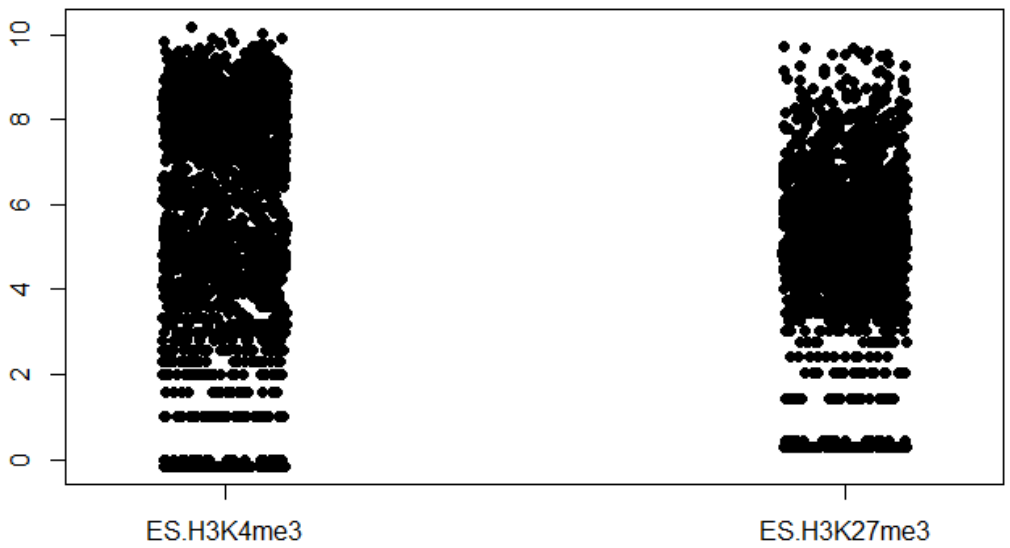


# Quantitative data: Scatterplot/stripchart

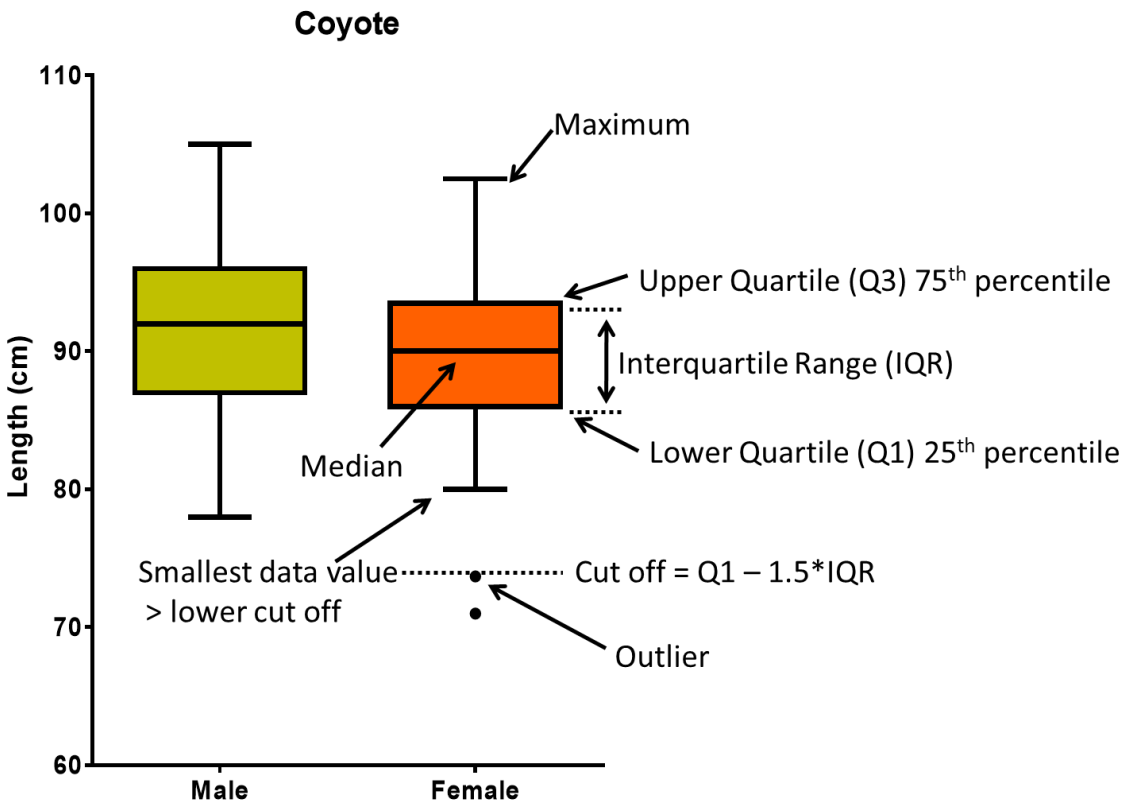


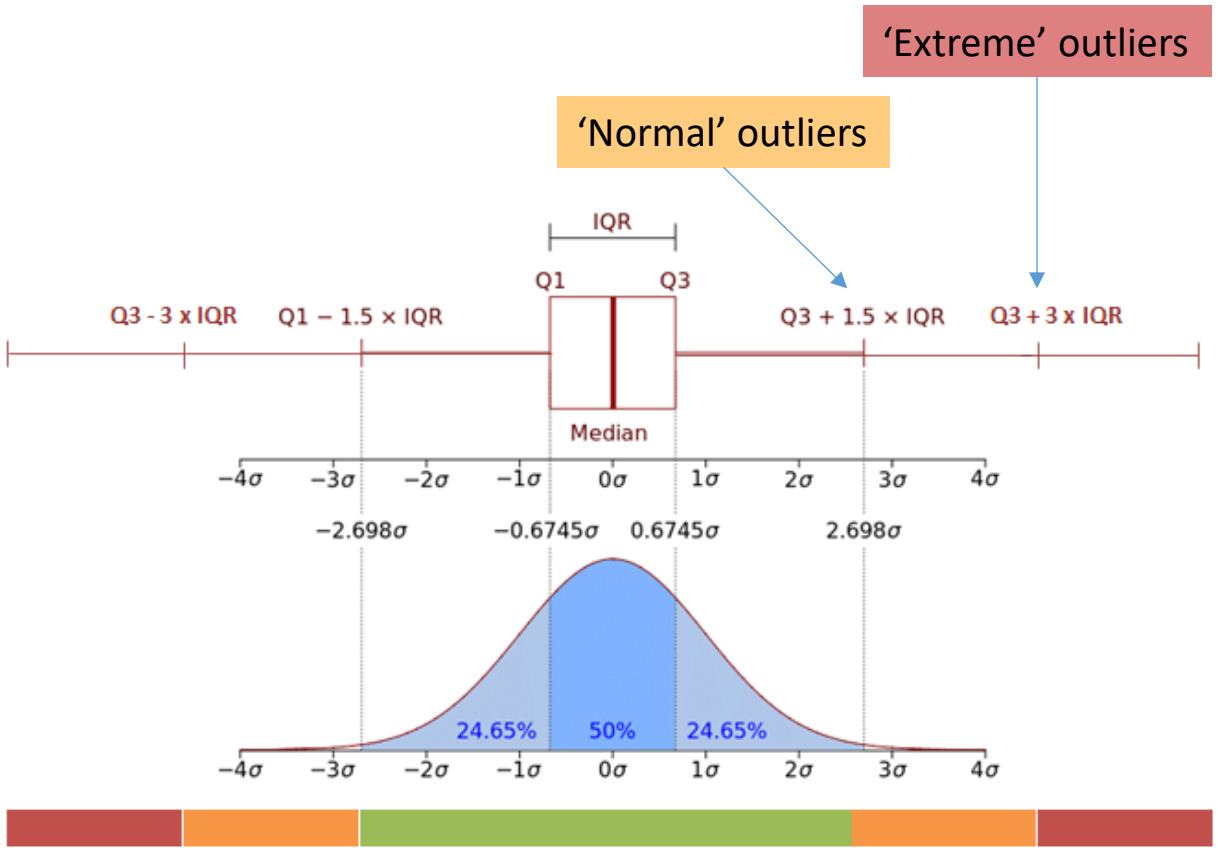

Small sample




Big sample

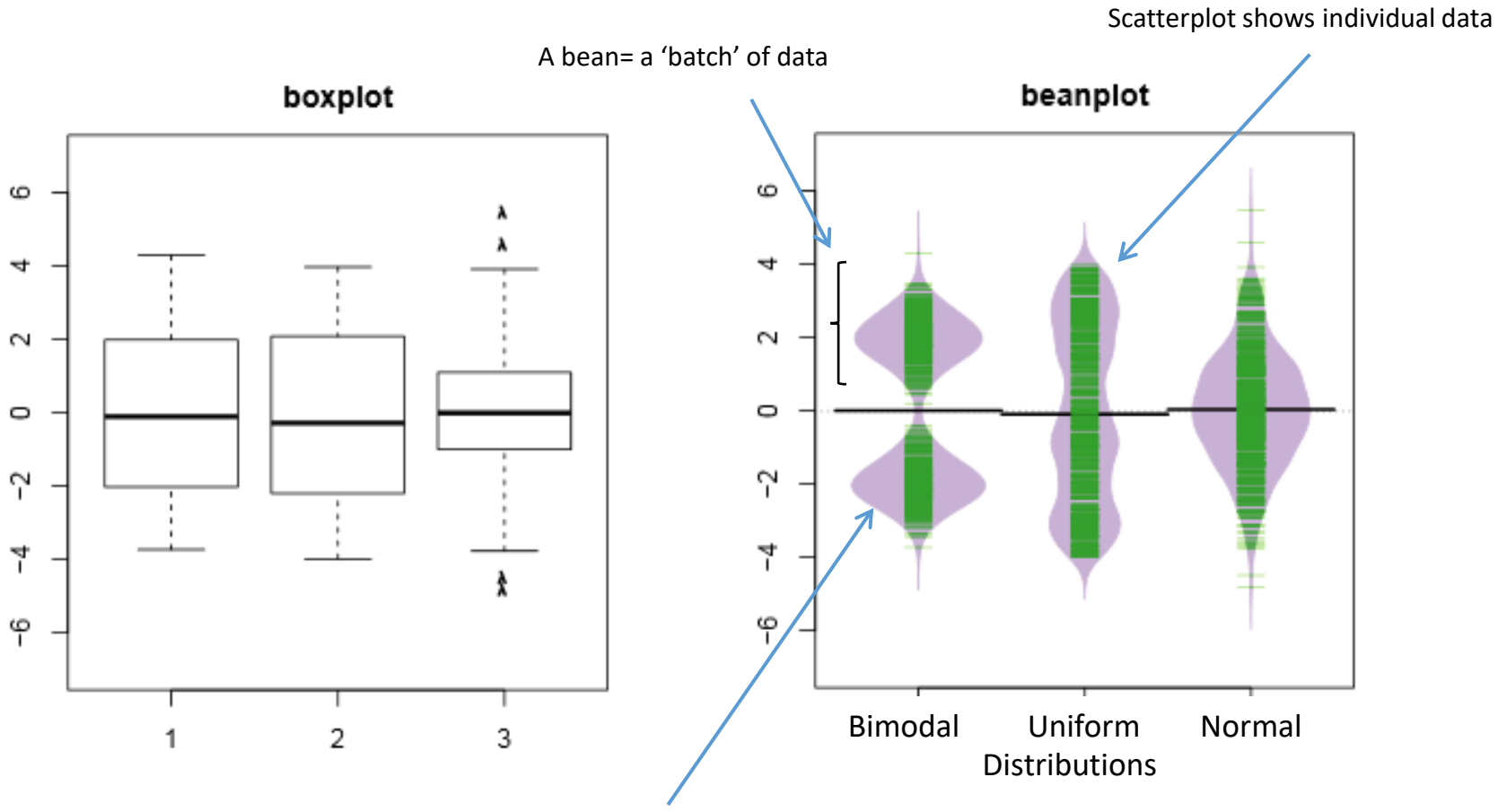
# Quantitative data: Boxplot



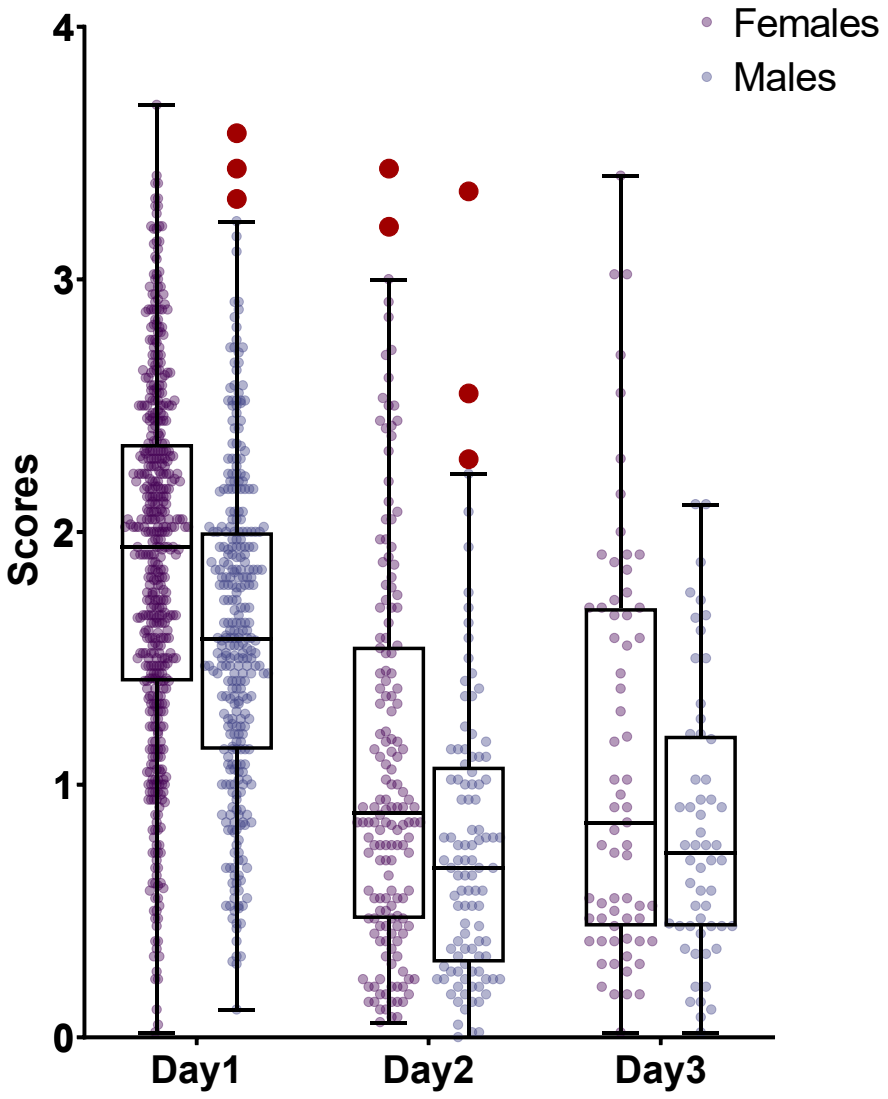

https://www.researchgate.net/publication/328818609_Outcomes_and_features_of_the_inspection_of_receiver_tubes_ITR_system_for_improved_OM_in_parabolic_trough_plants

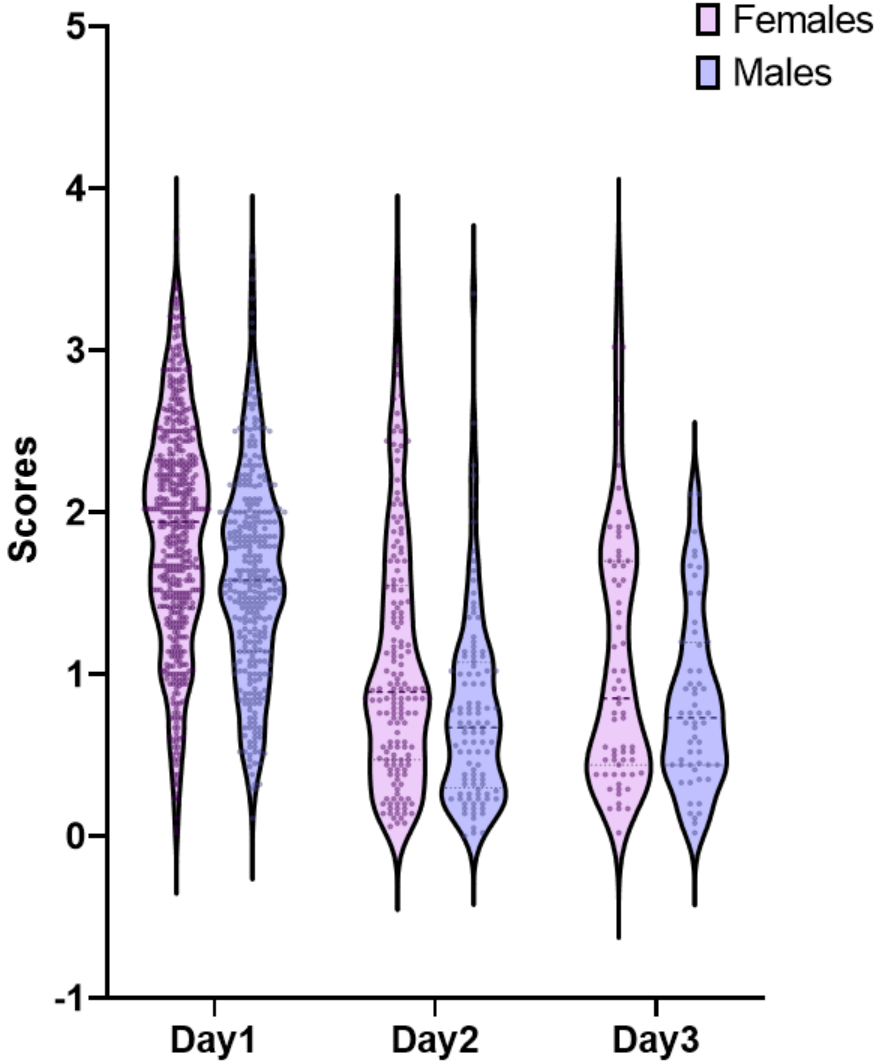
# Quantitative data: Boxplot or Beanplot (aka Violinplot)



# Quantitative data:

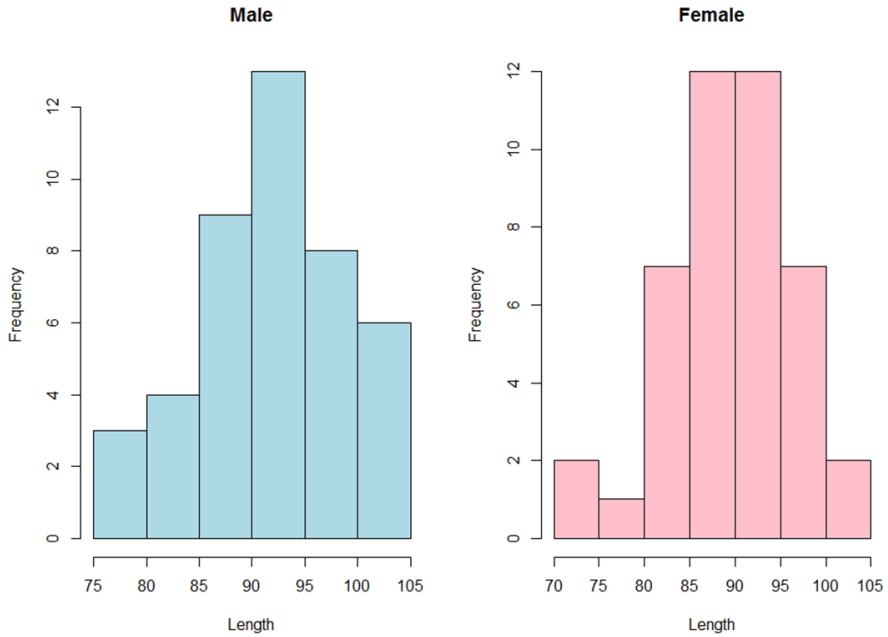
## Boxplot and Violinplot and Scatterplot



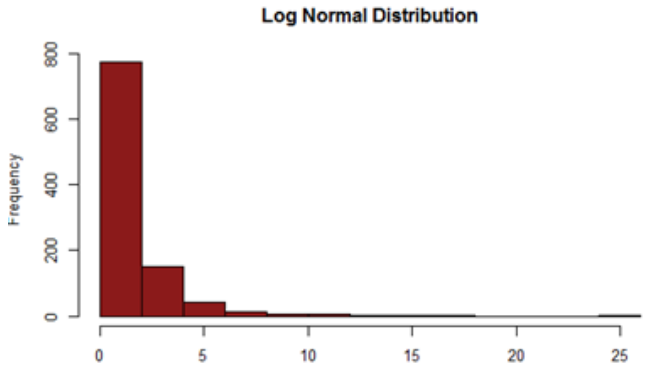
# Quantitative data: Histogram

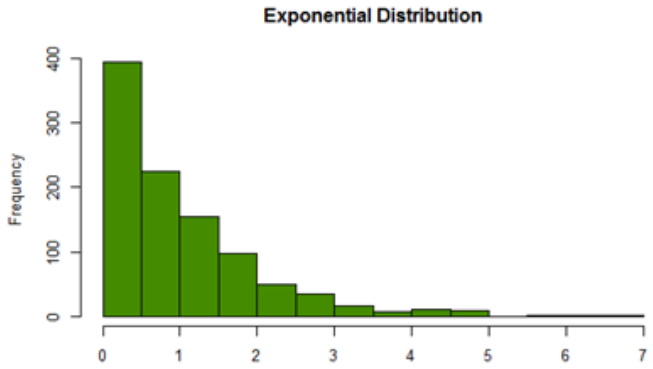
Big sample

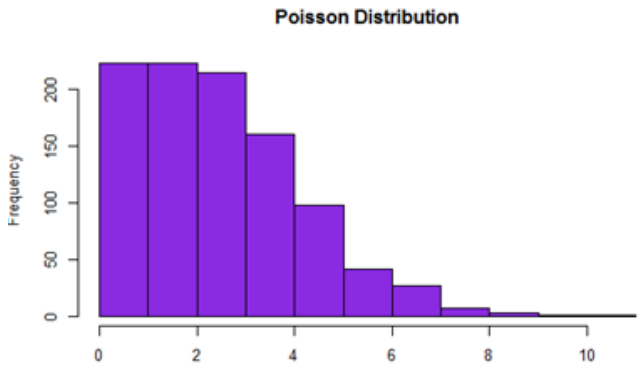


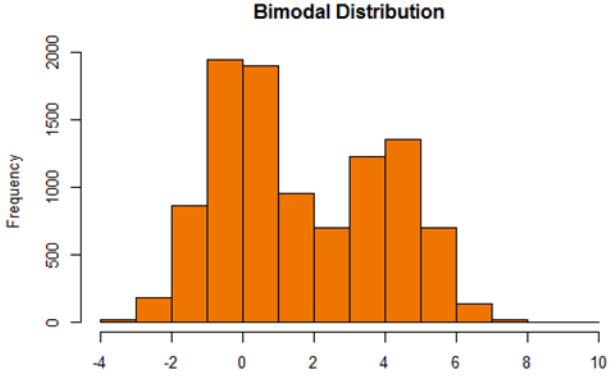

Small sample

# Quantitative data: Histogram (distribution)



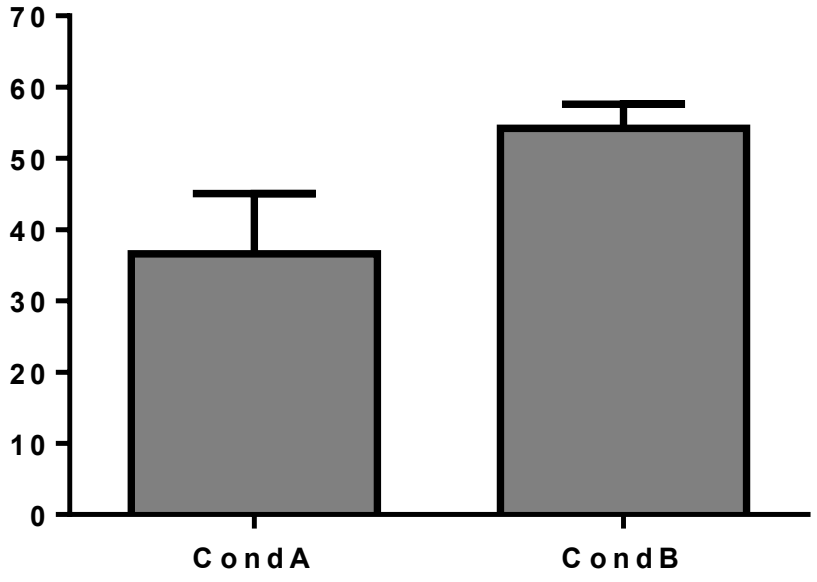



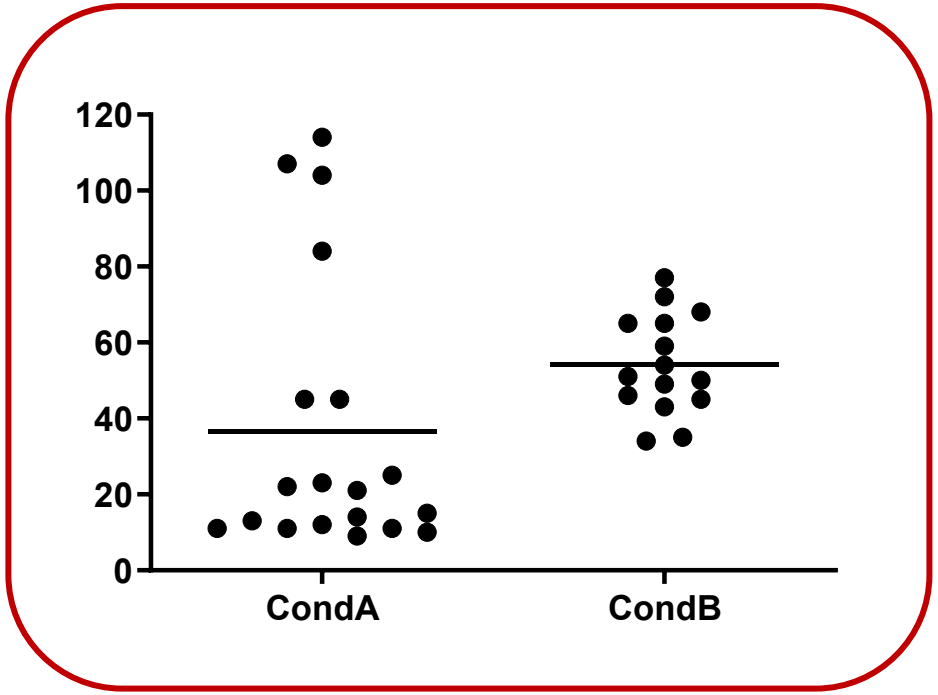
# Data exploration ≠ plotting data

https://github.com/allisonhorst/stats-illustrations#other-stats-artwork

# Plotting is not inherently the same thing as exploring

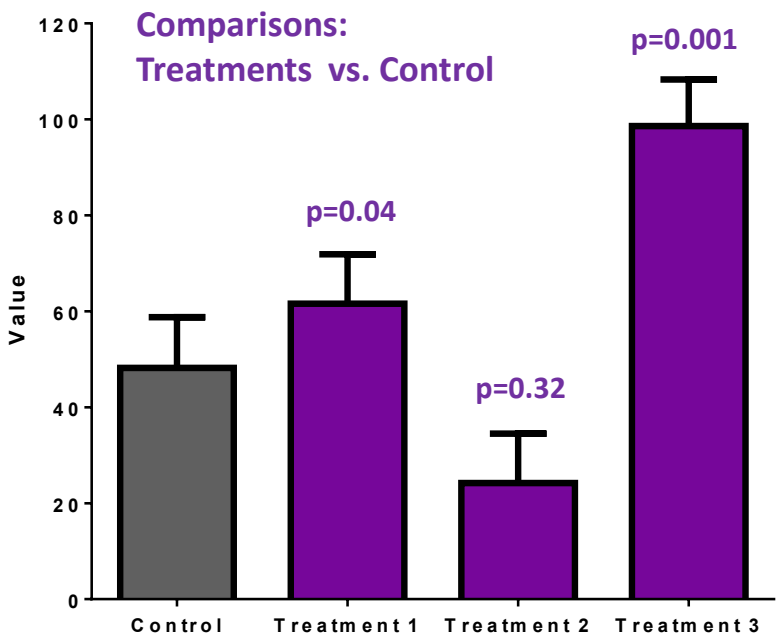
- <u>One experiment</u>: change in the variable of interest between CondA to CondB.
  - ❖ Data plotted as a **bar chart**.

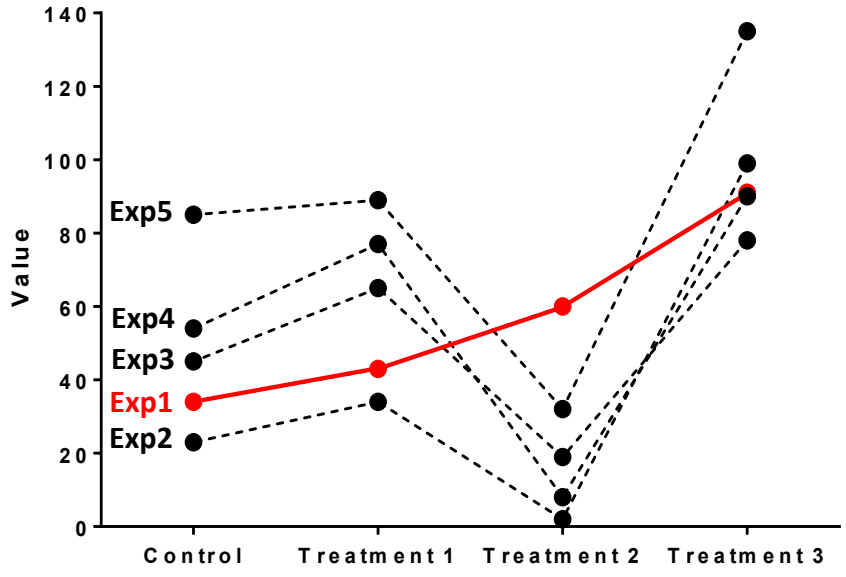


# Plotting is not inherently the same thing as exploring

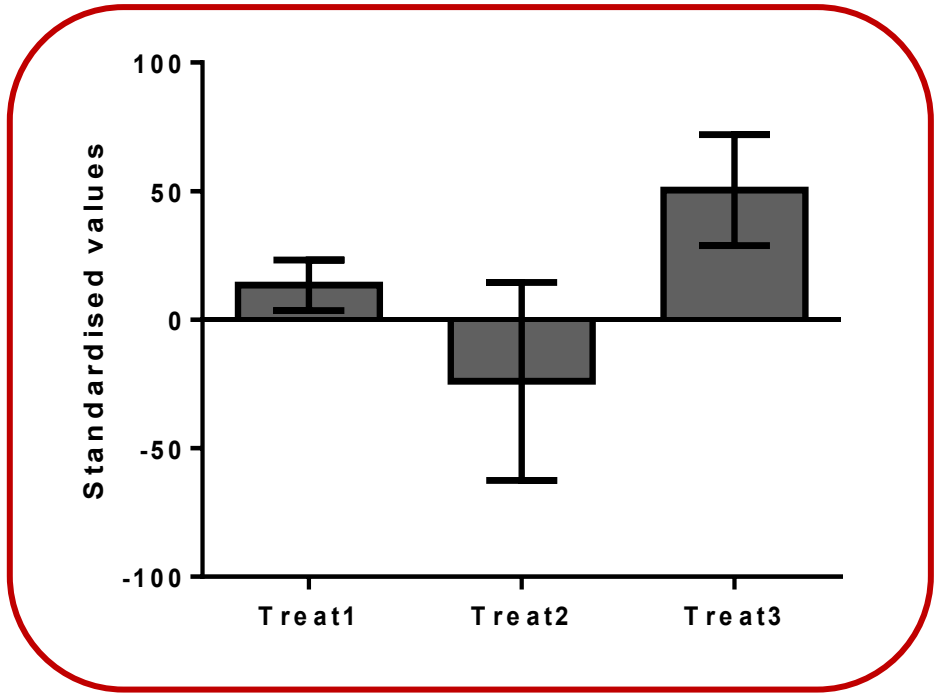
- Five experiments: change in the variable of interest between 3 treatments and a control.
  - ❖ Data plotted as a **bar chart**.

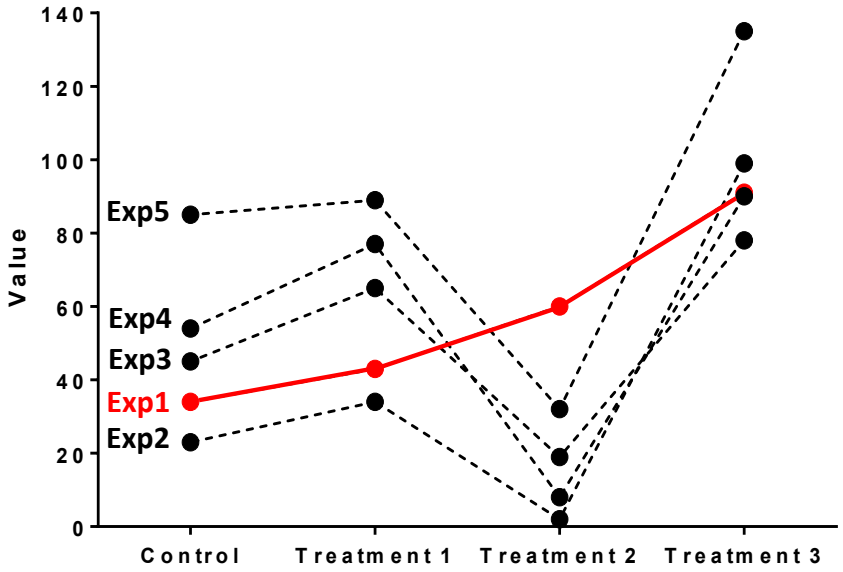


# Plotting is not inherently the same thing as exploring

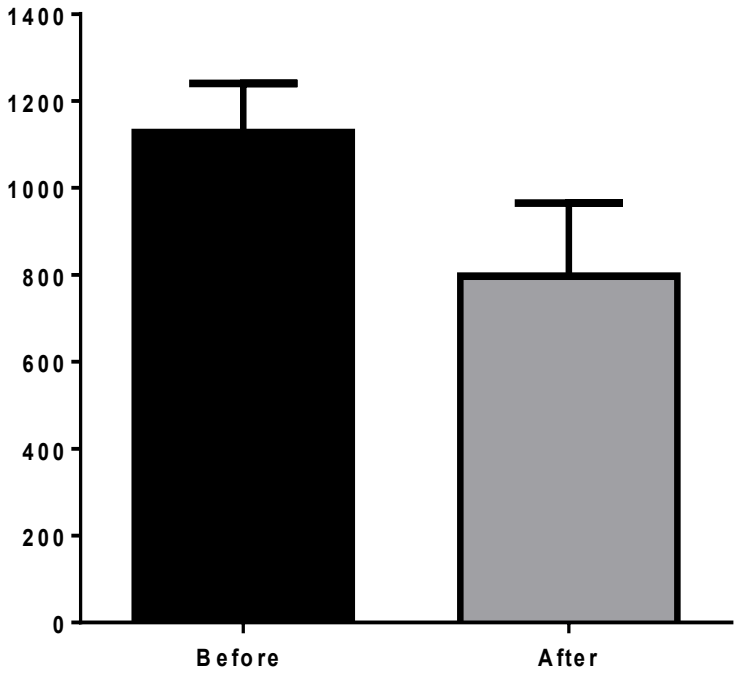
- Five experiments: change in the variable of interest between 3 treatments and a control.
  - ❖ Data plotted as a **bar chart**.

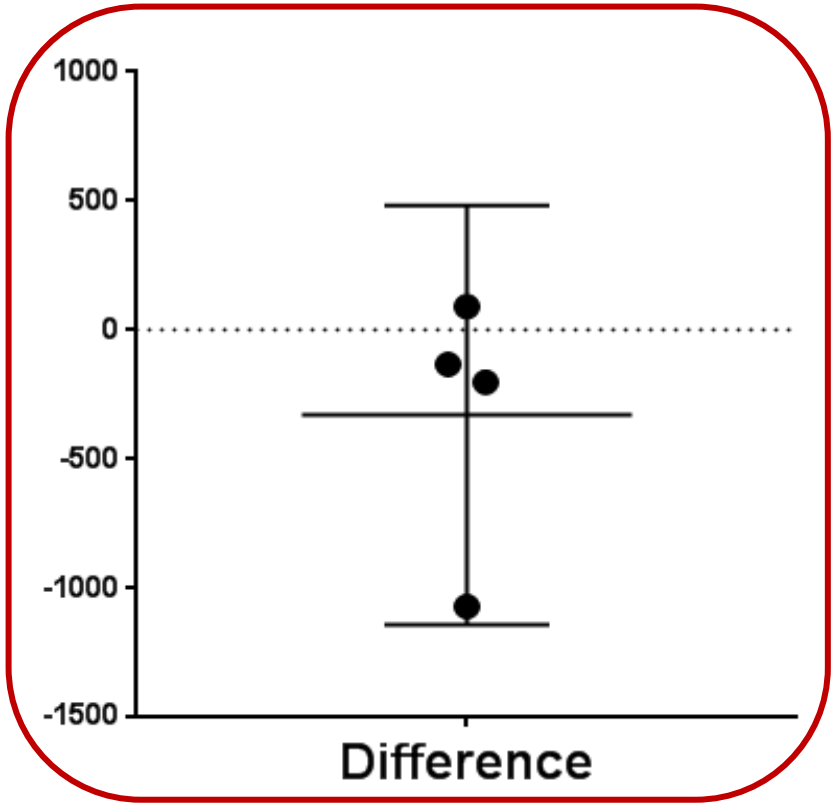


# Plotting is not inherently the same thing as exploring

- Four experiments: Before-After treatment effect on a variable of interest.
- Hypothesis: Applying a treatment will decrease the levels of the variable of interest.
  - ❖ Data plotted as a **bar chart**.



# Data exploration ≠ plotting data

https://github.com/allisonhorst/stats-illustrations#other-stats-artwork

Babraham Institute

# Key concepts and Assumptions

## Analysis of Quantitative data
## Choice of a statistical test

Hayley Carr & Anne Segonds-Pichon

v2024-05

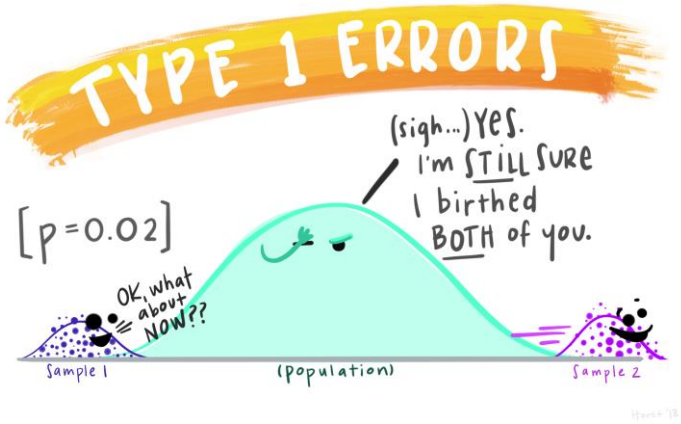
Babraham Bioinformatics

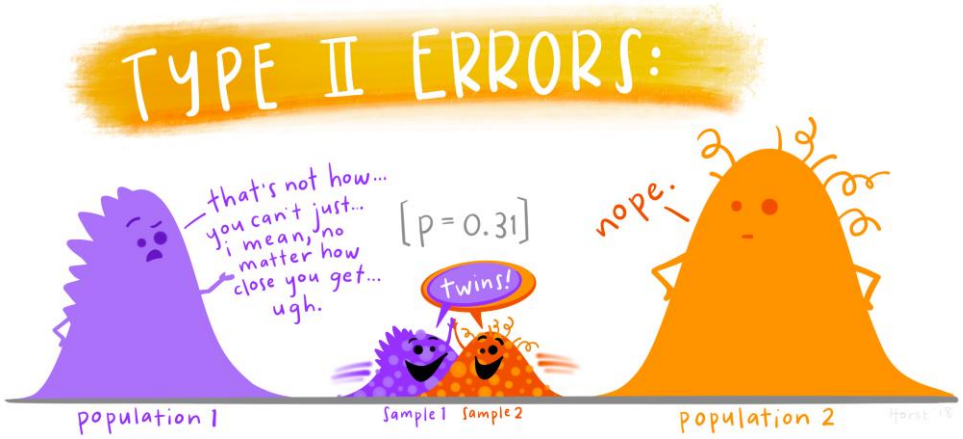
# The null hypothesis and error types

- The null hypothesis ($H_0$): $H_0$ = no effect
- The aim of a statistical test is to reject or not $H_0$.

- **High specificity** = low **False Positives** = low **Type I error**
- **High sensitivity** = low **False Negatives** = low **Type II error**

| Statistical decision | True state of $H_0$ | |
|---|---|---|
| | **$H_0$ true (no effect)** | **$H_0$ false (effect)** |
| **Reject $H_0$** | Type I error α<br>False positive | Correct<br>True positive |
| **Do not reject $H_0$** | Correct<br>True negative | Type II error β<br>False negative |




https://github.com/allisonhorst/stats-illustrations#other-stats-artwork

Sample → **Statistical inference** → Population

## Difference

| Is it meaningful? | Is it real? |
|---|---|

Determined scientifically

## Statistical test

| Difference | Variation | Sample size |
|---|---|---|

## Result

| Statistic (t, F, …) | Is difference big enough? |
|---|---|

Significance

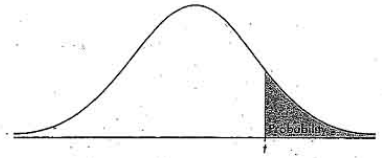



**TABLE B: t-DISTRIBUTION CRITICAL VALUES**

| | Tail probability p | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| df | .25 | .20 | .15 | .10 | .05 | .025 | .02 | .01 | .005 | .0025 | .001 | .0005 |
| 1 | 1.000 | 1.376 | 1.963 | 3.078 | 6.314 | 12.71 | 15.89 | 31.82 | 63.66 | 127.3 | 318.3 | 636.6 |
| 2 | .816 | 1.061 | 1.386 | 1.886 | 2.920 | 4.303 | 4.849 | 6.965 | 9.925 | 14.09 | 22.33 | 31.60 |
| 3 | .765 | .978 | 1.250 | 1.638 | 2.353 | 3.182 | 3.482 | 4.541 | 5.841 | 7.453 | 10.21 | 12.92 |
| 4 | .741 | .941 | 1.190 | 1.533 | 2.132 | 2.776 | 2.999 | 3.747 | 4.604 | 5.598 | 7.173 | 8.610 |
| 5 | .727 | .920 | 1.156 | 1.476 | 2.015 | 2.571 | 2.757 | 3.365 | 4.032 | 4.773 | 5.893 | 6.869 |
| 6 | .718 | .906 | 1.134 | 1.440 | 1.943 | 2.447 | 2.612 | 3.143 | 3.707 | 4.317 | 5.208 | 5.959 |
| 7 | .711 | .896 | 1.119 | 1.415 | 1.895 | 2.365 | 2.517 | 2.998 | 3.499 | 4.029 | 4.785 | 5.408 |
| 8 | .706 | .889 | 1.108 | 1.397 | 1.860 | 2.306 | 2.449 | 2.896 | 3.355 | 3.833 | 4.501 | 5.041 |
| 9 | .703 | .883 | 1.100 | 1.383 | 1.833 | 2.262 | 2.398 | 2.821 | 3.250 | 3.690 | 4.297 | 4.781 |
| 10 | .700 | .879 | 1.093 | 1.372 | 1.812 | 2.228 | 2.359 | 2.764 | 3.169 | 3.581 | 4.144 | 4.587 |
| 11 | .697 | .876 | 1.088 | 1.363 | 1.796 | 2.201 | 2.328 | 2.718 | 3.106 | 3.497 | 4.025 | 4.437 |
| 12 | .695 | .873 | 1.083 | 1.356 | 1.782 | 2.179 | 2.303 | 2.681 | 3.055 | 3.428 | 3.930 | 4.318 |
| 13 | .694 | .870 | 1.079 | 1.350 | 1.771 | 2.160 | 2.282 | 2.650 | 3.012 | 3.372 | 3.852 | 4.221 |
| 14 | .692 | .868 | 1.076 | 1.345 | 1.761 | 2.145 | 2.264 | 2.624 | 2.977 | 3.326 | 3.787 | 4.140 |
| 15 | .691 | .866 | 1.074 | 1.341 | 1.753 | 2.131 | 2.249 | 2.602 | 2.947 | 3.286 | 3.733 | 4.073 |
| 16 | .690 | .865 | 1.071 | 1.337 | 1.746 | 2.120 | 2.235 | 2.583 | 2.921 | 3.252 | 3.686 | 4.015 |
| 17 | .689 | .863 | 1.069 | 1.333 | 1.740 | 2.110 | 2.224 | 2.567 | 2.898 | 3.222 | 3.646 | 3.965 |
| 18 | .688 | .862 | 1.067 | 1.330 | 1.734 | 2.101 | 2.214 | 2.552 | 2.878 | 3.197 | 3.611 | 3.922 |
| 19 | .688 | .861 | 1.066 | 1.328 | 1.729 | 2.093 | 2.205 | 2.539 | 2.861 | 3.174 | 3.579 | 3.883 |
| 20 | .687 | .860 | 1.064 | 1.325 | 1.725 | 2.086 | 2.197 | 2.528 | 2.845 | 3.153 | 3.552 | 3.850 |
| 21 | .686 | .859 | 1.063 | 1.323 | 1.721 | 2.080 | 2.189 | 2.518 | 2.831 | 3.135 | 3.527 | 3.819 |
| 22 | .686 | .858 | 1.061 | 1.321 | 1.717 | 2.074 | 2.183 | 2.508 | 2.819 | 3.119 | 3.505 | 3.792 |

# Signal-to-noise ratio

- Stats are all about understanding and controlling variation
- The ratio of signal to noise determines the significance

| Statistical test | | |
|---|---|---|
| Difference | Variation | Sample size |

$$\frac{\text{Difference (signal)}}{\text{Variation (noise)}}$$

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

If the **noise** (interindividual variation) **is low** then the **signal is detectable**

→ statistical significance

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

If the **noise is large** the **same signal will not be detected**

→ no statistical significance

# Choice of a statistical test

There are many statistical tests. Which one we use depends on:

- What we want to do
  - The questions asked
  - Correct **statistical test** to answer our questions
- What sort of data we have
  - The type and behaviour
  - Correct **statistical family**
    - There are 2 families of statistical tests:
      - Parametric tests with 4 assumptions to be met
      - Non-parametric tests with no or few assumptions and/or for qualitative data

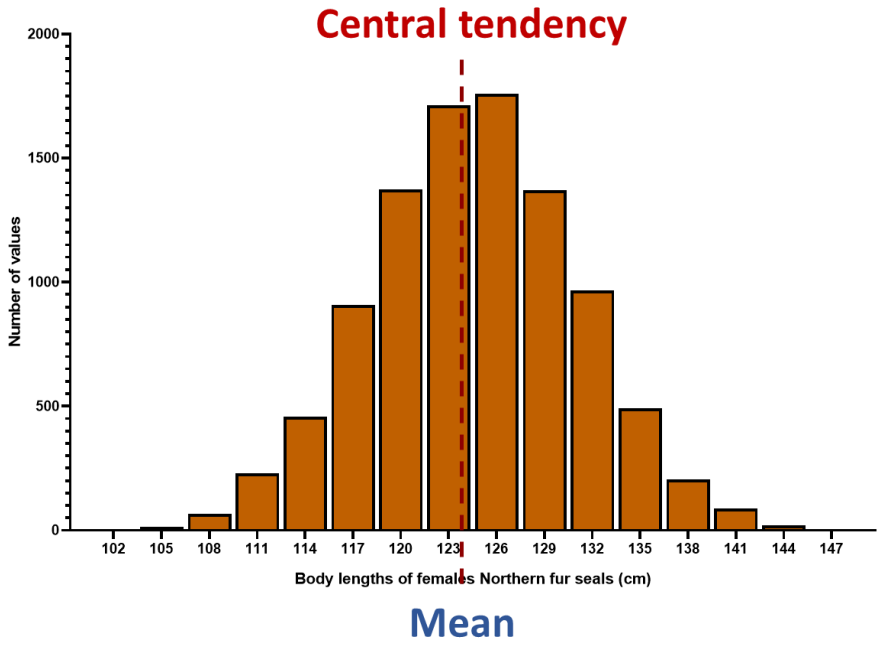
# Assumptions of Parametric Data

## What sort of data we have

- All parametric tests have 4 basic assumptions that must be met for the test to be accurate.

### First assumption: Normality

- Normal shape, bell shape, Gaussian shape

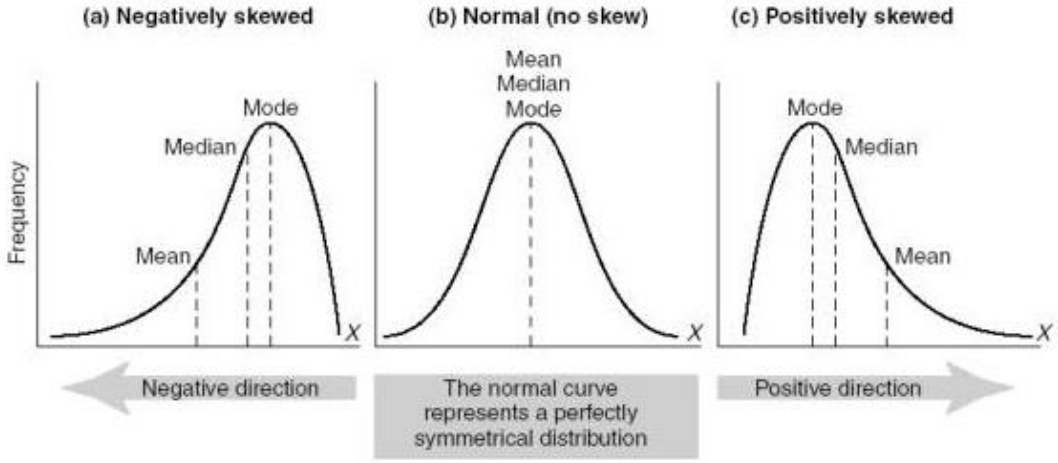


# Assumptions of Parametric Data

## What sort of data we have

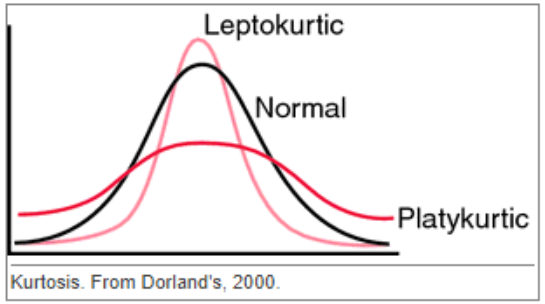
- Frequent departures from normality:
  - Skewness: lack of symmetry of a distribution

Skewness < 0 Skewness = 0 Skewness > 0




  - Kurtosis: measure of the degree of 'peakedness'
    - Same variance and same skew but differ markedly in kurtosis

More peaked distribution: kurtosis > 0




Kurtosis. From Dorland's, 2000.

Flatter distribution: kurtosis < 0

# Assumptions of Parametric Data

## What sort of data we have

**Second assumption: Homoscedasticity (Homogeneity in variance)**

- The variance should not change systematically throughout the data

**Third assumption: Interval data (linearity)**

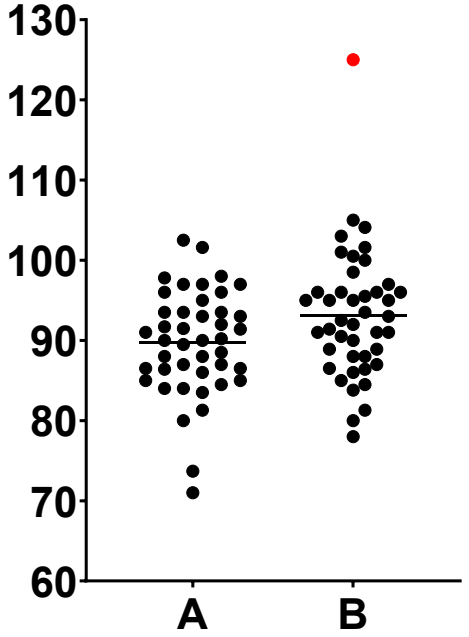
- The distance between points of the scale should be equal at all parts along the scale

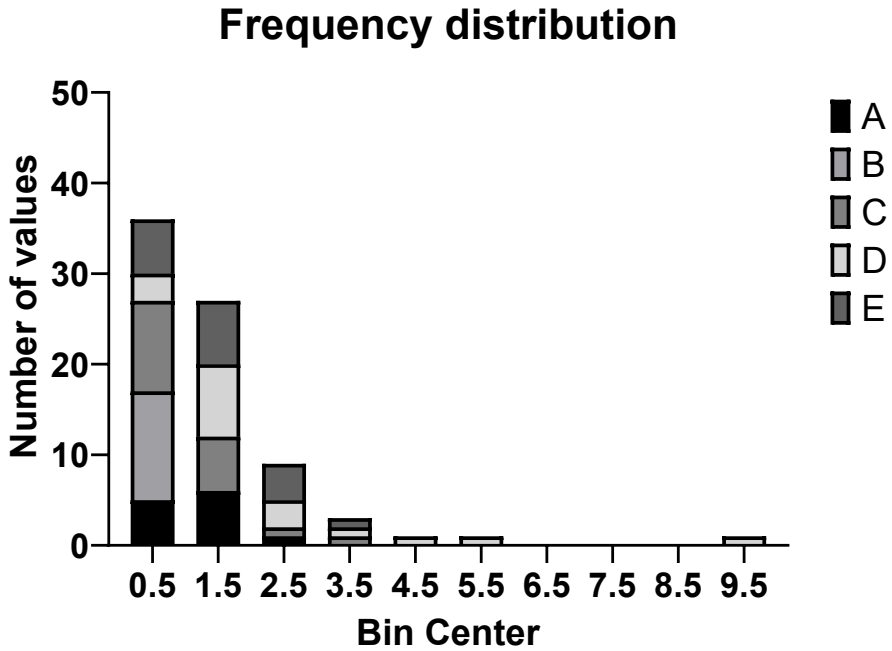
**Fourth assumption: Independence**

- Data from different subjects are independent
  - Each data point in the sample is independent from all the others = Values corresponding to one subject do not influence the values corresponding to another subject
  - Important in repeated measures experiments

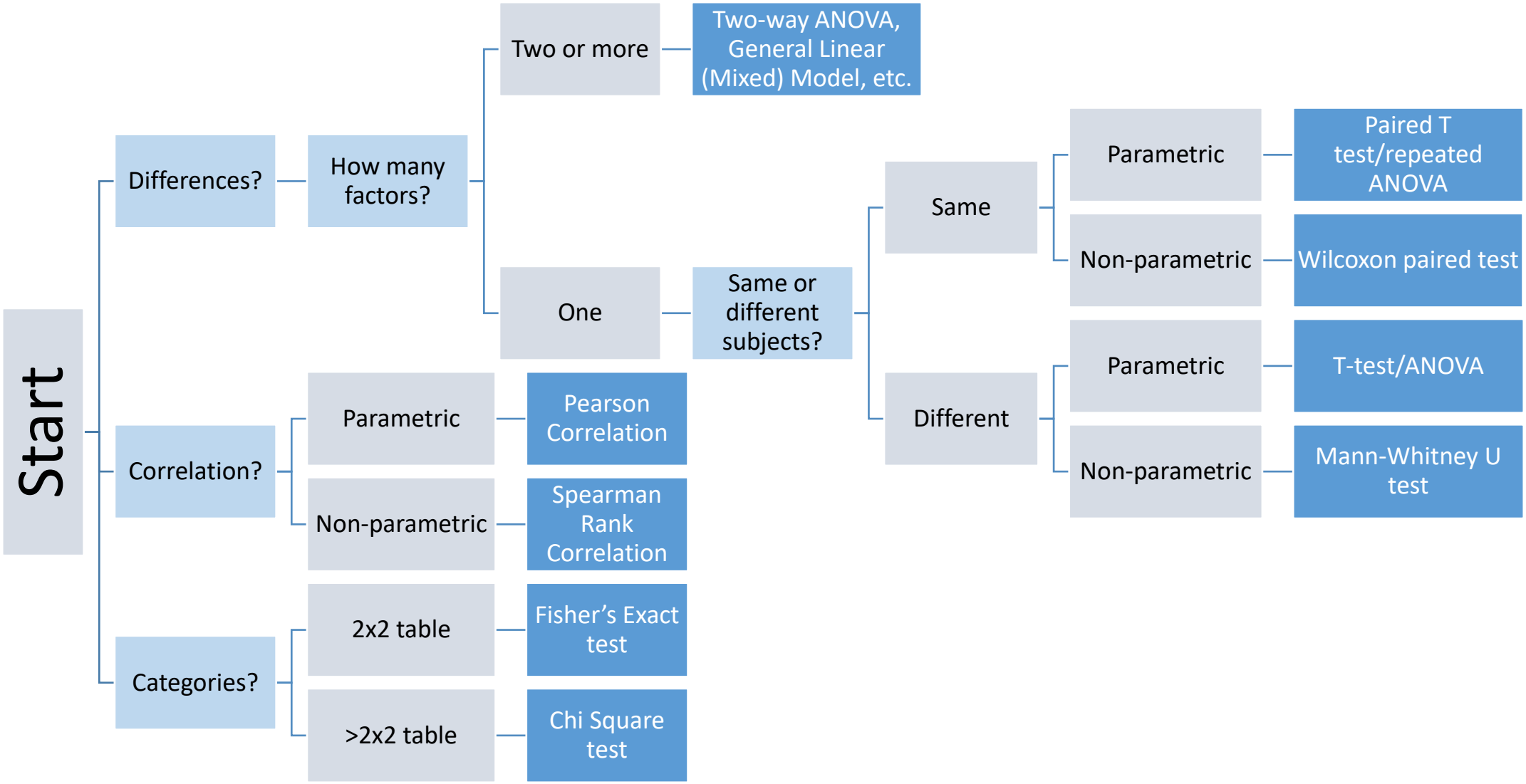
# Non-parametric tests

- General principle: original data are transformed into **ranks**
- **Not meeting the assumptions for parametric tests is not enough** to switch to a non-parametric approach
- Data exploration is key:
  - Outliers?
  - Possible transformation?
  - Parametric with corrections?
- If outcome is a rank or a score with limited possible values: often non-parametric approach



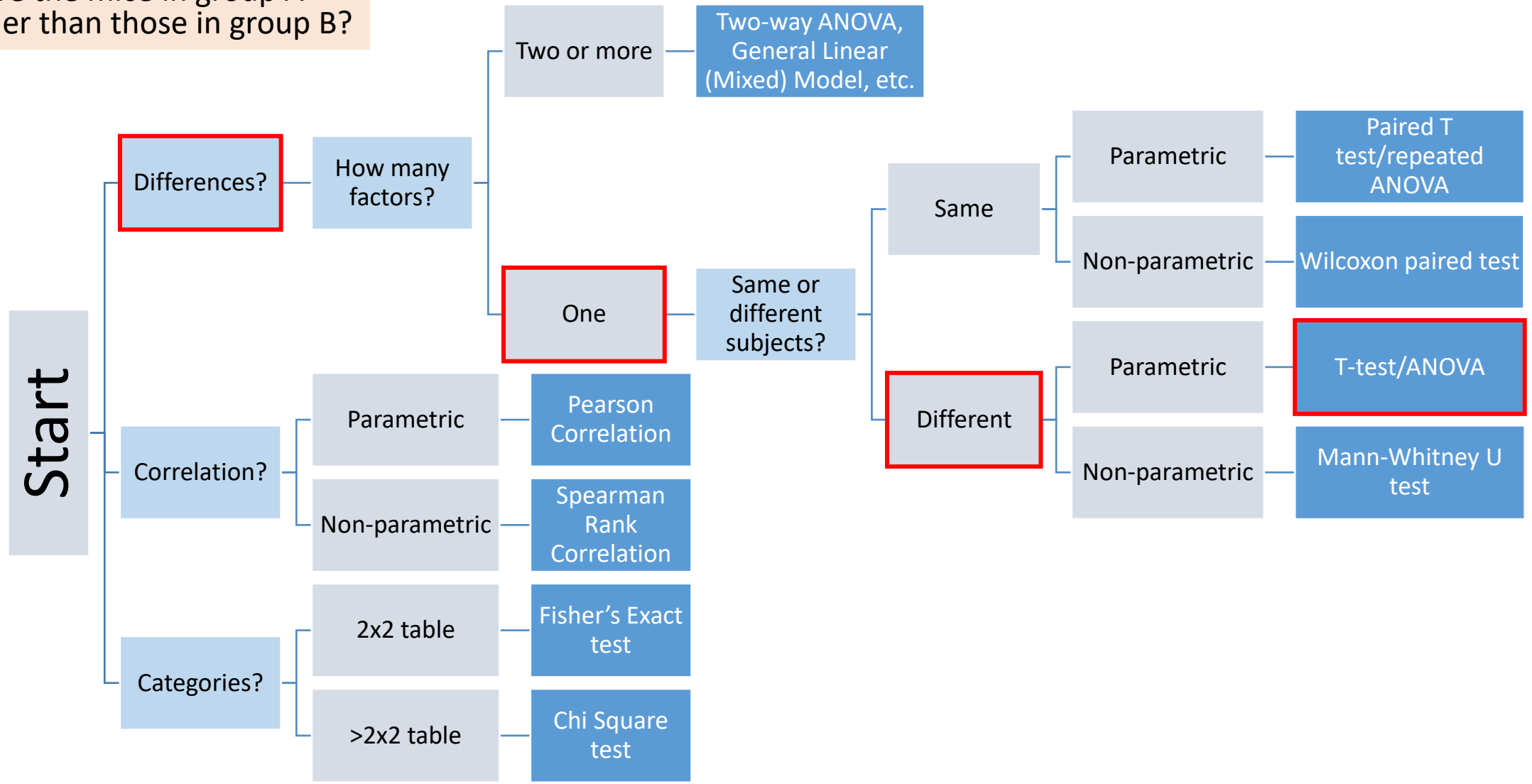



# Choice of a statistical test



# Choice of a statistical test

Are the mice in group A heavier than those in group B?



Babraham Institute

# Analysis of Quantitative data

## Student's *t*-test

Hayley Carr & Anne Segonds-Pichon

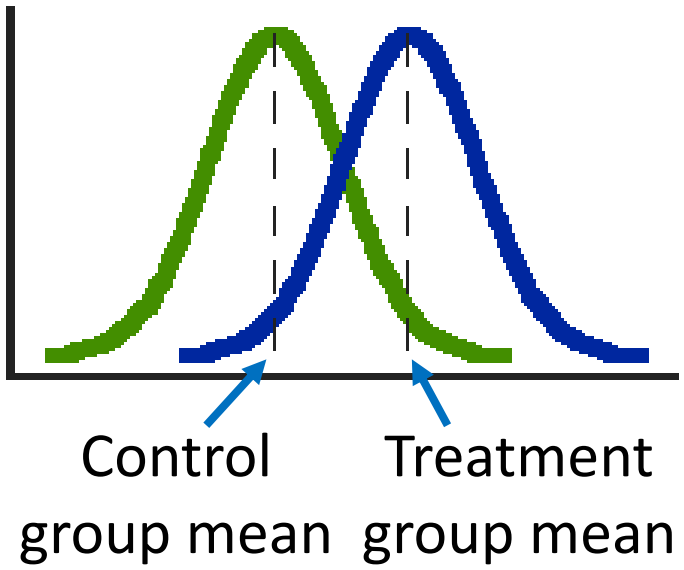
v2023-09

Babraham Bioinformatics

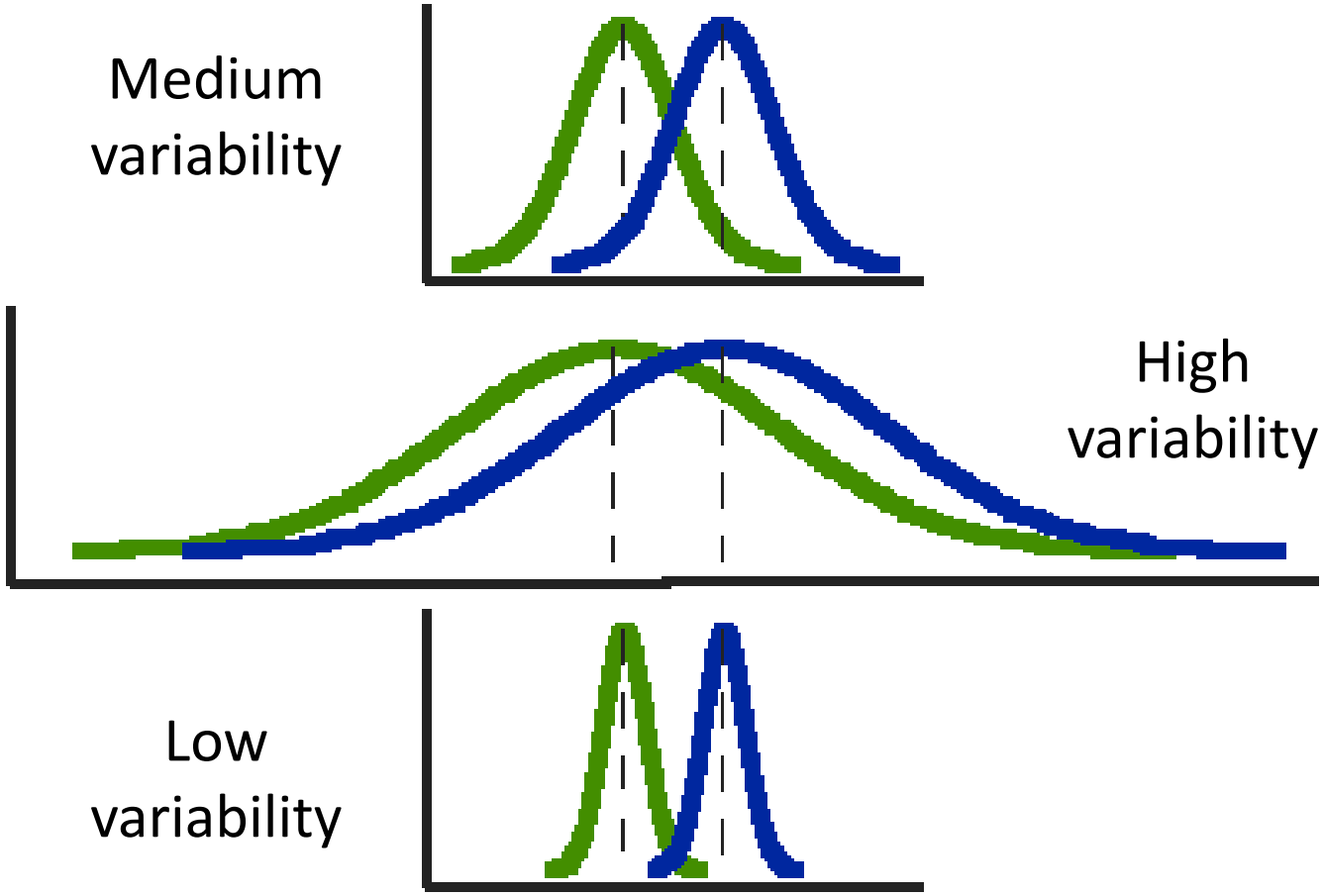
# Comparison between 2 groups

## Student's *t*-test

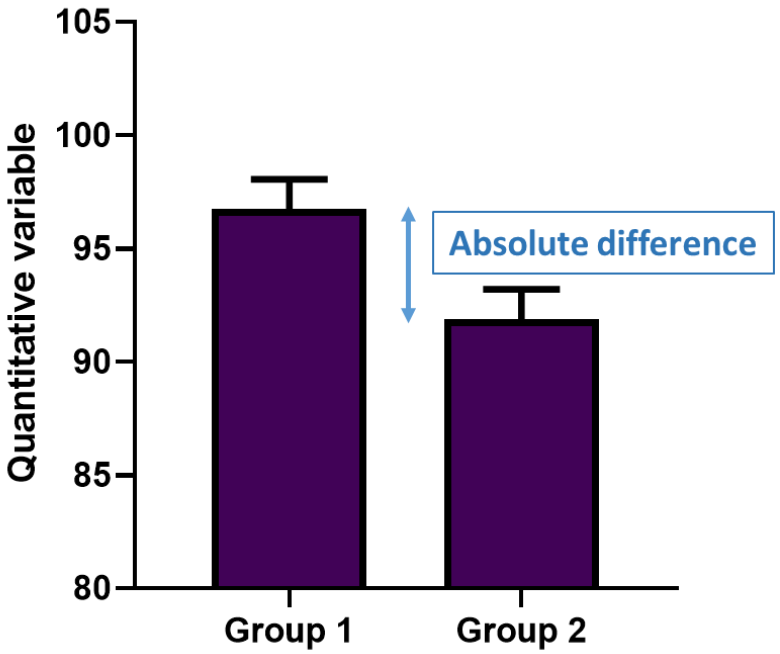
- **Basic idea**:
  - Comparison between 2 means accounting for variability
    - Absolute difference vs. variability



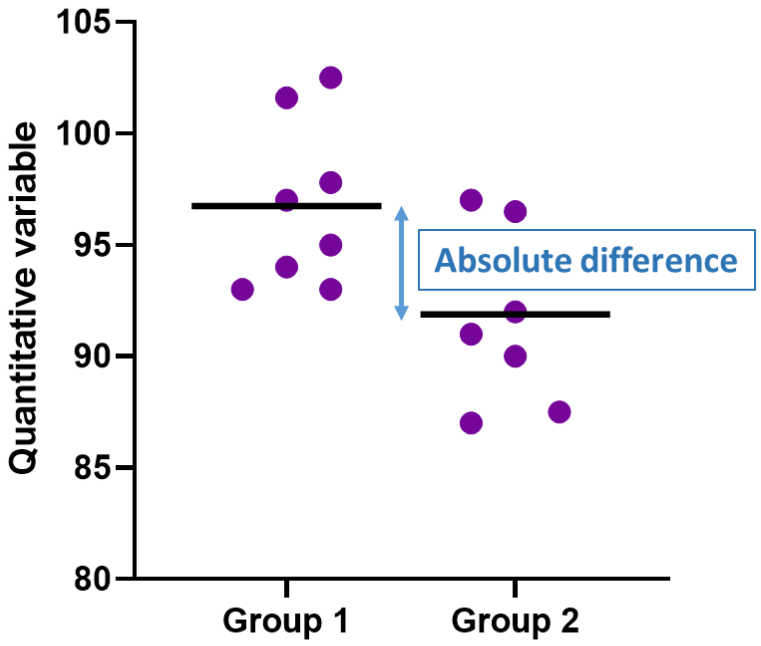
# Variability does matter



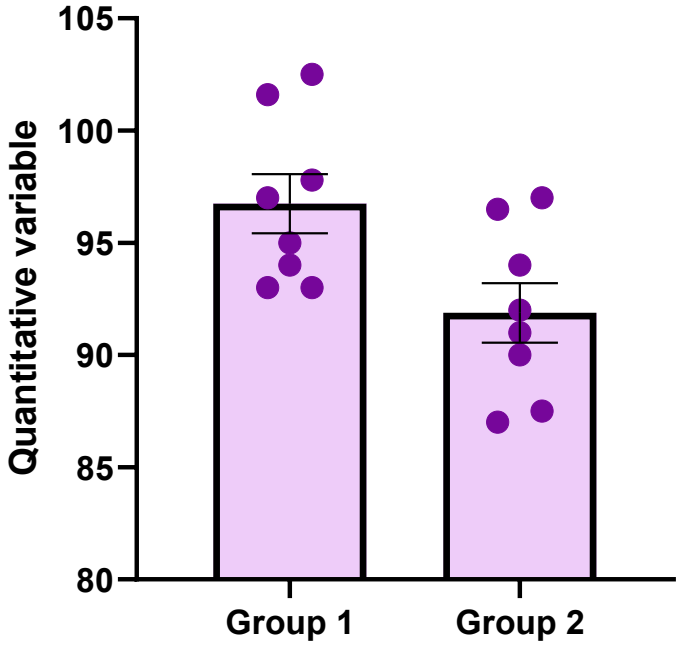
# Variability does matter




Bar chart ☹️

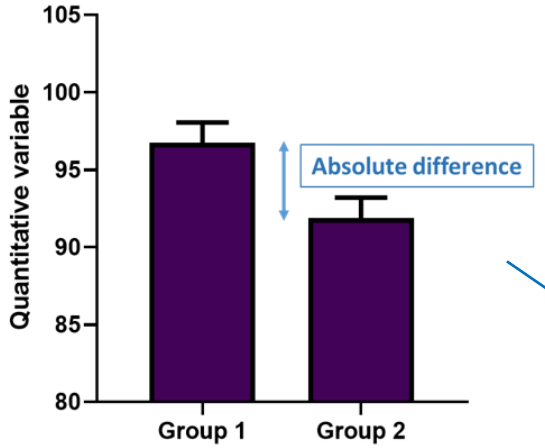


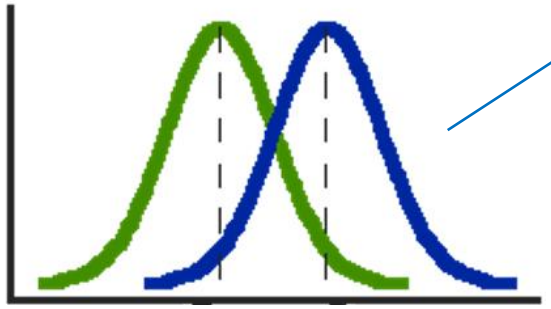

Scatter plot 🙂




🙂 🙂 🙂

# Signal-to-noise ratio



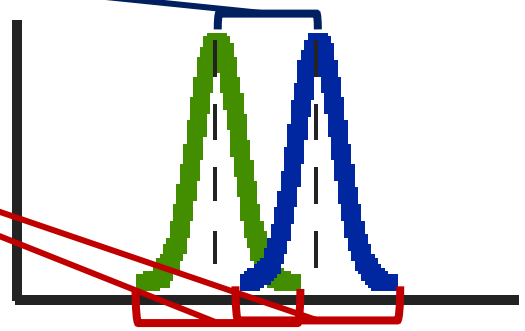
$$\frac{\text{Difference (signal)}}{\text{Variation (noise)}}$$

$$\frac{\text{Signal}}{\text{Noise}} = \textbf{statistical significance}$$

$$\frac{\text{Signal}}{\text{Noise}} = \textbf{no statistical significance}$$

# Student's *t*-test

$$\frac{\text{Signal}}{\text{Noise}} = \frac{\text{Difference between group means}}{\text{Variability of groups}}$$



$$= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

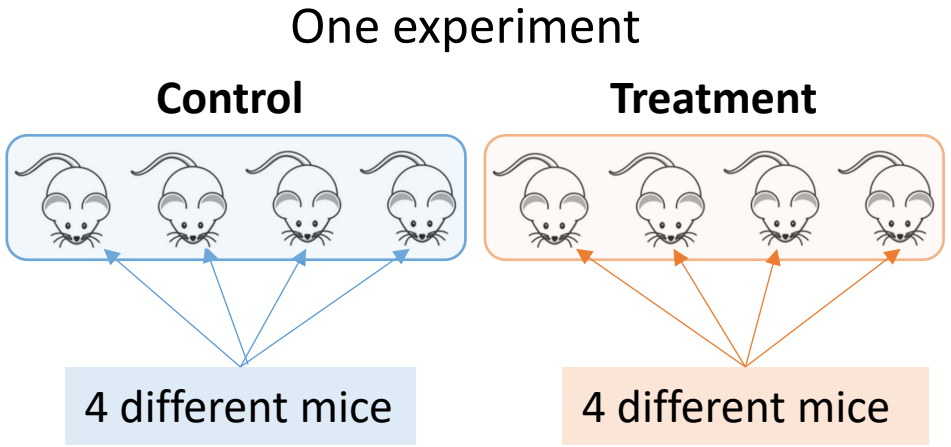
Basically the combination of the 2 SEM

$$= \textbf{t-value}$$

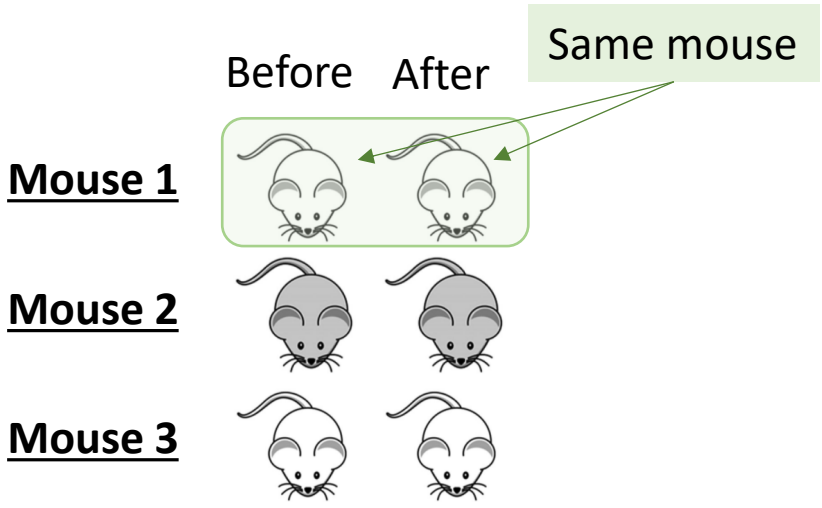
# Student's *t*-test

- 3 types, depending on experimental design

- **Independent *t*-test**
  - Difference between 2 means of one variable for two independent groups




- **Paired *t*-test**
  - Difference between two measures of one variable for one group



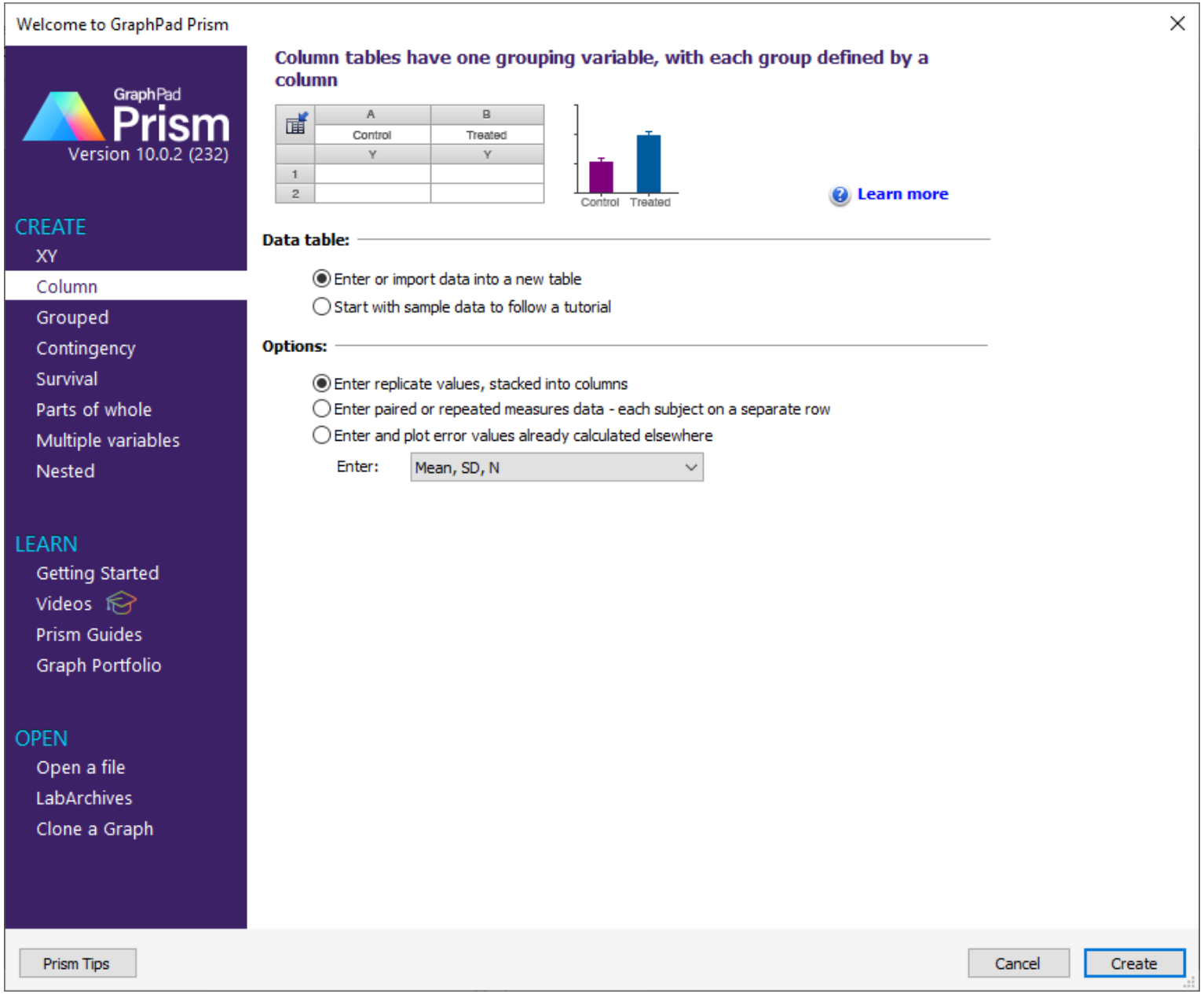

- **One-Sample *t*-test**
  - Difference between the mean of a single variable and a specified constant

# Example: coyotes.xlsx

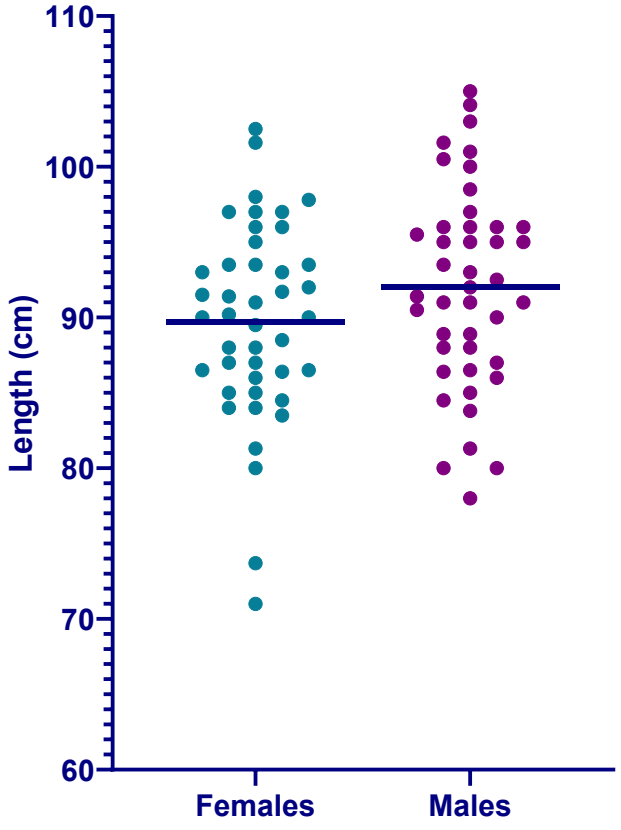
- Do male and female coyotes differ in size?
- The file contains individual body length of male and female coyotes.
- Steps:
  - Data exploration
  - Check the assumptions for parametric test
  - Statistical analysis: Independent *t*-test

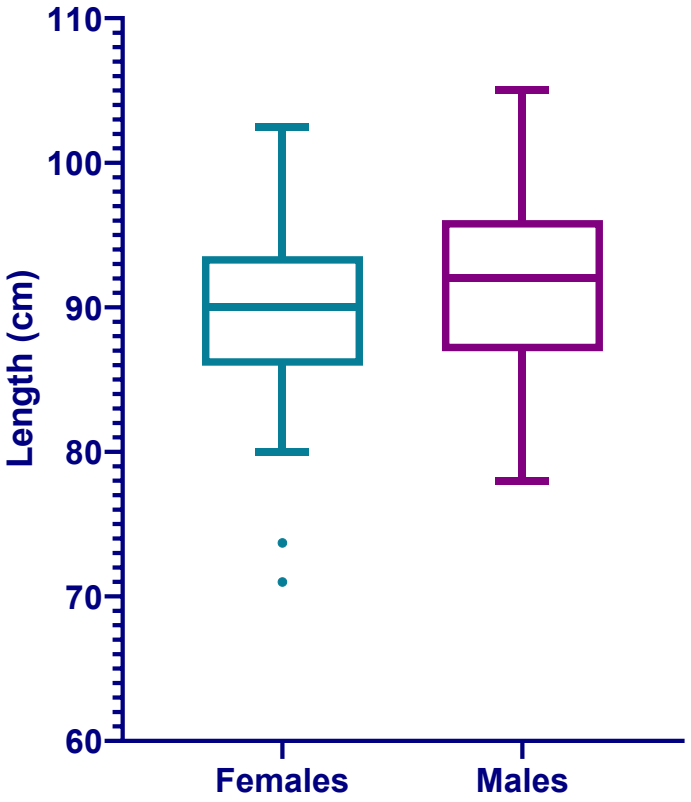
# Data exploration
## coyotes.xlsx

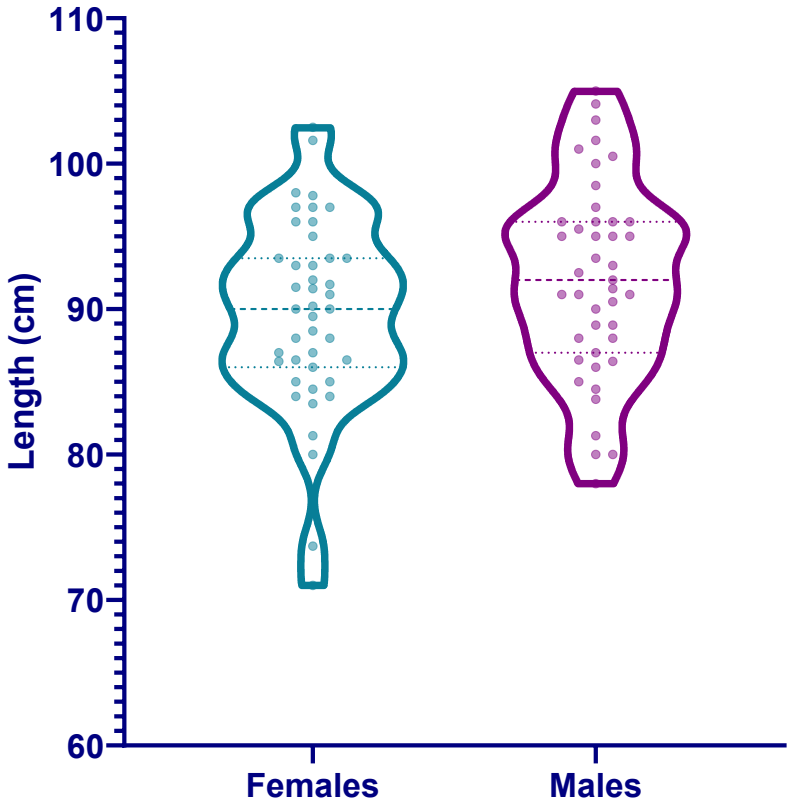
- Do male and female coyotes differ in size?
- Load in and explore the data



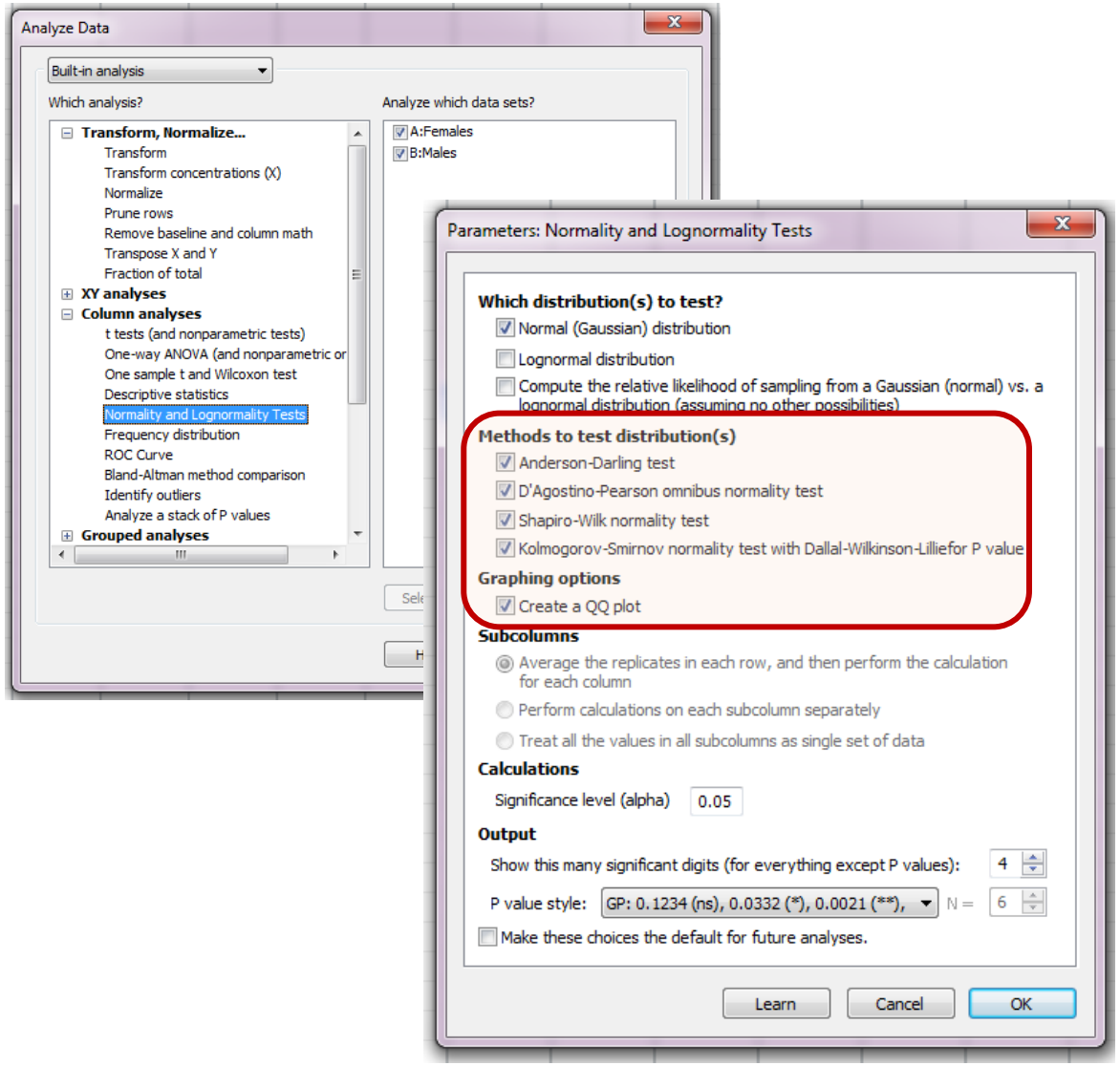
# Exercise: Exploring data - *Answers*



# Checking normality assumption with formal tests



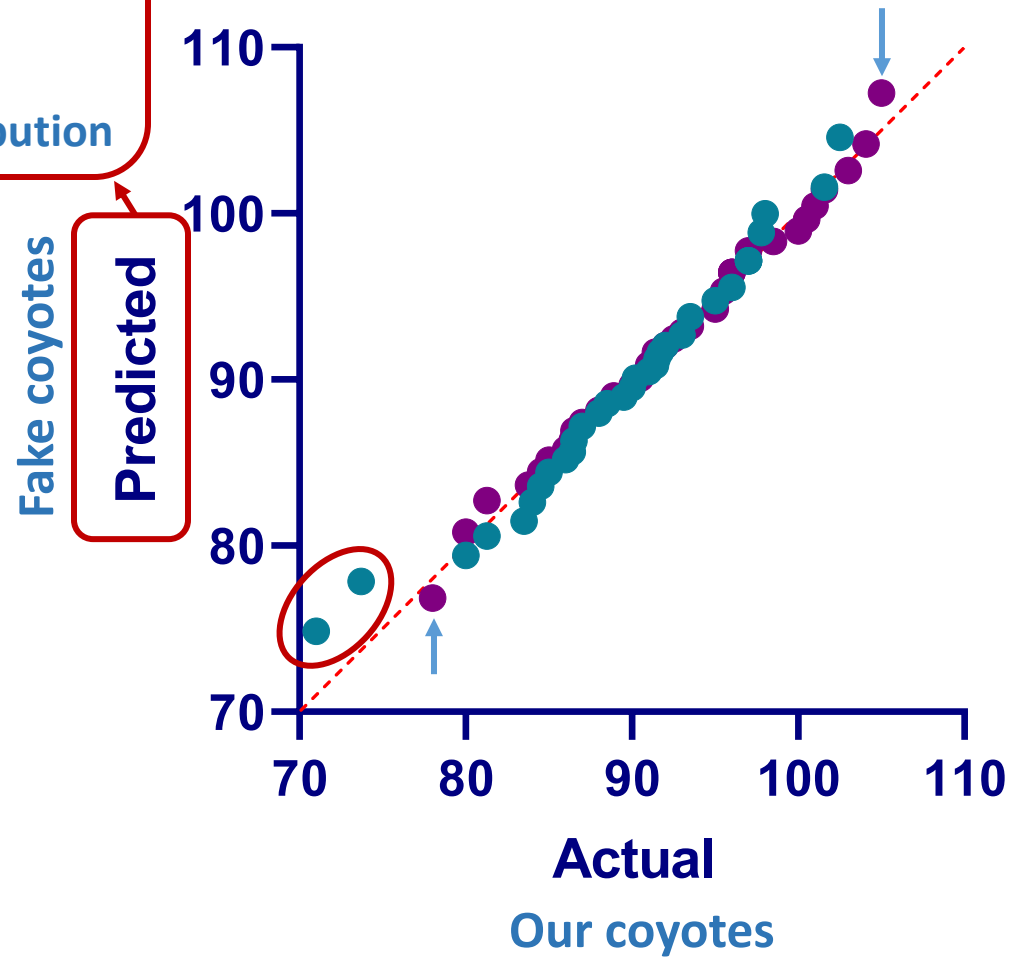

| | A | B |
|---|---|---|
| | Females | Males |
| 1 | Test for normal distribution | |
| 2 | Anderson-Darling test | |
| 3 | A2* | 0.3158 | 0.1750 |
| 4 | P value | 0.5294 | 0.9192 |
| 5 | Passed normality test (alpha=0.05)? | Yes | Yes |
| 6 | P value summary | ns | ns |
| 7 | | | |
| 8 | D'Agostino & Pearson test | | |
| 9 | K2 | 4.203 | 0.5080 |
| 10 | P value | 0.1223 | 0.7757 |
| 11 | Passed normality test (alpha=0.05)? | Yes | Yes |
| 12 | P value summary | ns | ns |
| 13 | | | |
| 14 | Shapiro-Wilk test | | |
| 15 | W | 0.9700 | 0.9845 |
| 16 | P value | 0.3164 | 0.8190 |
| 17 | Passed normality test (alpha=0.05)? | Yes | Yes |
| 18 | P value summary | ns | ns |
| 19 | | | |
| 20 | Kolmogorov-Smirnov test | | |
| 21 | KS distance | 0.07845 | 0.08853 |
| 22 | P value | >0.1000 | >0.1000 |
| 23 | Passed normality test (alpha=0.05)? | Yes | Yes |
| 24 | P value summary | ns | ns |
| 25 | | | |
| 26 | Number of values | 43 | 43 |

Normality ☑

# Checking normality assumption with a QQ plot

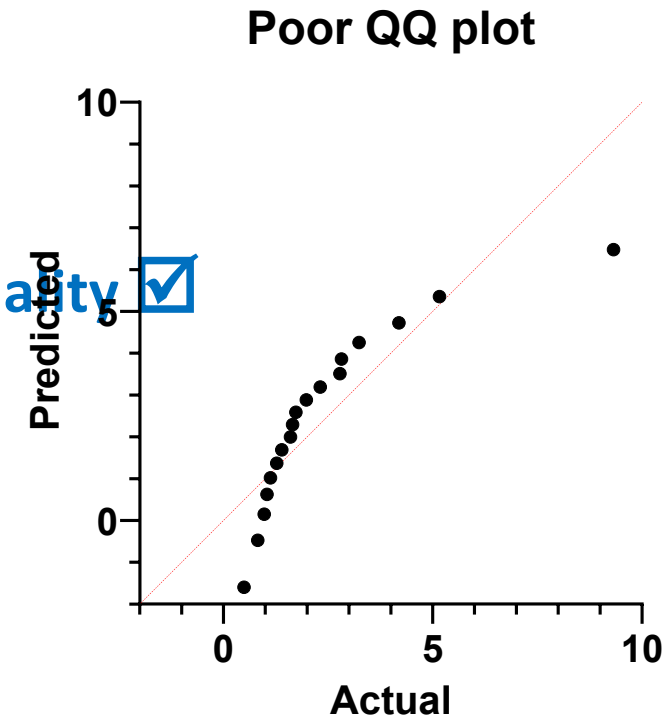
Same mean
Same SD
Same sample size
BUT
perfectly normal distribution

**Normal QQ plot = Quantile – Quantile plot**

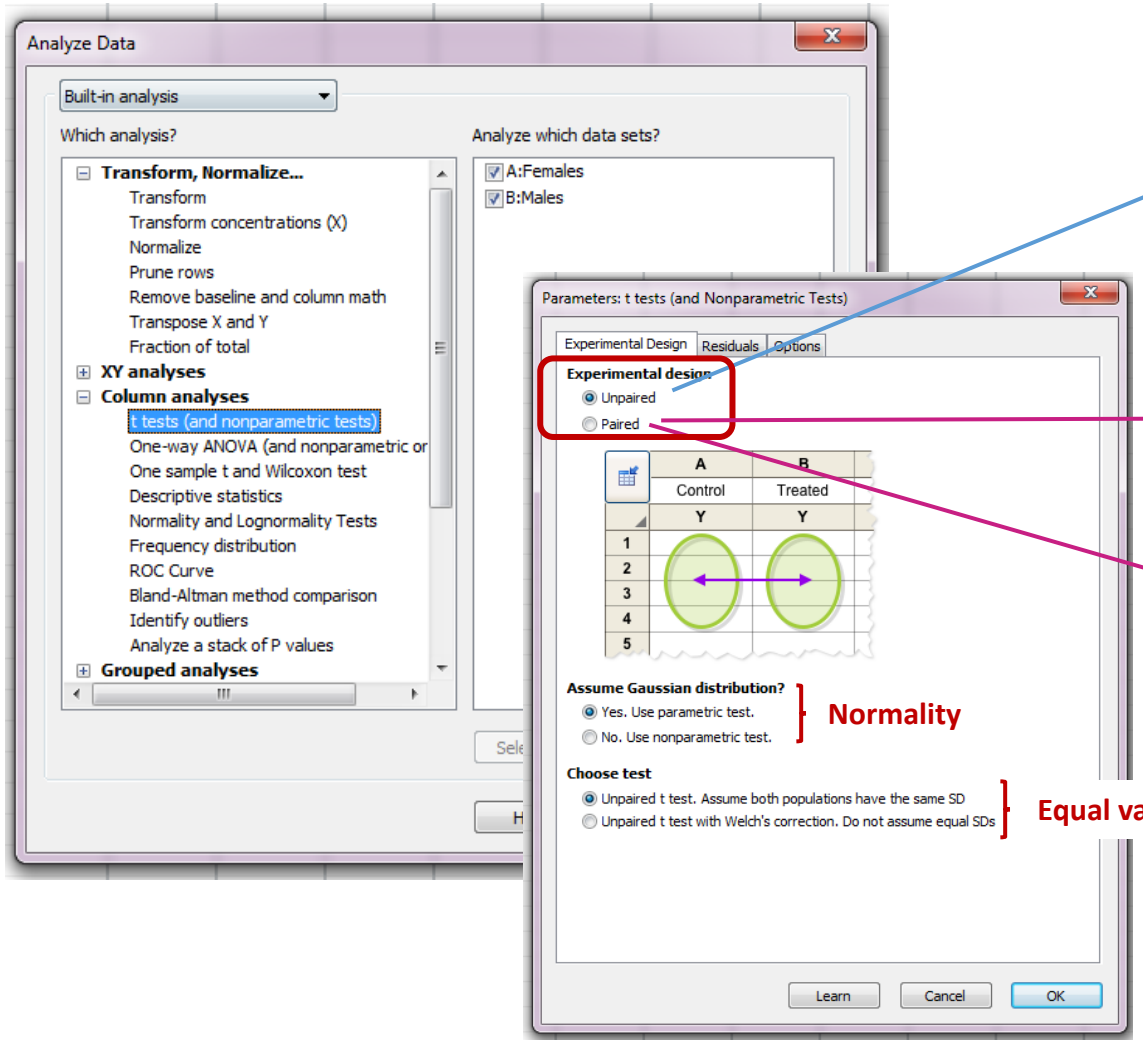


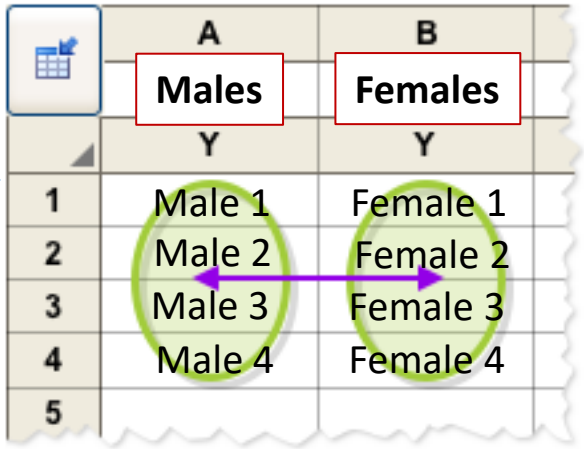

Females
Males

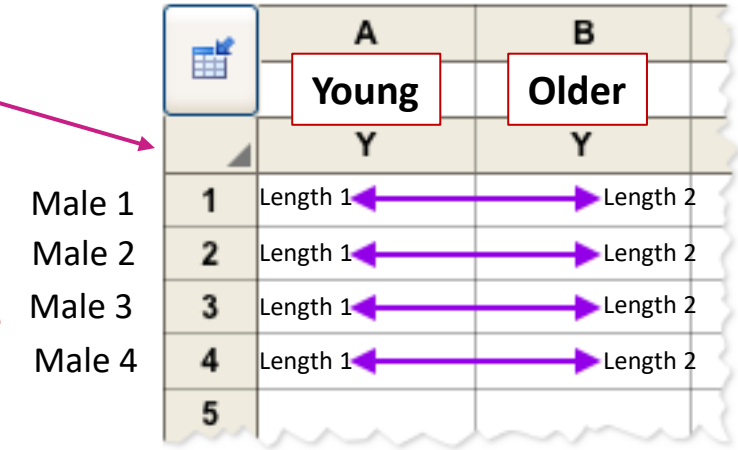
Normality ☑



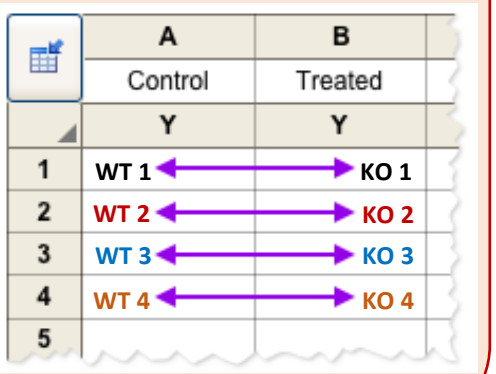
# Student's *t*-test with Prism



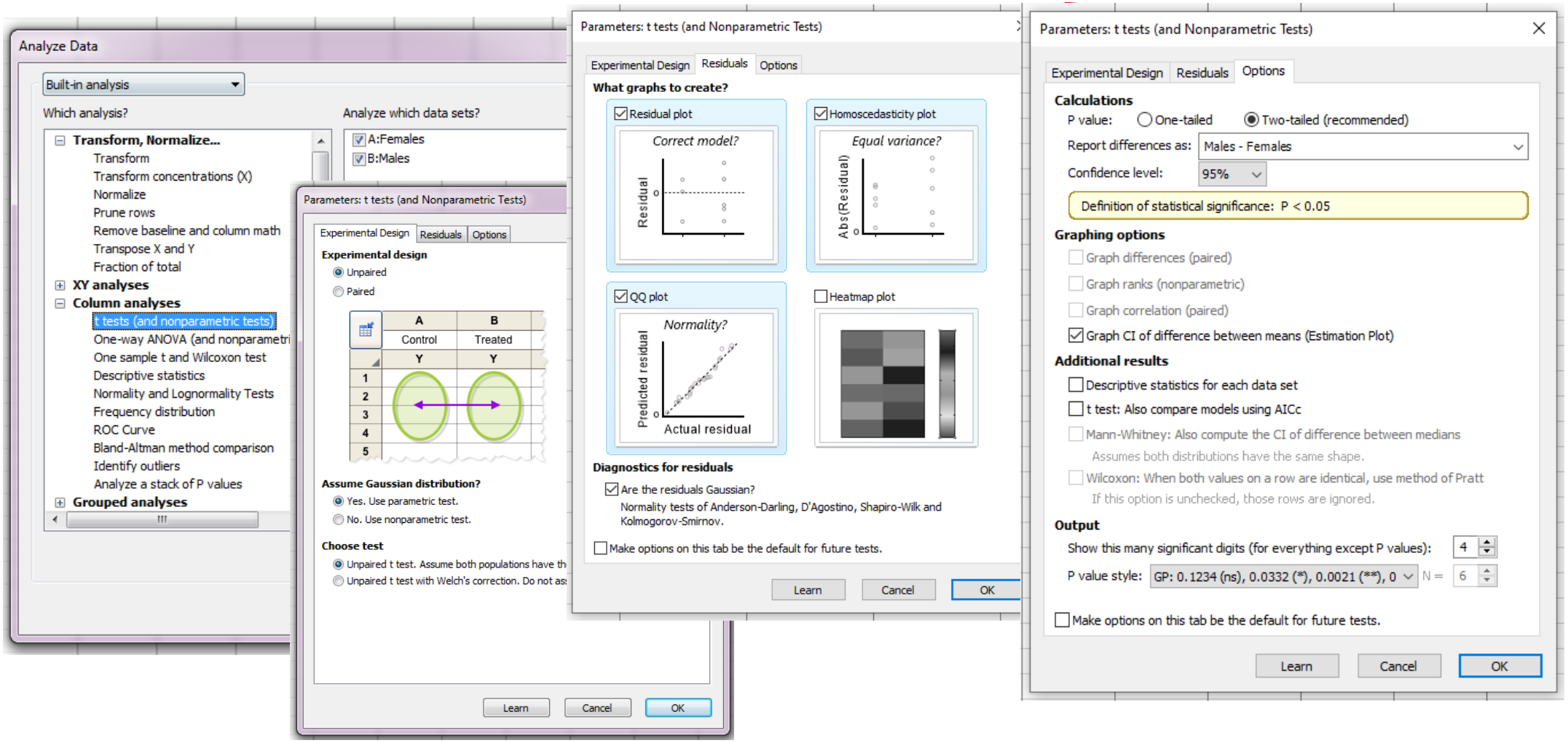




Could also be
4 independent experiments



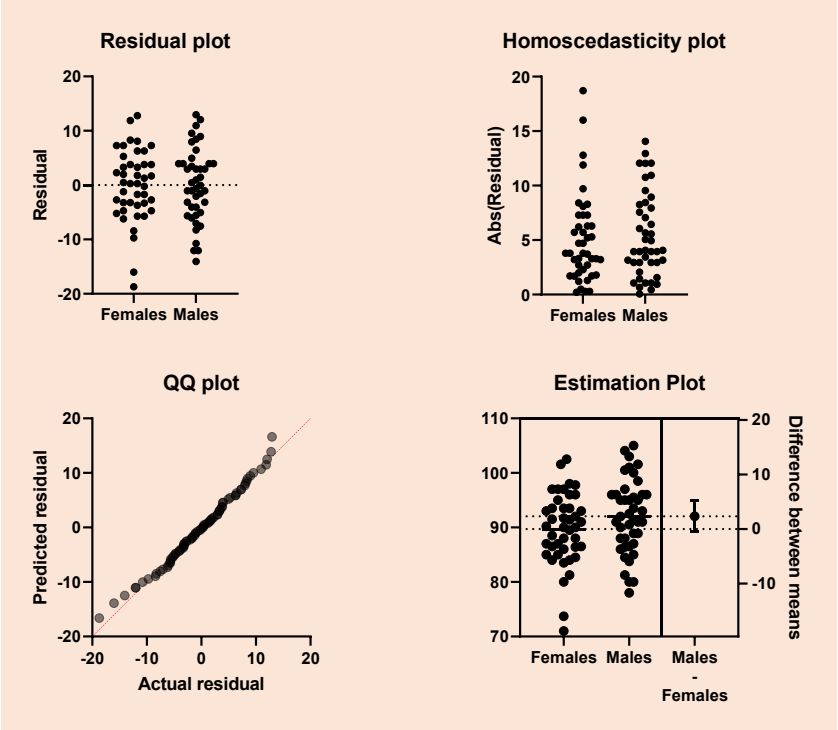
# Student's *t*-test: Residuals




**Have a go!**

# Independent *t*-test: Results

| | Unpaired t test | |
|---|---|---|
| 1 | Table Analyzed | Coyote |
| 2 | | |
| 3 | Column A | Females |
| 4 | vs. | vs. |
| 5 | Column B | Males |
| 6 | | |
| 7 | **Unpaired t test** | |
| 8 | P value | 0.1045 |
| 9 | P value summary | ns |
| 10 | Significantly different (P < 0.05)? | No |
| 11 | One- or two-tailed P value? | Two-tailed |
| 12 | t, df | t=1.641, df=84 |
| 13 | | |
| 14 | **How big is the difference?** | |
| 15 | Mean of column A | 89.71 |
| 16 | Mean of column B | 92.06 |
| 17 | Difference between means (A - B) ± SEM | -2.344 ± 1.428 |
| 18 | 95% confidence interval | -5.185 to 0.4964 |
| 19 | R squared (eta squared) | 0.03107 |
| 20 | | |
| 21 | **F test to compare variances** | |
| 22 | F, DFn, Dfd | 1.045, 42, 42 |
| 23 | P value | 0.8870 |
| 24 | P value summary | ns |
| 25 | Significantly different (P < 0.05)? | No |
| 26 | | |
| 27 | **Data analyzed** | |
| 28 | Sample size, column A | 43 |
| 29 | Sample size, column B | 43 |
| 30 | | |




**Normality ☑**

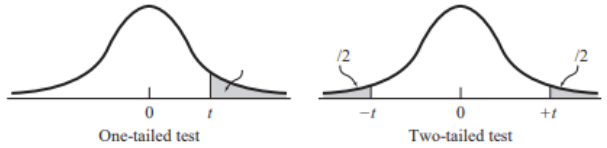
| Normality of Residuals | | | | |
|---|---|---|---|---|
| **Test name** | **Statistics** | **P value** | **Passed normality test (alpha=0.05)?** | **P value summary** |
| Anderson-Darling (A2*) | 0.2183 | 0.8347 | Yes | ns |
| D'Agostino-Pearson omnibus (K2) | 1.690 | 0.4295 | Yes | ns |
| Shapiro-Wilk (W) | 0.9873 | 0.5683 | Yes | ns |
| Kolmogorov-Smirnov (distance) | 0.05634 | 0.1000 | Yes | ns |

**Homogeneity in variance ☑**

# Independent *t*-test: Results

## *The old-fashion way*

| Unpaired t test | |
|---|---|
| P value | 0.1045 |
| P value summary | ns |
| Significantly different (P < 0.05)? | No |
| One- or two-tailed P value? | Two-tailed |
| t, df | t=1.641, df=84 |



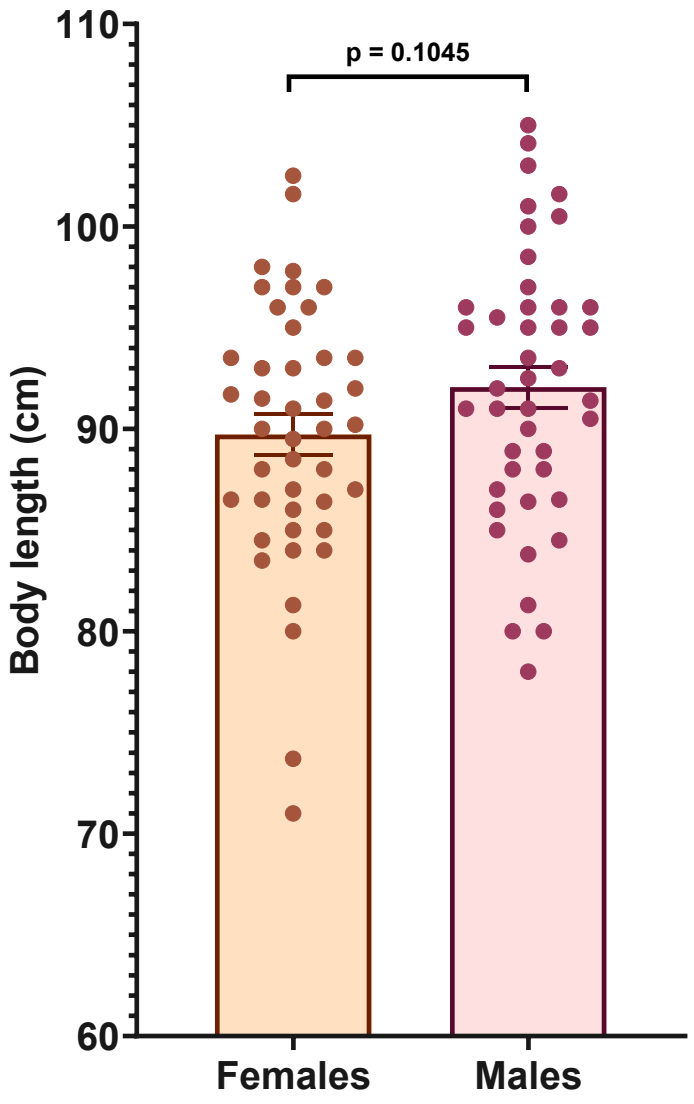

| | Level of Significance for One-Tailed Test | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *0.25* | *0.20* | *0.15* | *0.10* | *0.05* | *0.025* | *0.01* | *0.005* | *0.0005* |
| | Level of Significance for Two-Tailed Test | | | | | | | | |
| *df* | *0.50* | *0.40* | *0.30* | *0.20* | *0.10* | *0.05* | *0.02* | *0.01* | *0.001* |
| 1 | 1.000 | 1.376 | 1.963 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 636.620 |
| 2 | 0.816 | 1.061 | 1.386 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 31.599 |
| 3 | 0.765 | 0.978 | 1.250 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 12.924 |
| 4 | 0.741 | 0.941 | 1.190 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 8.610 |
| 5 | 0.727 | 0.920 | 1.156 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 6.869 |
| 6 | 0.718 | 0.906 | 1.134 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.959 |
| 7 | 0.711 | 0.896 | 1.119 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 5.408 |
| 8 | 0.706 | 0.889 | 1.108 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 5.041 |
| 9 | 0.703 | 0.883 | 1.100 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.781 |
| 10 | 0.700 | 0.879 | 1.093 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.587 |
| 11 | 0.697 | 0.876 | 1.088 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.437 |
| 12 | 0.695 | 0.873 | 1.083 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 4.318 |
| 13 | 0.694 | 0.870 | 1.079 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 4.221 |
| 14 | 0.692 | 0.868 | 1.076 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 4.140 |
| 15 | 0.691 | 0.866 | 1.074 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 4.073 |
| 16 | 0.690 | 0.865 | 1.071 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 4.015 |
| 17 | 0.689 | 0.863 | 1.069 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.965 |
| 18 | 0.688 | 0.862 | 1.067 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.922 |
| 19 | 0.688 | 0.861 | 1.066 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.883 |
| 20 | 0.687 | 0.860 | 1.064 | 1.325 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |
| 21 | 0.686 | 0.859 | 1.063 | 1.323 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |
| 22 | 0.686 | 0.858 | 1.061 | 1.321 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |
| 23 | 0.685 | 0.858 | 1.060 | 1.319 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |
| 24 | 0.685 | 0.857 | 1.059 | 1.318 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |
| 25 | 0.684 | 0.856 | 1.058 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.725 |
| 26 | 0.684 | 0.856 | 1.058 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.707 |
| 27 | 0.684 | 0.855 | 1.057 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.690 |
| 28 | 0.683 | 0.855 | 1.056 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.674 |
| 29 | 0.683 | 0.854 | 1.055 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.659 |
| 30 | 0.683 | 0.854 | 1.055 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.646 |
| 40 | 0.681 | 0.851 | 1.050 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 3.551 |
| 50 | 0.679 | 0.849 | 1.047 | 1.299 | 1.676 | 2.009 | 2.403 | 2.678 | 3.496 |
| 100 | 0.677 | 0.845 | 1.042 | 1.290 | 1.660 | 1.984 | 2.364 | 2.626 | 3.390 |
| [illegible] | 0.674 | 0.842 | 1.036 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.291 |

**$t$ = 1.641 < 1.984: not significant**

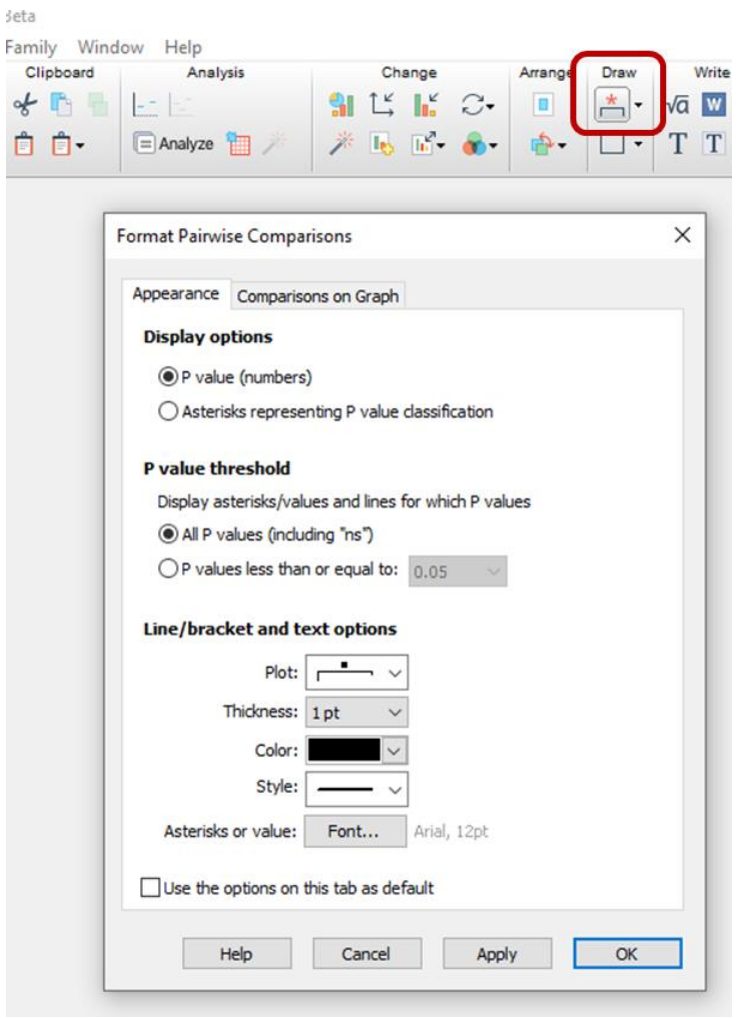
Critical value

**What about power?**

# Independent *t*-test: Results

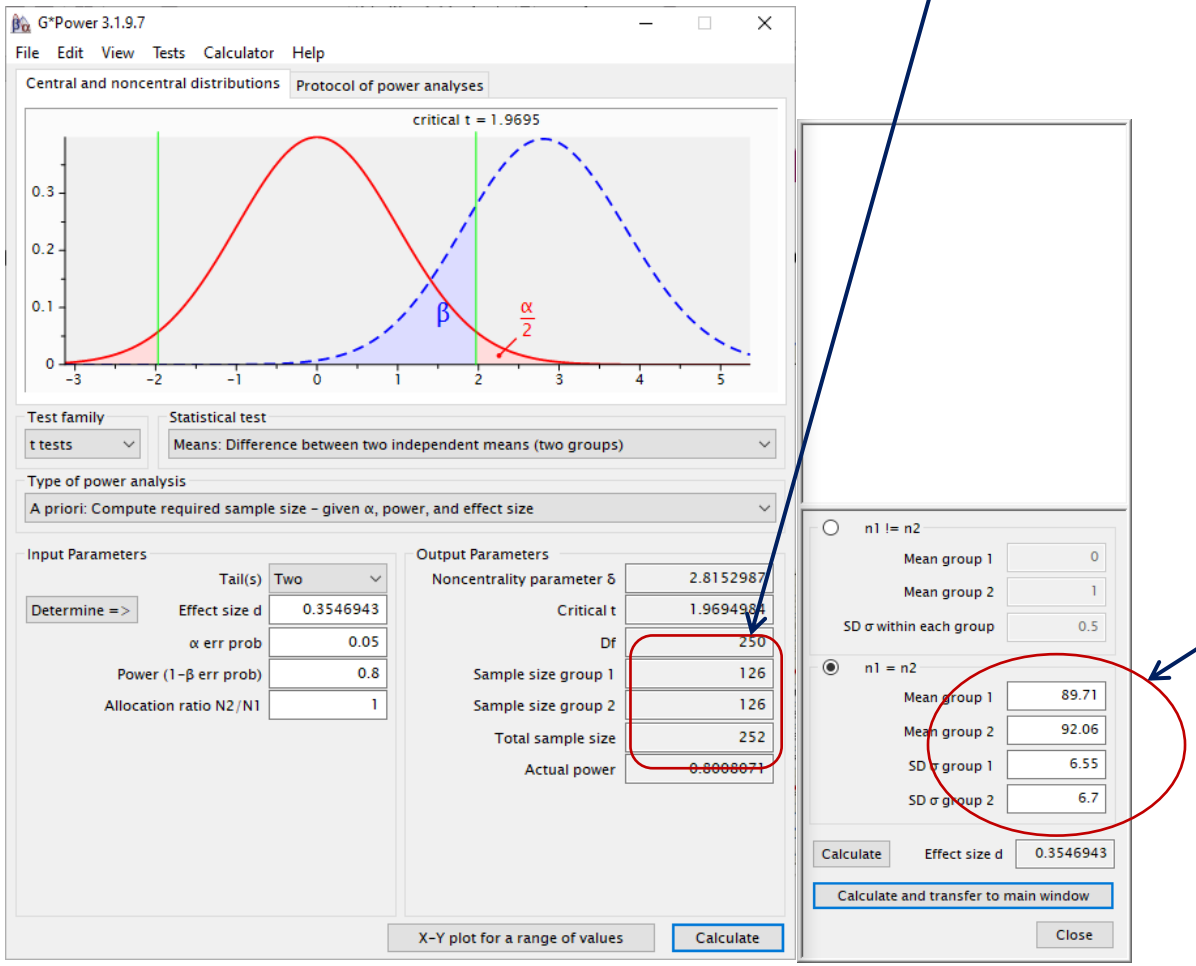



**Males tend to be longer than females but not significantly so (p=0.1045).**



# Power analysis

You would need a sample 3 times bigger to reach the accepted power of 80%




| | Col. stats | A | B |
|---|---|---|---|
| | | Females | Males |
| 1 | Number of values | 43 | 43 |
| 2 | | | |
| 3 | Minimum | 71.00 | 78.00 |
| 4 | 25% Percentile | 86.00 | 87.00 |
| 5 | Median | 90.00 | 92.00 |
| 6 | 75% Percentile | 93.50 | 96.00 |
| 7 | Maximum | 102.5 | 105.0 |
| 8 | | | |
| 9 | Mean | 89.71 | 92.06 |
| 10 | Std. Deviation | 6.550 | 6.696 |
| 11 | Std. Error of Mean | 0.9988 | 1.021 |
| 12 | | | |
| 13 | Lower 95% CI of mean | 87.70 | 90.00 |
| 14 | Upper 95% CI of mean | 91.73 | 94.12 |
| 15 | | | |
| 16 | Sum | 3858 | 3958 |
| 17 | | | |
| 18 | **D'Agostino & Pearson normality test** | | |
| 19 | K2 | 4.203 | 0.5080 |
| 20 | P value | 0.1223 | 0.7757 |
| 21 | Passed normality test (alpha=0.05)? | Yes | Yes |
| 22 | P value summary | ns | ns |
| 23 | | | |
| 24 | **Shapiro-Wilk normality test** | | |
| 25 | W | 0.9700 | 0.9845 |
| 26 | P value | 0.3164 | 0.8190 |
| 27 | Passed normality test (alpha=0.05)? | Yes | Yes |
| 28 | P value summary | ns | ns |

**But is a 2.3 cm difference between sexes biologically relevant (<3%) ?**

# Exercise 2

Babraham Institute

# Analysis of Quantitative data
# One-Way ANOVA

Hayley Carr & Anne Segonds-Pichon

v2024-05

Babraham Bioinformatics

# Analysis of Quantitative data
## One-Way + Two-Way ANOVA

- One-way ANOVA
  - Independent design
  - Repeated measures design

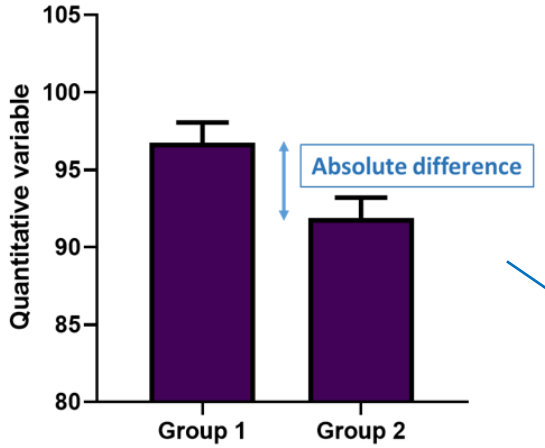
- Two-way ANOVA (two factors/predictors)
  - Tests **each factor** and **interactions** between them
  - Independent design
  - Repeated measures design (time series)

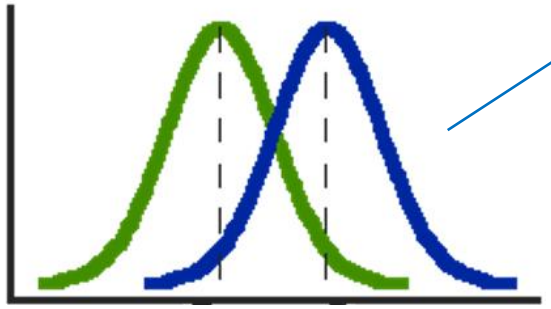
# Comparison between more than 2 groups

## One factor = One predictor

## One-Way ANOVA

# Signal-to-noise ratio



$$\frac{\text{Difference (signal)}}{\text{Variation (noise)}}$$

$$\frac{\text{Signal}}{\text{Noise}} = \textbf{statistical significance}$$

$$\frac{\text{Signal}}{\text{Noise}} = \textbf{no statistical significance}$$

# Analysis of variance: how does it work?

$$\frac{\text{Signal}}{\text{Noise}} = \frac{\text{Difference between the means}}{\text{Variability in the groups}}$$

$$= \text{F ratio}$$

- If the variance amongst sample means is greater than the error/random variance, then F>1
  - In an ANOVA, we test whether F is significantly higher than 1 or not

# One-Way Analysis of variance

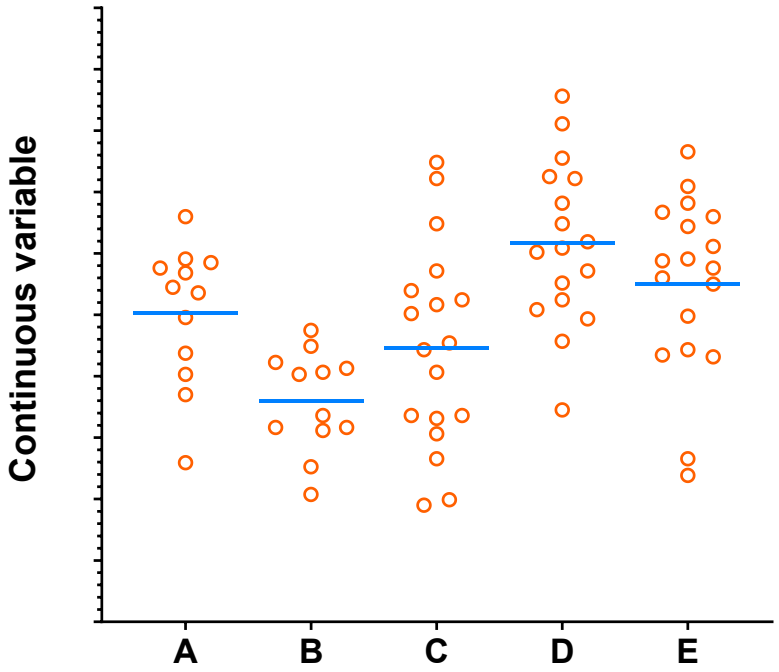
**Step 1: Omnibus test**

- It tells us if there is a difference between the means but not which means are significantly different from which other ones

**Step 2: Post-hoc tests**

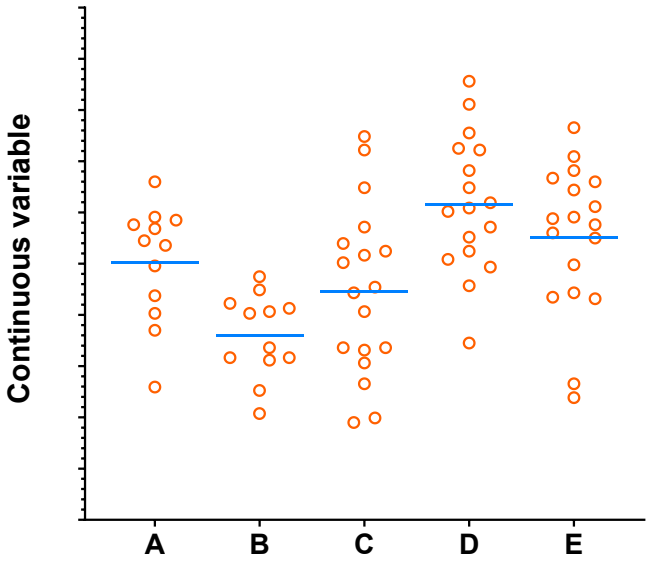
- Tell us if there are differences between the group means pairwise
- A correction for multiple comparisons will be applied on the p-values
- *Should only be used when the ANOVA finds a significant effect*

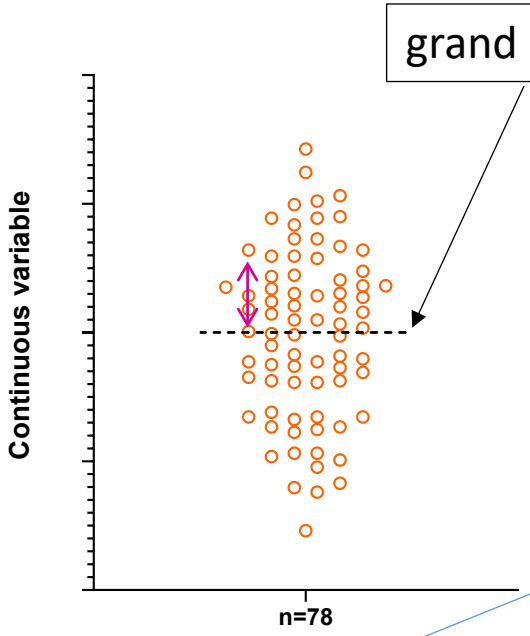
# Analysis of variance: how does it work?




| Source of variation | Sum of Squares | df | Mean Square | F | p-value |
|---|---|---|---|---|---|
| Between Groups | 18.1 | 4 | 4.5 | 6.32 | 0.0002 |
| Within Groups | 51.8 | 73 | 0.71 | | |
| Total | 69.9 | | | | |

# Analysis of variance: how does it work?




grand mean

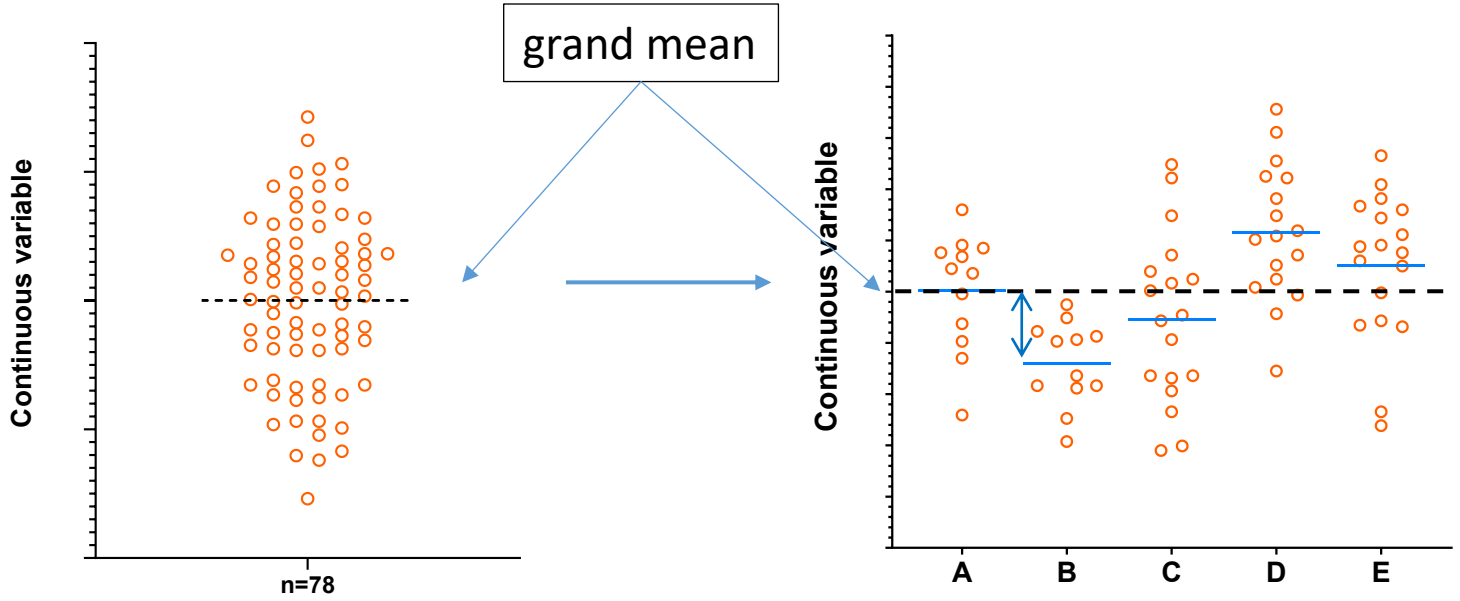
78 differences: $\sum_1^{78} (\text{value}_n - \text{grand mean})^2$

=

**Sum of squared errors**

| Source of variation | Sum of Squares | df | Mean Square | F | p-value |
|---|---|---|---|---|---|
| Between Groups | | | | | |
| Within Groups | | | | | |
| Total | 69.9 | | | | |

# Analysis of variance: how does it work?




5 differences: $\sum_1^5 (\text{mean}_n - \text{grand mean})^2$

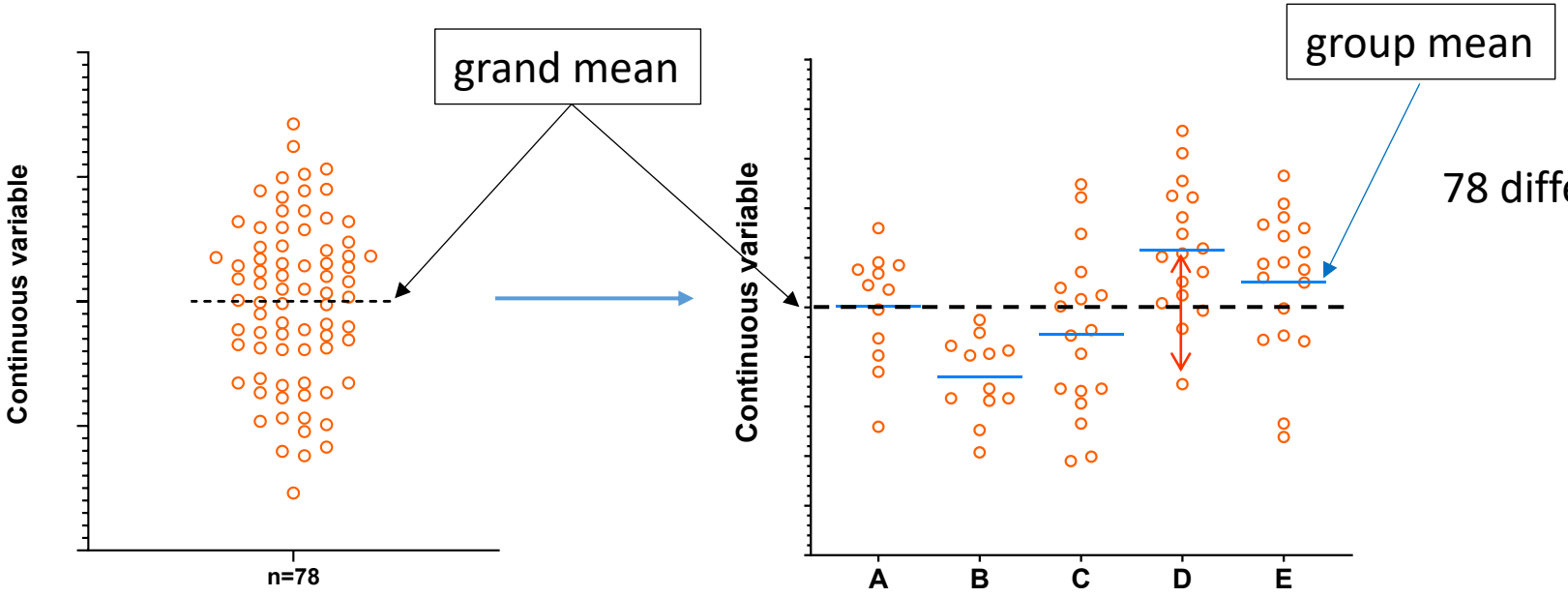
=

**Sum of squared errors**

**Between the groups**

| Source of variation | Sum of Squares | df | Mean Square | F | p-value |
|---|---|---|---|---|---|
| Between Groups | 18.1 | | | | |
| Within Groups | | | | | |
| Total | 69.9 | | | | |

# Analysis of variance: how does it work?




78 differences: $\sum_{1}^{78} (\text{value}_n - \text{group mean})^2$

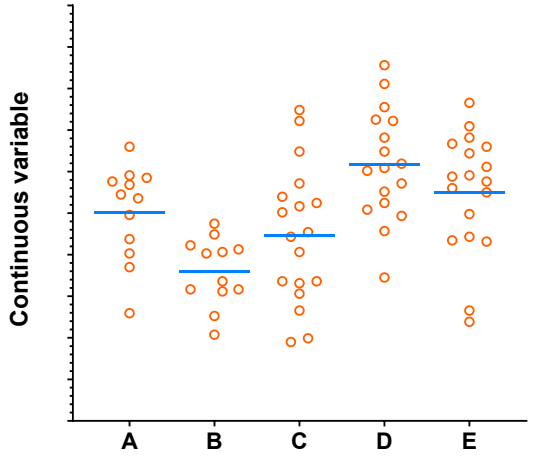
=

**Sum of squared errors**

**Within the Groups**

| Source of variation | Sum of Squares | df | Mean Squares | F | p-value |
|---|---|---|---|---|---|
| Between Groups | 18.1 | | | | |
| Within Groups | 51.8 | | | | |
| Total | 69.9 | | | | |

# Analysis of variance: how does it work?




| | Source of variation | Sum of Squares | df | Mean Squares | F ratio | p-value |
|---|---|---|---|---|---|---|
| Signal | Between Groups | 18.1 | k-1 | | | |
| Noise | Within Groups | 51.8 | n-k | | | |
| | Total | 69.9 | | | | |

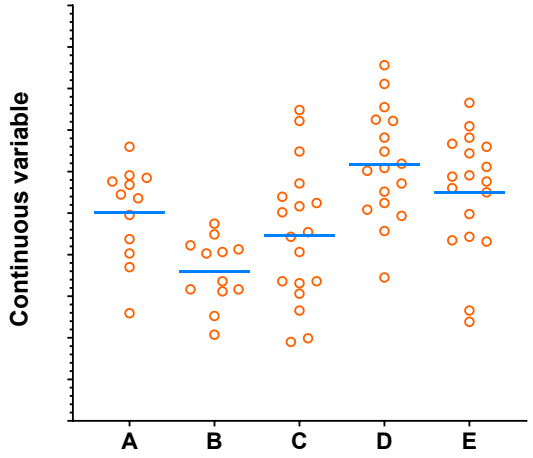
df: degree of freedom with df = n-1

n = number of values, k = number of groups

Between groups: df = 4 (k-1)

Within groups: df = 73 ($n-k = n_1-1 + \ldots + n_5-1$)

# Analysis of variance: how does it work?



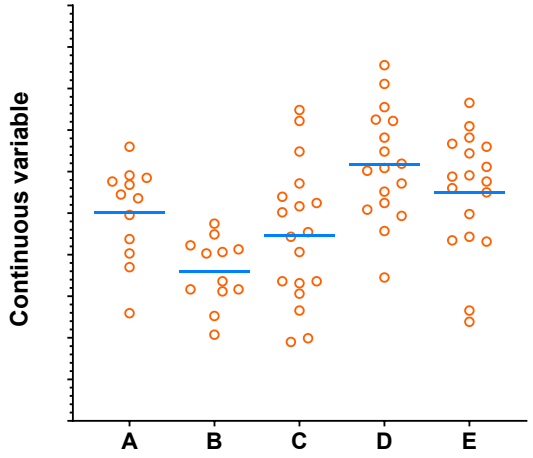

| | Source of variation | Sum of Squares | df | Mean Squares | F ratio | p-value |
|---|---|---|---|---|---|---|
| Signal | Between Groups | 18.1 | 4 | 4.5 | | |
| Noise | Within Groups | 51.8 | 73 | 0.71 | | |
| | Total | 69.9 | | | | |

df: degree of freedom with df = n-1

**18.2/4 = 4.5** **51.8/73 = 0.71**

Mean squares = **Sum of Squares / n-1 = Variance!**

# Analysis of variance: how does it work?




| Source of variation | Sum of Squares | df | Mean Squares | F ratio | p-value |
|---|---|---|---|---|---|
| Between Groups | 18.1 | 4 | 4.5 | 6.34 | 0.0002 |
| Within Groups | 51.8 | 73 | 0.71 | | |
| Total | 69.9 | | | | |

Mean squares = **Sum of Squares / n-1 = Variance**

$$\text{F ratio} = \frac{\text{Variance between the groups}}{\text{Variance within the groups (individual variability)}} = \frac{4.5}{0.71} = \mathbf{6.34}$$

# One-Way Analysis of variance

**Step 1: Omnibus test**

- It tells us if there is a difference between the means but not which means are significantly different from which other ones

**Step 2: Post-hoc tests**

- Tell us if there are differences between the group means pairwise
- A correction for multiple comparisons will be applied on the p-values
- *Should only be used when the ANOVA finds a significant effect*

# Comparison of more than 2 means

- Running multiple tests on the same data increases the **familywise error rate**

  = error rate across tests on the same experimental data

- One of the basic rules ('laws') of probability:
  - The Multiplicative Rule: The probability of the joint occurrence of 2 or more independent events is the product of the individual probabilities

$$\boxed{P(A,B) = P(A) \times P(B)}$$

For example:

$$P(2 \text{ heads}) = P(\text{head}) \times P(\text{head}) = 0.5 \times 0.5 = 0.25$$

# Familywise error rate

- **<u>Example</u>**: All pairwise comparisons between 3 groups A, B and C:
  = A-B, A-C and B-C

- Probability of making the Type I Error: **5%**
  → probability of not making the Type I Error is 95% (= 1 – 0.05)

- Multiplicative Rule:
  - Overall probability of no Type I errors = 0.95*0.95*0.95 = 0.857

- Probability of making at least one Type I Error = 1 – 0.857 = 0.143 or **14.3%**
  - Probability has increased from **5% → 14.3%**

- For comparisons between 5 groups, the familywise error rate is **40%** (=1-$(0.95)^n$)

# Familywise error rate

- Solution to increased familywise error rate = correction for multiple comparisons

  → post-hoc tests

- Many different approaches:
  - Different statisticians addressed different issues
    - e.g. unbalanced design, heterogeneity of variance, liberal vs conservative

- Two main ways to address the multiple testing problem:
  - **Familywise Error Rate** (**FWER**) and **False Discovery Rate** (**FDR**)

- In all cases:

  More tests → higher familywise error rate → more stringent correction

# Multiple testing problem

- **Difference between FWER and FDR**:
  - FWER: a p-value of 0.05 implies that 5% of **all tests** will result in false positives
  - FDR: an adjusted p-value (or **q-value**) of 0.05 implies that 5% of **significant tests** will result in false positives
- **FWER**: **Bonferroni**: $\alpha_{adjust}$ = 0.05/n comparisons, e.g. 3 comparisons: 0.05/3=0.016
  - Problem: very conservative leading to **loss of power** (lots of false negative)
  - 10 comparisons: threshold for significance = 0.05/10 = 0.005
  - Pairwise comparisons across 20,000 genes = 0.05/20,000 = $2.5 \times 10^{-6}$
- **FDR**: **Benjamini-Hochberg**: controls the expected proportion of "discoveries" (significant tests) that are false (false positive)
  - Correction applied only on the significant tests
  - **More power** but increased Type I Errors

# Repeated measures One-Way ANOVA

- **A new assumption:**
  - That the variances of the differences between all combinations of related conditions (or group levels) are equal – known as the **assumption of sphericity**
  - The **Mauchly's test of sphericity** is used to assess whether the assumption of sphericity is met
  - If the assumption of sphericity is not met, a correction is applied
    - Often the default as the assumption is seldom met
  - Most common correction: **Greenhouse-Geisser correction**

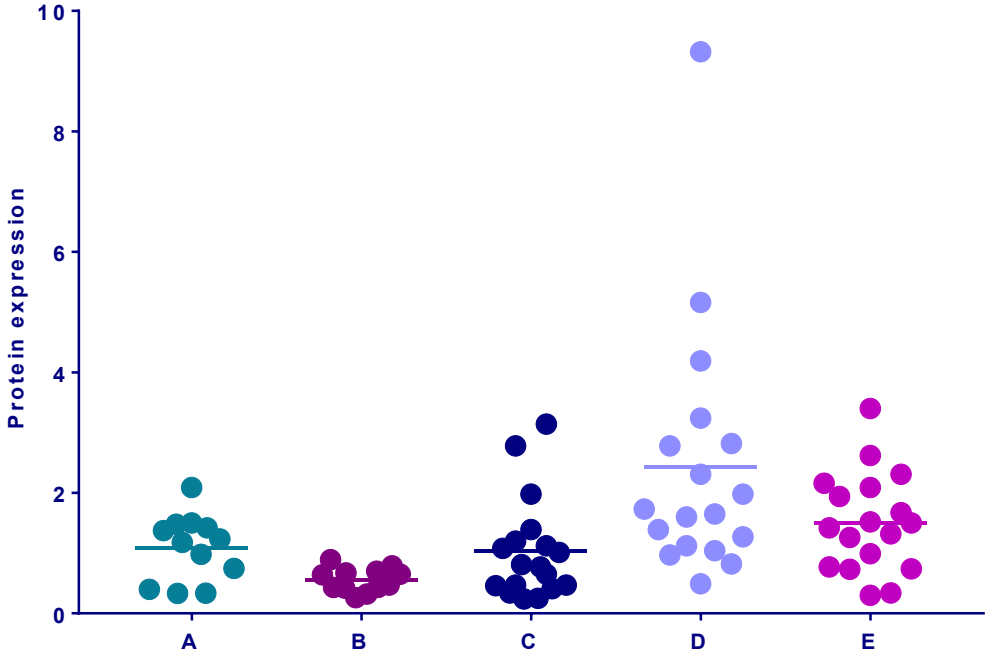
# Exercise: One-way ANOVA

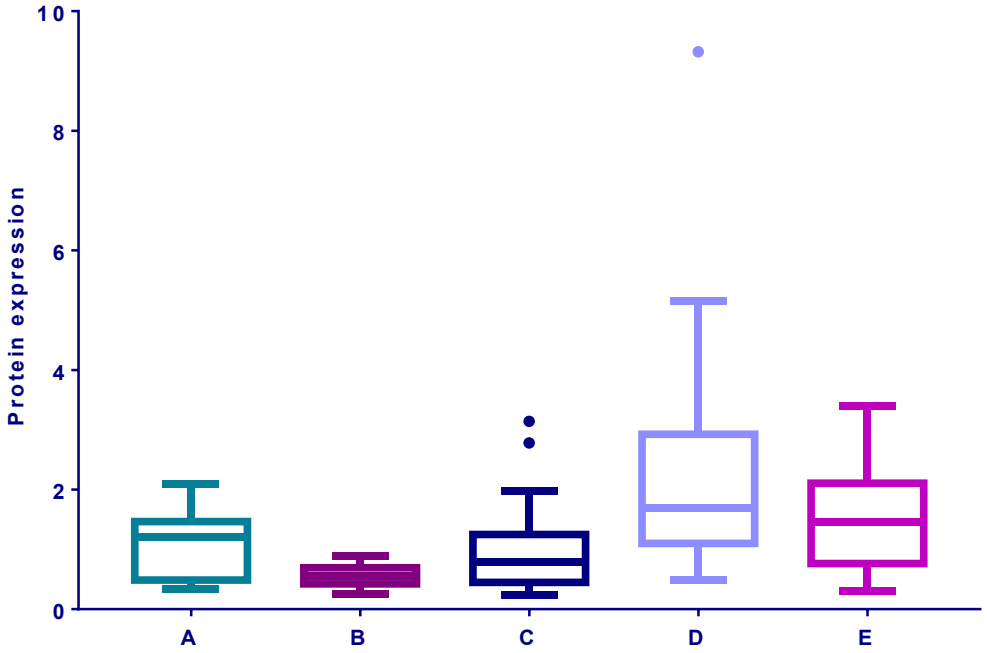
## protein expression.xlsx

- **Question**: is there a difference in protein expression between the 5 cell lines?

- **1 Plot the data**

- **2 Check the assumptions for parametric test**

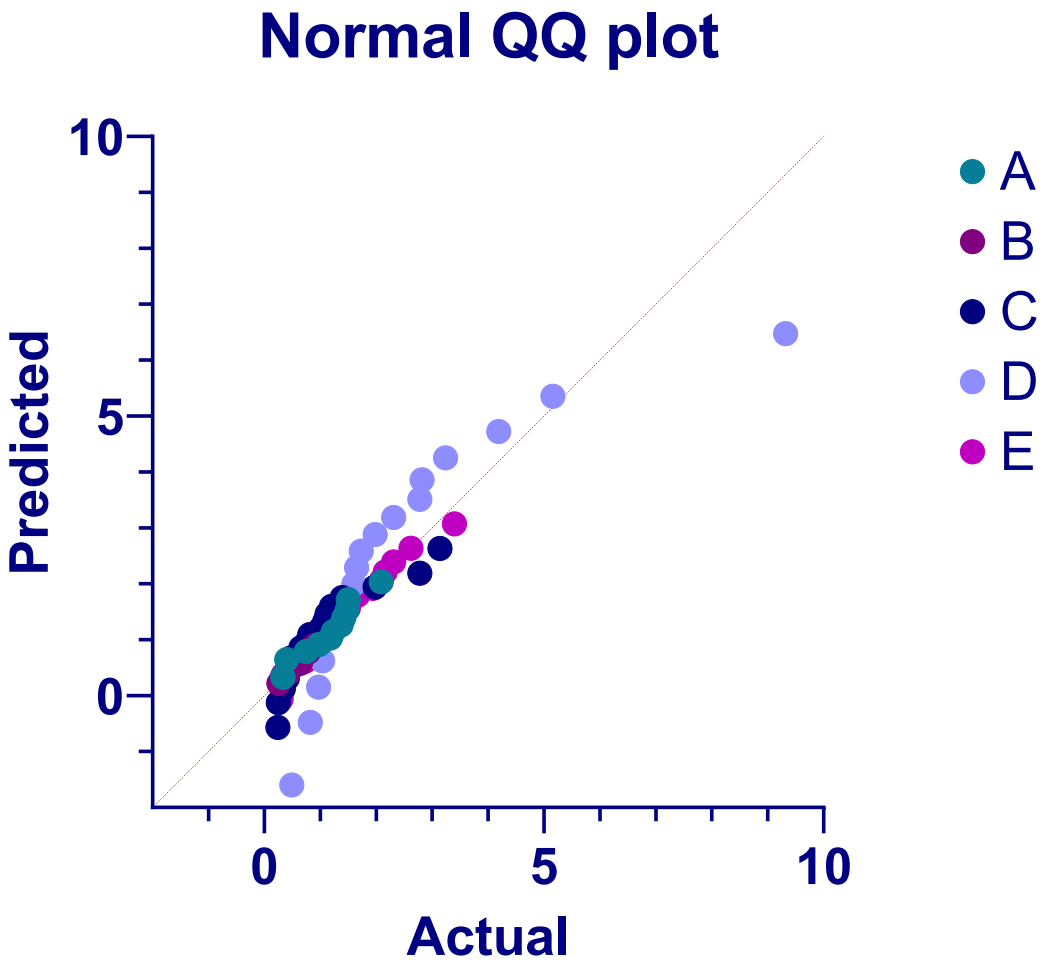

Protein expression
0
2
4
6
8
10
A
B
C
D
E


Protein expression
0
2
4
6
8
10
A
B
C
D
E

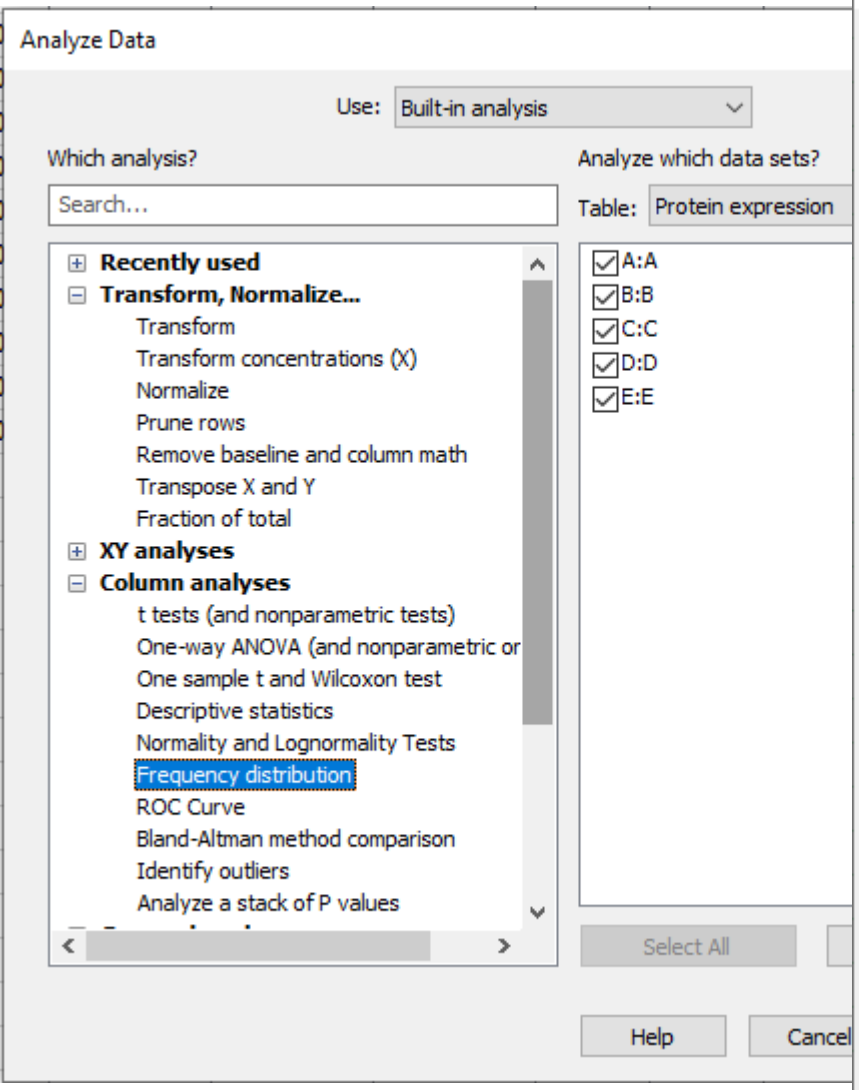
# Parametric tests assumptions

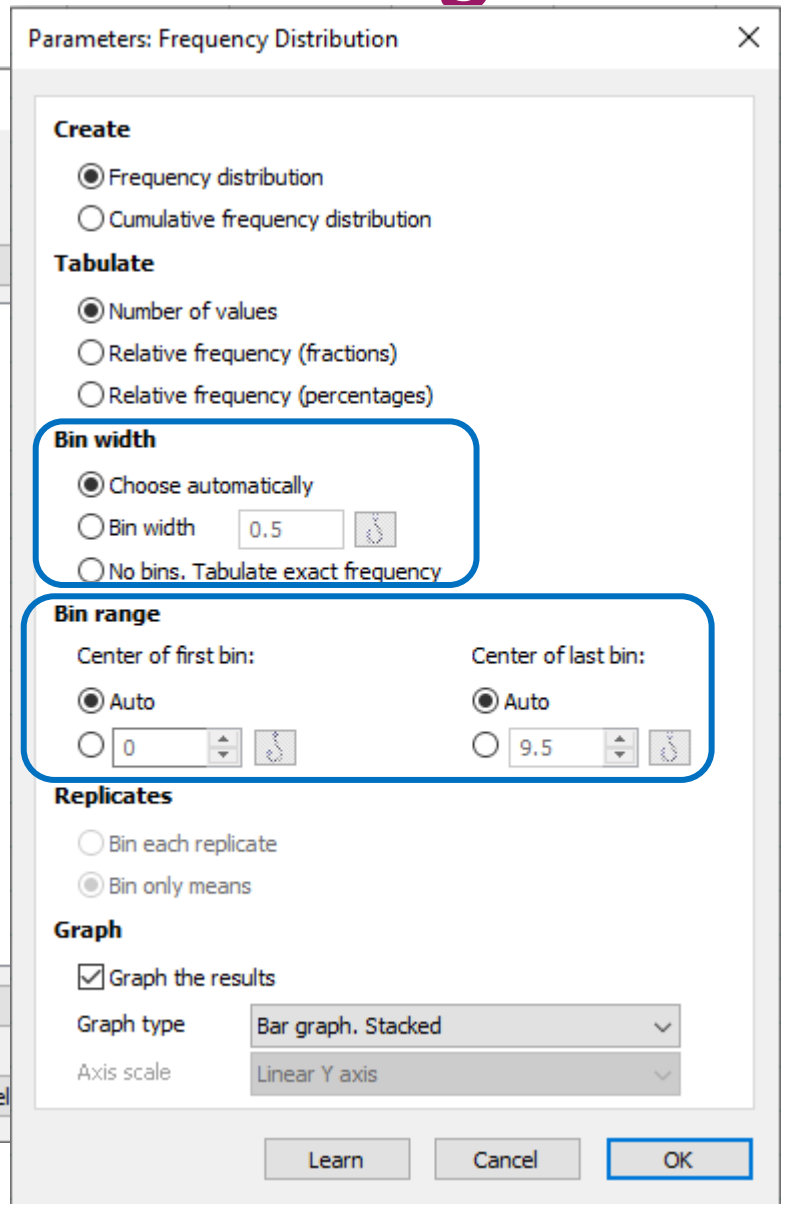
Tabular results

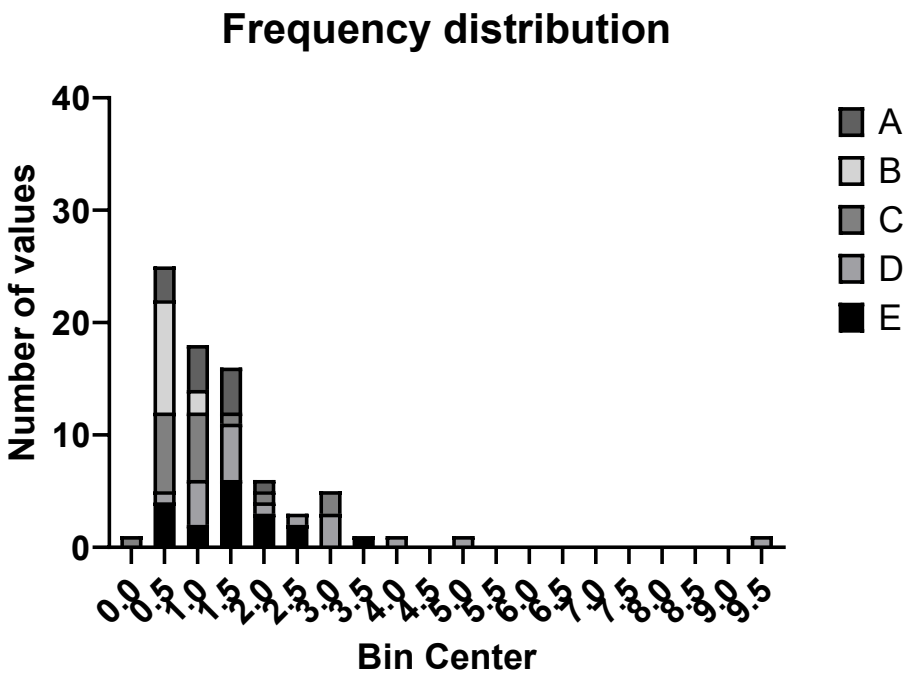
| Normality and Lognormality Tests<br>Tabular results | A<br>A | B<br>B | C<br>C | D<br>D | E<br>E |
|---|---|---|---|---|---|
| **Test for normal distribution** | | | | | |
| **Anderson-Darling test** | | | | | |
| A2* | 0.3797 | 0.3141 | 1.166 | 1.439 | 0.2011 |
| P value | 0.3446 | 0.5029 | 0.0035 | 0.0007 | 0.8590 |
| Passed normality test (alpha=0.05)? | Yes | Yes | No | No | Yes |
| P value summary | ns | ns | ** | *** | ns |
| | | | | | |
| **D'Agostino & Pearson test** | | | | | |
| K2 | 0.1236 | 0.7508 | 9.375 | 22.59 | 1.280 |
| P value | 0.9401 | 0.6870 | 0.0092 | <0.0001 | 0.5274 |
| Passed normality test (alpha=0.05)? | Yes | Yes | No | No | Yes |
| P value summary | ns | ns | ** | **** | ns |
| | | | | | |
| **Shapiro-Wilk test** | | | | | |
| W | 0.9295 | 0.9535 | 0.8197 | 0.7531 | 0.9671 |
| P value | 0.3752 | 0.6888 | 0.0029 | 0.0004 | 0.7411 |
| Passed normality test (alpha=0.05)? | Yes | Yes | No | No | Yes |
| P value summary | ns | ns | ** | *** | ns |
| | | | | | |
| **Kolmogorov-Smirnov test** | | | | | |
| KS distance | 0.1485 | 0.1704 | 0.1980 | 0.2058 | 0.1035 |
| P value | >0.1000 | >0.1000 | 0.0603 | 0.0424 | >0.1000 |
| Passed normality test (alpha=0.05)? | Yes | Yes | Yes | No | Yes |
| P value summary | ns | ns | ns | * | ns |
| | | | | | |
| **Number of values** | 12 | 12 | 18 | 18 | 18 |



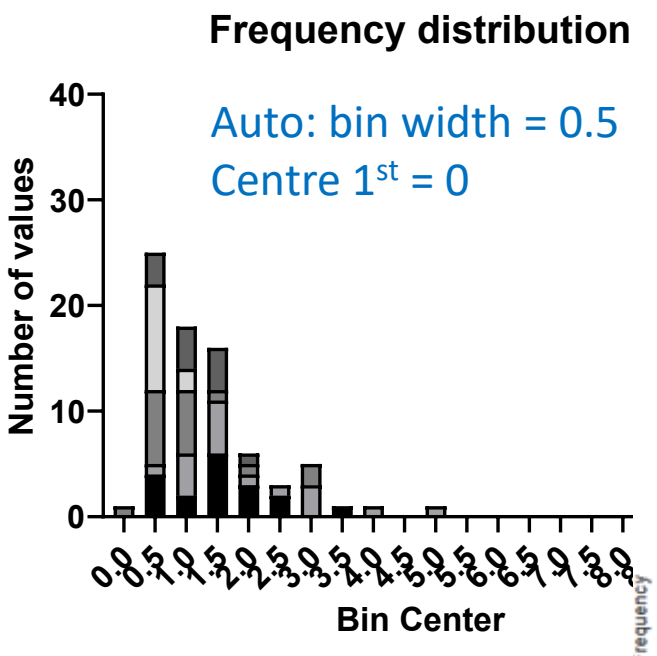
# Histogram

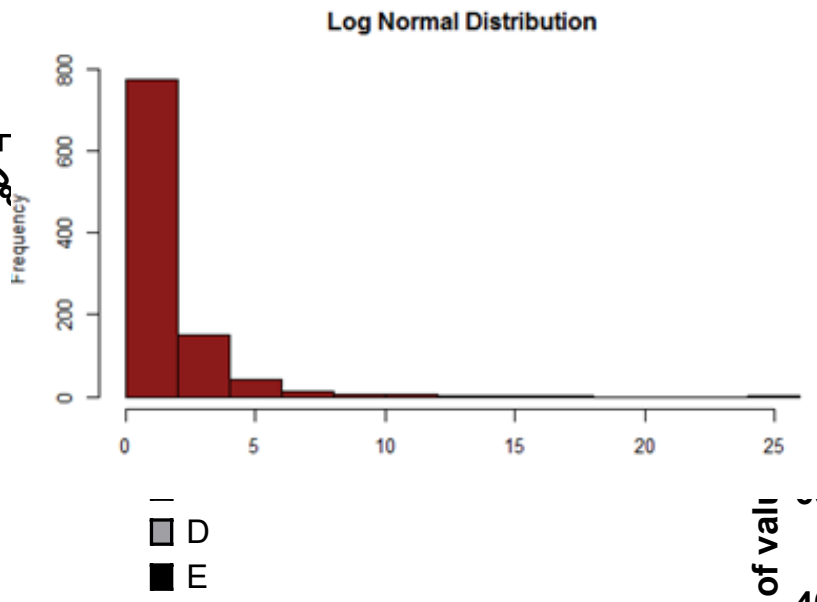

Analyze Data
Use: Built-in analysis
Which analysis?
Search...
Recently used
Transform, Normalize...
Transform
Transform concentrations (X)
Normalize
Prune rows
Remove baseline and column math
Transpose X and Y
Fraction of total
XY analyses
Column analyses
t tests (and nonparametric tests)
One-way ANOVA (and nonparametric or
One sample t and Wilcoxon test
Descriptive statistics
Normality and Lognormality Tests
Frequency distribution
ROC Curve
Bland-Altman method comparison
Identify outliers
Analyze a stack of P values
Analyze which data sets?
Table: Protein expression
A:A
B:B
C:C
D:D
E:E
Select All
Help
Cancel

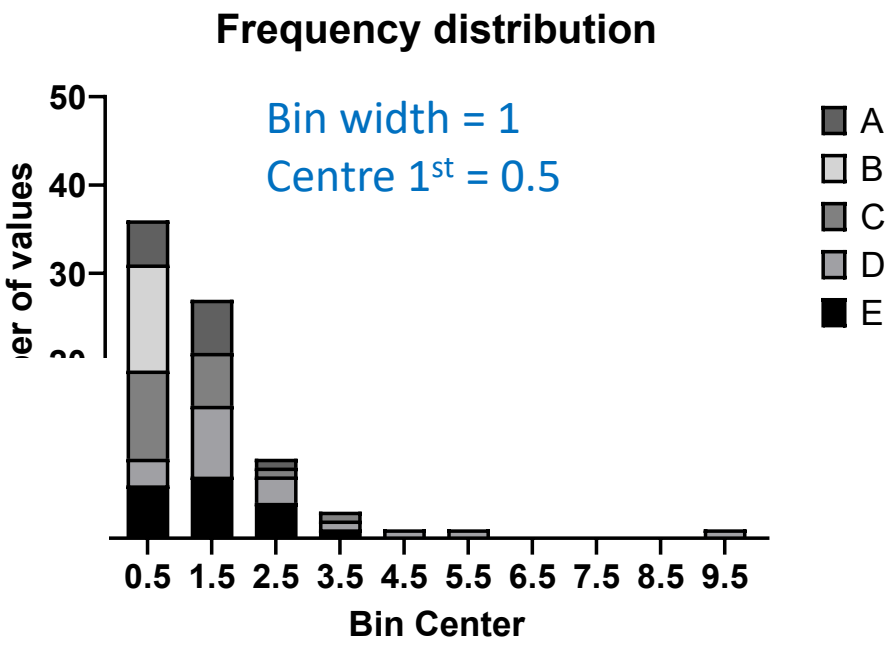

Parameters: Frequency Distribution
Create
Frequency distribution
Cumulative frequency distribution
Tabulate
Number of values
Relative frequency (fractions)
Relative frequency (percentages)
Bin width
Choose automatically
Bin width 0.5
No bins. Tabulate exact frequency
Bin range
Center of first bin:
Center of last bin:
Auto
0
Auto
9.5
Replicates
Bin each replicate
Bin only means
Graph
Graph the results
Graph type
Bar graph. Stacked
Axis scale
Linear Y axis
Learn
Cancel
OK

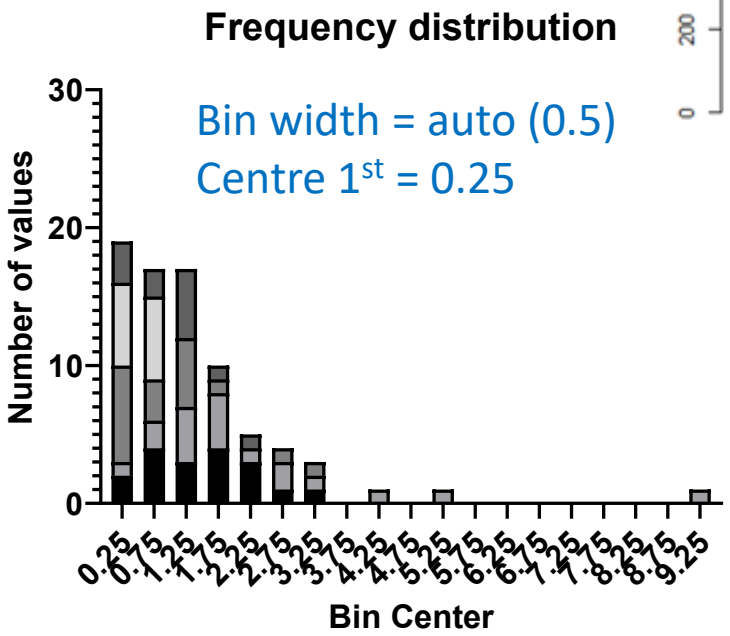

Frequency distribution
Number of values
40
30
20
10
0
Bin Center
A
B
C
D
E

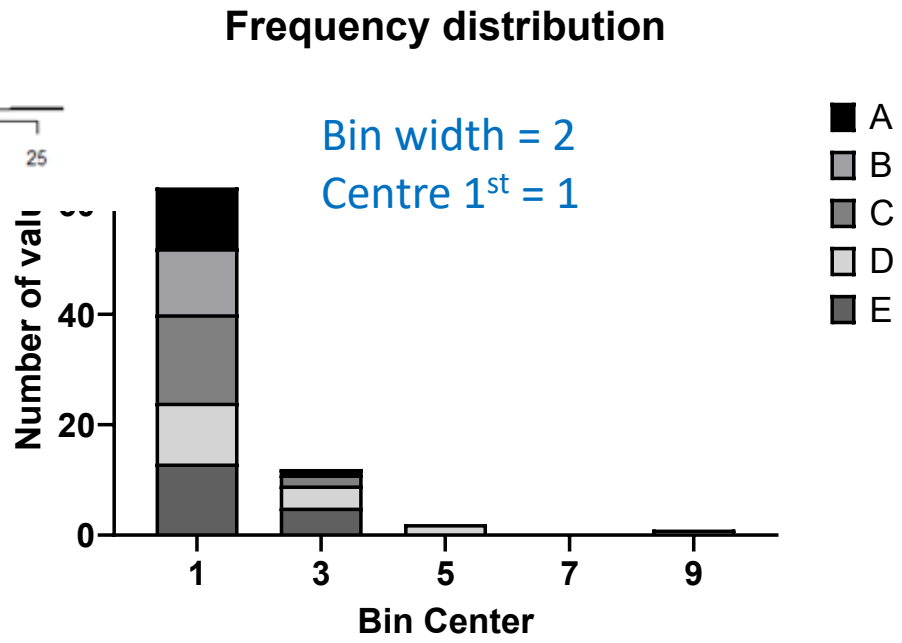
# Histogram

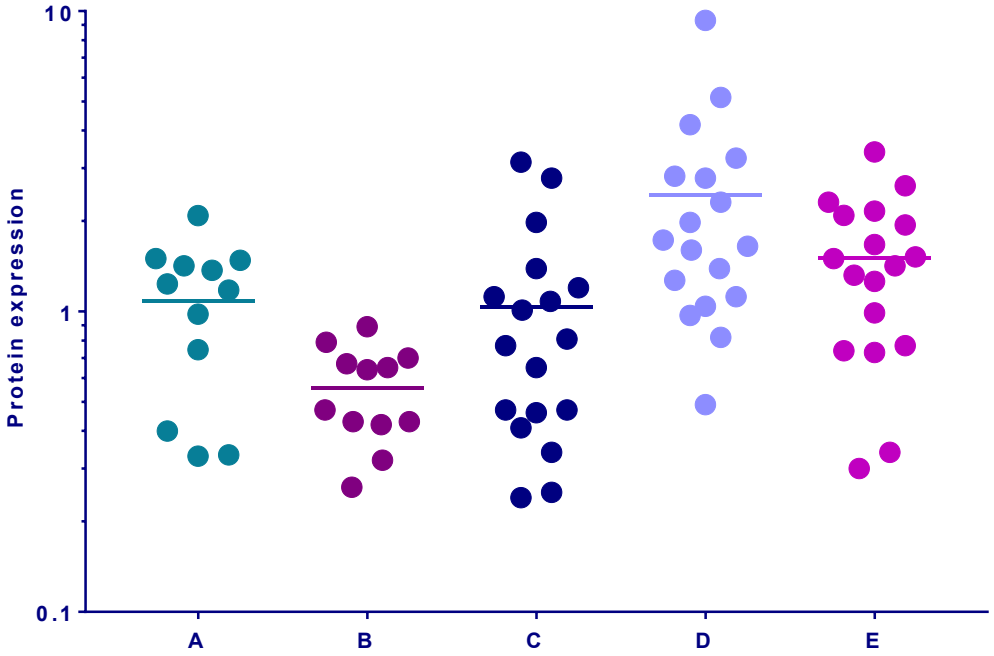

Frequency distribution
Auto: bin width = 0.5
Centre 1st = 0
Number of values
Bin Center
A
B
C
D
E

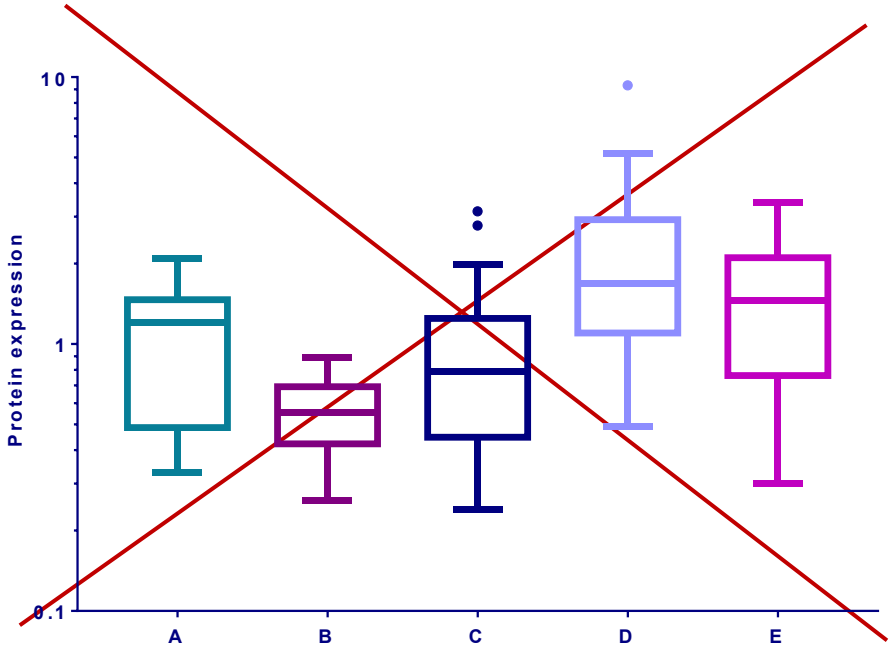

Log Normal Distribution
Frequency
D
E

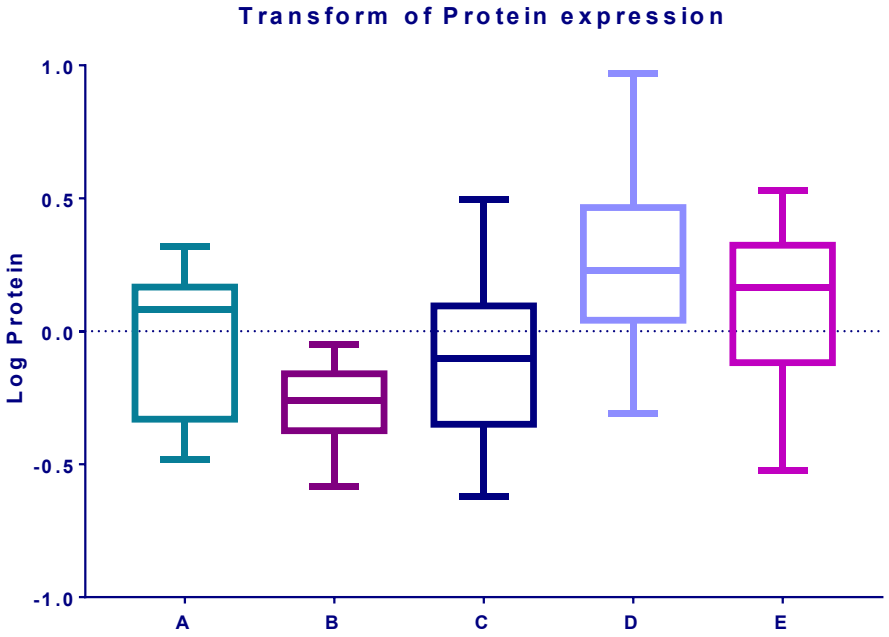

Frequency distribution
Bin width = 1
Centre 1st = 0.5
Bin Center
A
B
C
D
E


Frequency distribution
Bin width = auto (0.5)
Centre 1st = 0.25
Number of values
Bin Center


Frequency distribution
Bin width = 2
Centre 1st = 1
Bin Center
A
B
C
D
E


10
1
0.1
Protein expression
A
B
C
D
E

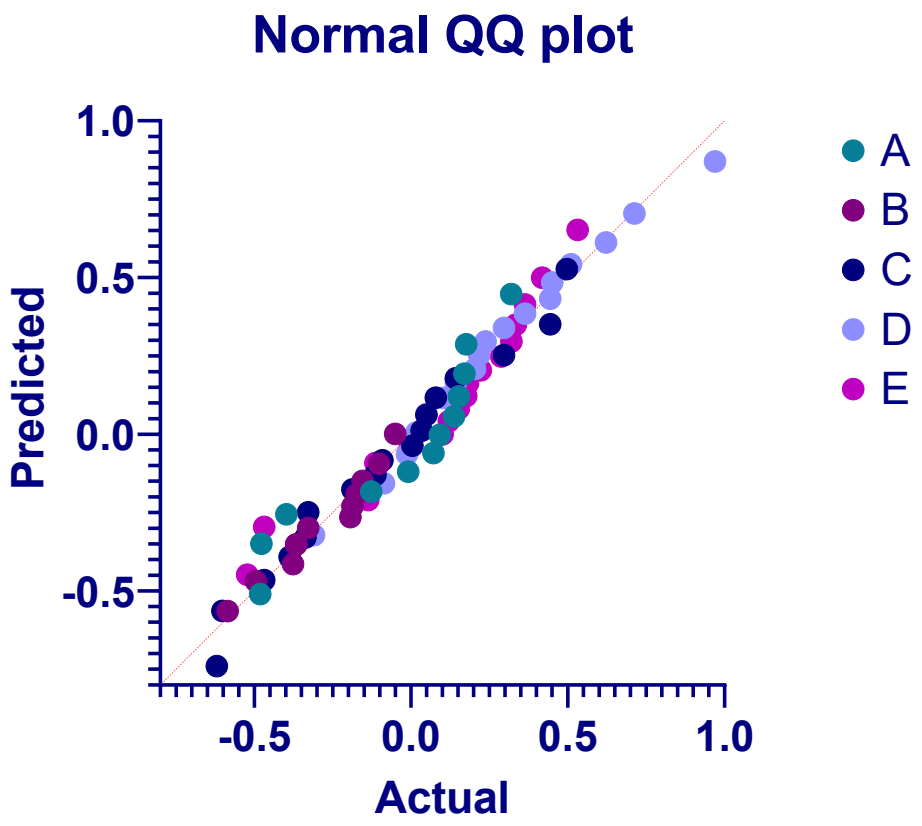

10
1
0.1
Protein expression
A
B
C
D
E


Transform of Protein expression
1.0
0.5
0.0
-0.5
-1.0
Log Protein
A
B
C
D
E

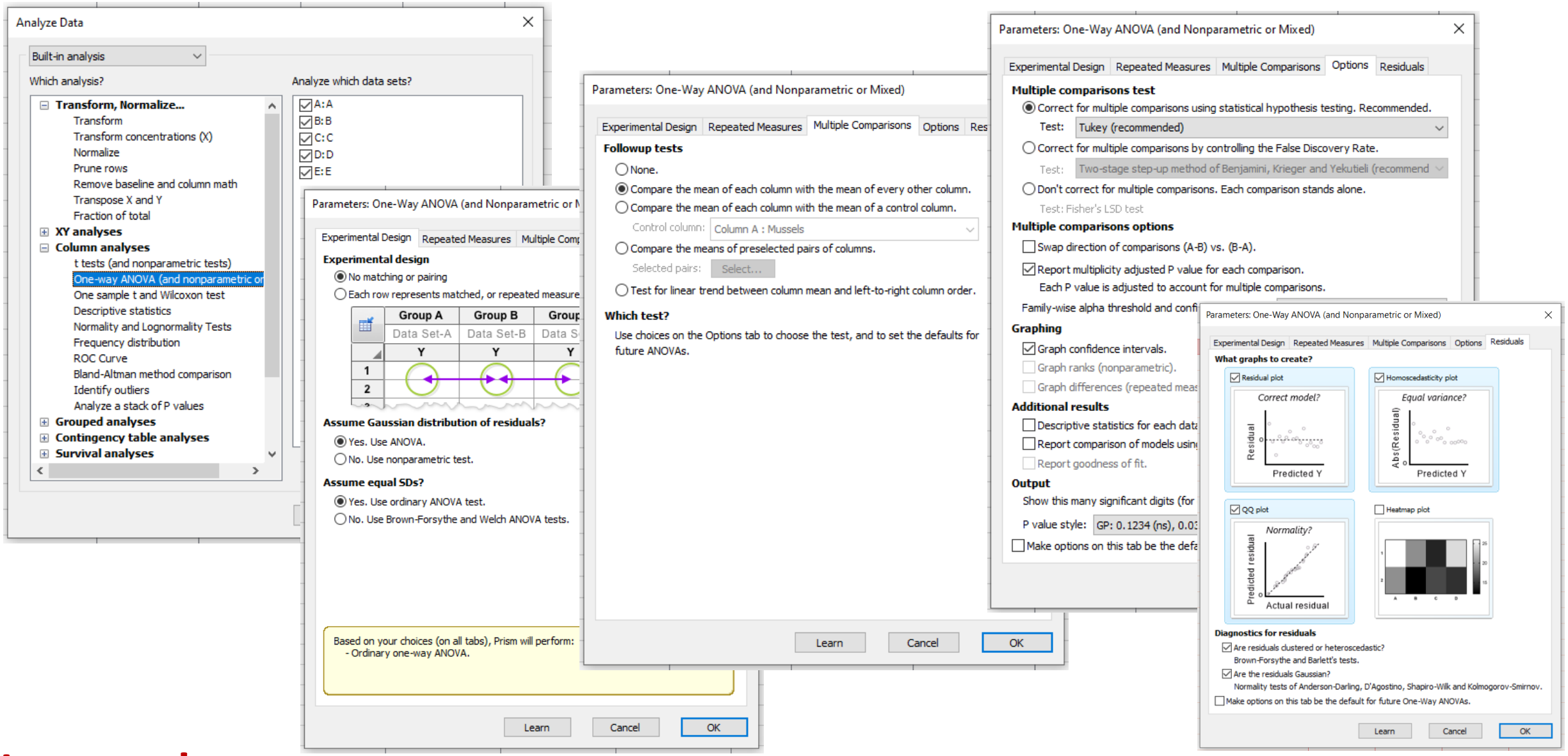
# Parametric tests assumptions

Tabular results

| Normality and Lognormality Tests<br>Tabular results | A<br>A | B<br>B | C<br>C | D<br>D | E<br>E |
|---|---|---|---|---|---|
| **Test for normal distribution** | | | | | |
| **Anderson-Darling test** | | | | | |
| A2* | 0.7849 | 0.3412 | 0.2086 | 0.1524 | 0.4727 |
| P value | 0.0295 | 0.4303 | 0.8386 | 0.9495 | 0.2138 |
| Passed normality test (alpha=0.05)? | No | Yes | Yes | Yes | Yes |
| P value summary | * | ns | ns | ns | ns |
| | | | | | |
| **D'Agostino & Pearson test** | | | | | |
| K2 | 2.037 | 0.6827 | 0.5884 | 0.8869 | 2.902 |
| P value | 0.3611 | 0.7108 | 0.7451 | 0.6418 | 0.2344 |
| Passed normality test (alpha=0.05)? | Yes | Yes | Yes | Yes | Yes |
| P value summary | ns | ns | ns | ns | ns |
| | | | | | |
| **Shapiro-Wilk test** | | | | | |
| W | 0.8553 | 0.9458 | 0.9657 | 0.9868 | 0.9313 |
| P value | 0.0427 | 0.5773 | 0.7142 | 0.9935 | 0.2050 |
| Passed normality test (alpha=0.05)? | No | Yes | Yes | Yes | Yes |
| P value summary | * | ns | ns | ns | ns |
| | | | | | |
| **Kolmogorov-Smirnov test** | | | | | |
| KS distance | 0.2278 | 0.2049 | 0.1373 | 0.1016 | 0.1646 |
| P value | 0.0857 | >0.1000 | >0.1000 | >0.1000 | >0.1000 |
| Passed normality test (alpha=0.05)? | Yes | Yes | Yes | Yes | Yes |
| P value summary | ns | ns | ns | ns | ns |
| | | | | | |
| **Number of values** | 12 | 12 | 18 | 18 | 18 |



# One-Way ANOVA in Prism




**Have a go!**

# Analysis of variance: Omnibus test

ANOVA results | Multiple comparisons

**Ordinary one-way ANOVA**
ANOVA results

| | | |
|---|---|---|
| 1 | Table Analyzed | Transform of Data 1 |
| 2 | Data sets analyzed | A-E |
| 3 | | |
| 4 | **ANOVA summary** | |
| 5 | F | 8.123 |
| 6 | P value | <0.0001 |
| 7 | P value summary | **** |
| 8 | Significant diff. among means (P < 0.05)? | Yes |
| 9 | R squared | 0.3080 |
| 10 | | |
| 11 | **Brown-Forsythe test** | |
| 12 | F (DFn, DFd) | 0.9821 (4, 73) |
| 13 | P value | 0.4227 |
| 14 | P value summary | ns |
| 15 | Are SDs significantly different (P < 0.05)? | No |
| 16 | | |
| 17 | **Bartlett's test** | |
| 18 | Bartlett's statistic (corrected) | 5.826 |
| 19 | P value | 0.2125 |
| 20 | P value summary | ns |
| 21 | Are SDs significantly different (P < 0.05)? | No |
| 22 | | |

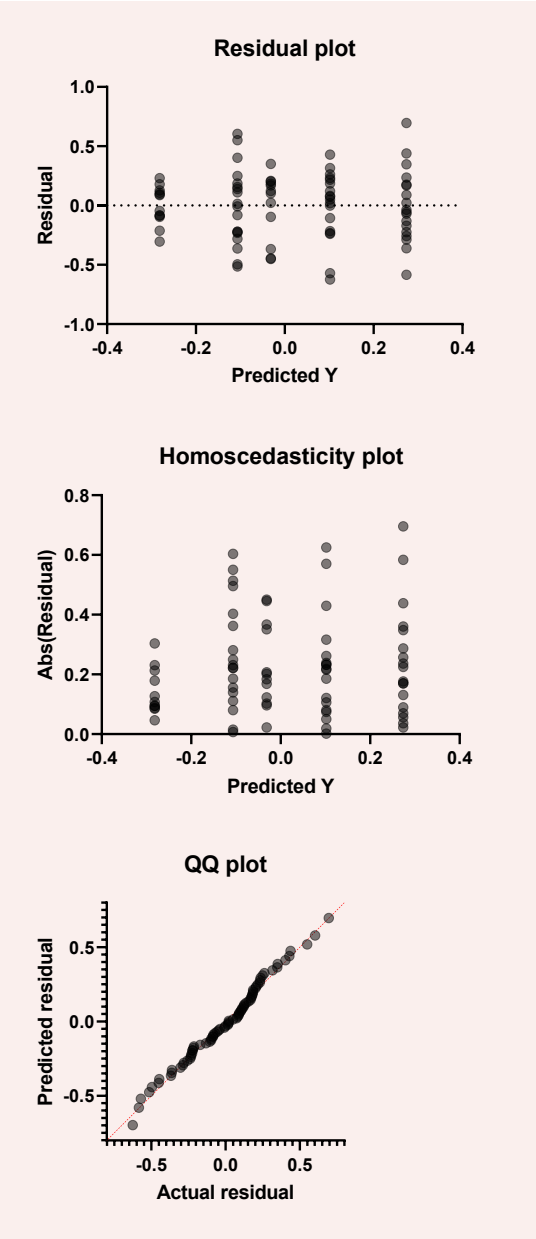
**Homogeneity of variance ☑**

**F=0.6728/0.08282=8.12**

| 23 | **ANOVA table** | **SS** | **DF** | **MS** | **F (DFn, DFd)** | **P value** |
|---|---|---|---|---|---|---|
| 24 | Treatment (between columns) | 2.691 | 4 | 0.6728 | F (4, 73) = 8.123 | P<0.0001 |
| 25 | Residual (within columns) | 6.046 | 73 | 0.08282 | | |
| 26 | Total | 8.737 | 77 | | | |

| 27 | | | | | |
|---|---|---|---|---|---|
| 28 | **Normality of Residuals** | | | | |
| 29 | **Test name** | **Statistics** | **P value** | **Passed normality test (alpha=0.05)?** | **P value summary** |
| 30 | Anderson-Darling (A2*) | 0.4214 | 0.3156 | Yes | ns |
| 31 | D'Agostino-Pearson omnibus (K2) | 0.1729 | 0.9172 | Yes | ns |
| 32 | Shapiro-Wilk (W) | 0.9862 | 0.5658 | Yes | ns |
| 33 | Kolmogorov-Smirnov (distance) | 0.07886 | 0.1000 | Yes | ns |

**Normality ☑**



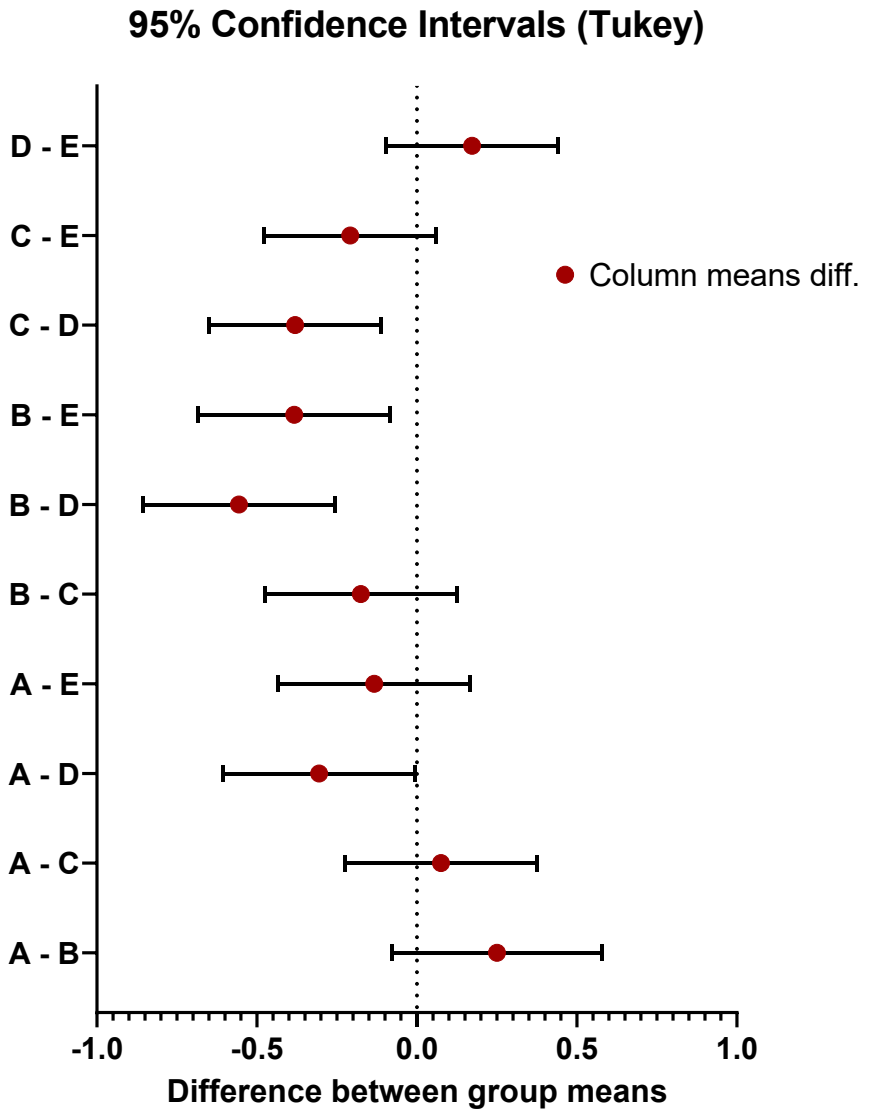
# Analysis of variance: Post Hoc tests

## Multiple comparisons: Tukey's correction

ANOVA results | Multiple comparisons

**Ordinary one-way ANOVA**
Multiple comparisons

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Number of families | 1 | | | | | | | |
| 2 | Number of comparisons per family | 10 | | | | | | | |
| 3 | Alpha | 0.05 | | | | | | | |
| 4 | | | | | | | | | |
| 5 | **Tukey's multiple comparisons test** | **Mean Diff.** | **95.00% CI of diff.** | **Below threshold?** | **Summary** | **Adjusted P Value** | | | |
| 6 | A vs. B | 0.2502 | -0.07839 to 0.5789 | No | ns | 0.2187 | A-B | | |
| 7 | A vs. C | 0.07500 | -0.2250 to 0.3750 | No | ns | 0.9560 | A-C | | |
| 8 | A vs. D | -0.3055 | -0.6055 to -0.005493 | Yes | * | 0.0439 | A-D | | |
| 9 | A vs. E | -0.1333 | -0.4333 to 0.1667 | No | ns | 0.7266 | A-E | | |
| 10 | B vs. C | -0.1753 | -0.4753 to 0.1247 | No | ns | 0.4809 | B-C | | |
| 11 | B vs. D | -0.5557 | -0.8557 to -0.2557 | Yes | **** | <0.0001 | B-D | | |
| 12 | B vs. E | -0.3835 | -0.6835 to -0.08352 | Yes | ** | 0.0055 | B-E | | |
| 13 | C vs. D | -0.3805 | -0.6488 to -0.1122 | Yes | ** | 0.0015 | C-D | | |
| 14 | C vs. E | -0.2083 | -0.4766 to 0.06006 | No | ns | 0.2023 | C-E | | |
| 15 | D vs. E | 0.1722 | -0.09611 to 0.4405 | No | ns | 0.3842 | D-E | | |
| 16 | | | | | | | | | |
| 17 | **Test details** | **Mean 1** | **Mean 2** | **Mean Diff.** | **SE of diff.** | **n1** | **n2** | **q** | **DF** |
| 18 | A vs. B | -0.03144 | -0.2817 | 0.2502 | 0.1175 | 12 | 12 | 3.012 | 73 |
| 19 | A vs. C | -0.03144 | -0.1064 | 0.07500 | 0.1073 | 12 | 18 | 0.9889 | 73 |
| 20 | A vs. D | -0.03144 | 0.2740 | -0.3055 | 0.1073 | 12 | 18 | 4.028 | 73 |
| 21 | A vs. E | -0.03144 | 0.1018 | -0.1333 | 0.1073 | 12 | 18 | 1.757 | 73 |
| 22 | B vs. C | -0.2817 | -0.1064 | -0.1753 | 0.1073 | 12 | 18 | 2.311 | 73 |
| 23 | B vs. D | -0.2817 | 0.2740 | -0.5557 | 0.1073 | 12 | 18 | 7.328 | 73 |
| 24 | B vs. E | -0.2817 | 0.1018 | -0.3835 | 0.1073 | 12 | 18 | 5.057 | 73 |
| 25 | C vs. D | -0.1064 | 0.2740 | -0.3805 | 0.09593 | 18 | 18 | 5.609 | 73 |
| 26 | C vs. E | -0.1064 | 0.1018 | -0.2083 | 0.09593 | 18 | 18 | 3.070 | 73 |
| 27 | D vs. E | 0.2740 | 0.1018 | 0.1722 | 0.09593 | 18 | 18 | 2.539 | 73 |



# Analysis of variance: Post Hoc tests

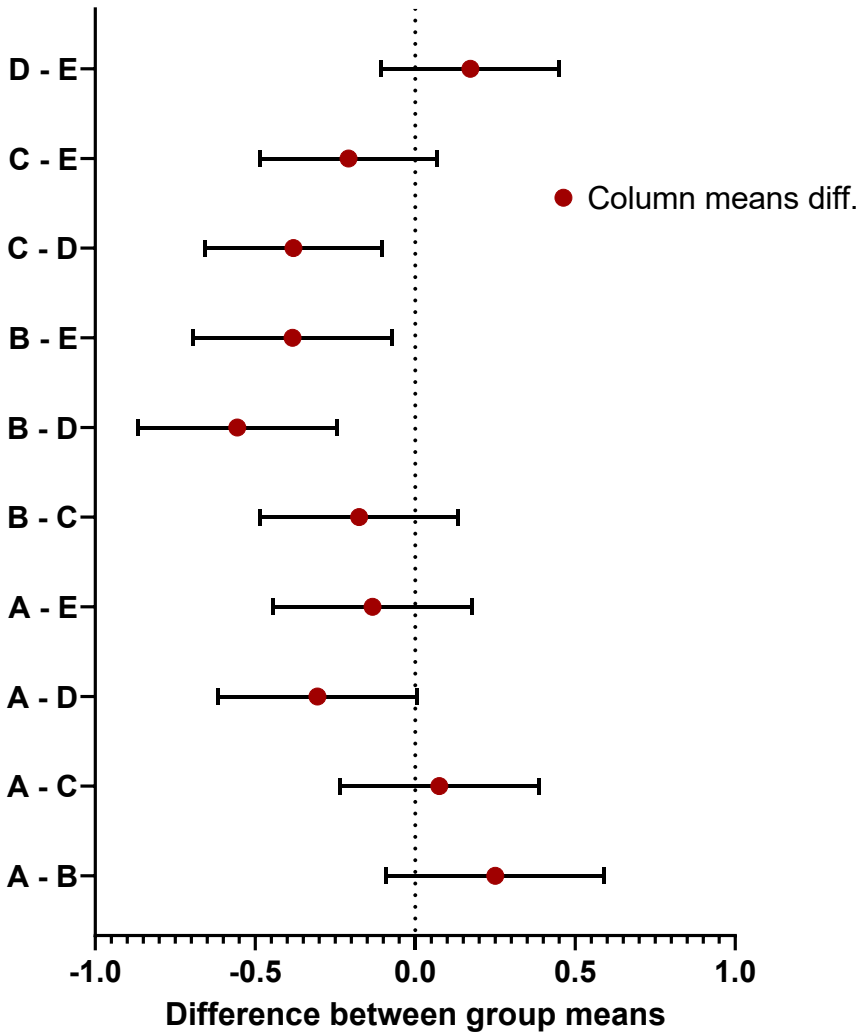
## Multiple comparisons: Bonferroni's correction

ANOVA results | Multiple comparisons

**Ordinary one-way ANOVA**
Multiple comparisons

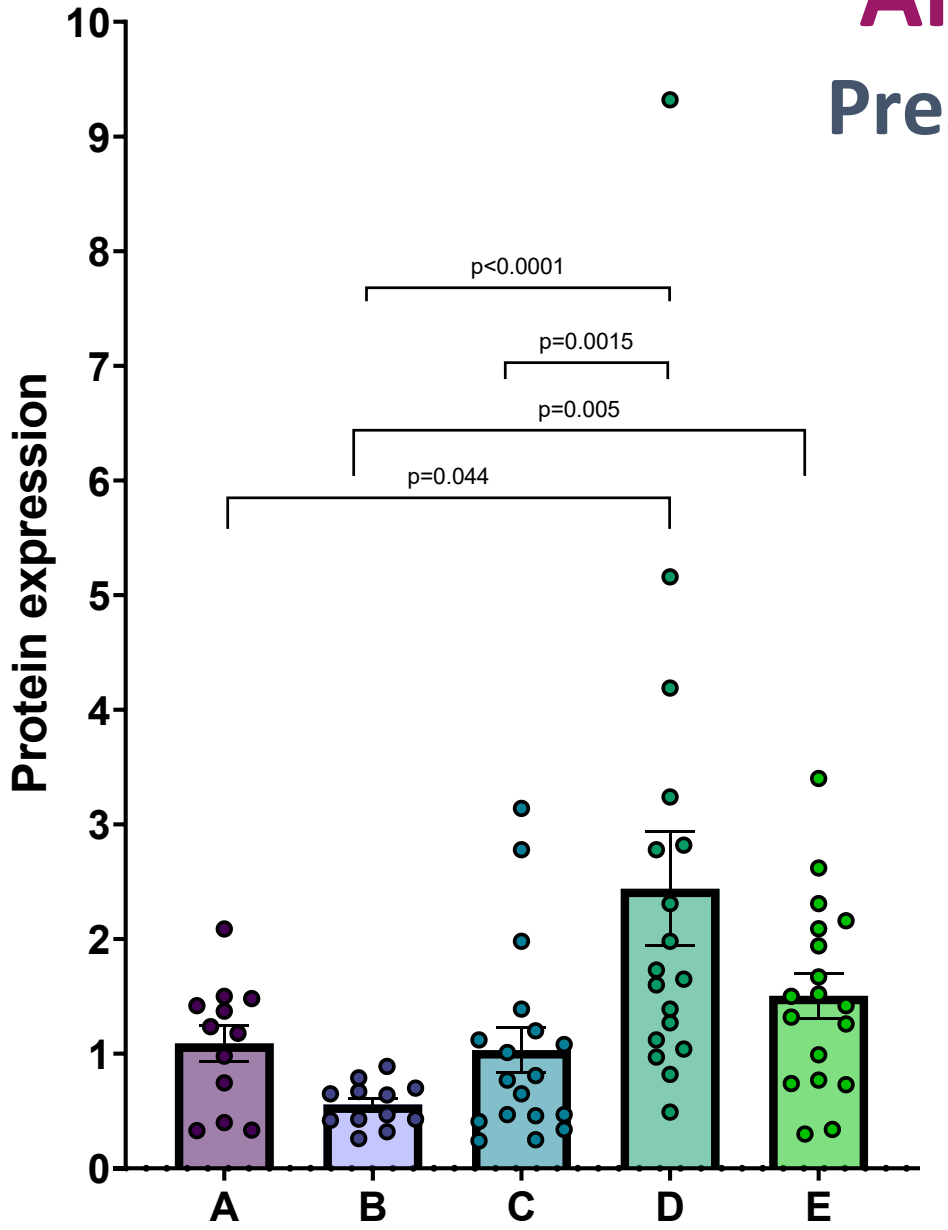
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Number of families | 1 | | | | | | | |
| 2 | Number of comparisons per family | 10 | | | | | | | |
| 3 | Alpha | 0.05 | | | | | | | |
| 4 | | | | | | | | | |
| 5 | **Bonferroni's multiple comparisons test** | **Mean Diff.** | **95.00% CI of diff.** | **Below threshold?** | **Summary** | **Adjusted P Value** | | | |
| 6 | A vs. B | 0.2502 | -0.08987 to 0.5904 | No | ns | 0.3655 | A-B | | |
| 7 | A vs. C | 0.07500 | -0.2355 to 0.3855 | No | ns | >0.9999 | A-C | | |
| 8 | A vs. D | -0.3055 | -0.6160 to 0.004988 | No | ns | 0.0571 | A-D | | |
| 9 | A vs. E | -0.1333 | -0.4438 to 0.1772 | No | ns | >0.9999 | A-E | | |
| 10 | B vs. C | -0.1753 | -0.4857 to 0.1352 | No | ns | >0.9999 | B-C | | |
| 11 | B vs. D | -0.5557 | -0.8662 to -0.2453 | Yes | **** | <0.0001 | B-D | | |
| 12 | B vs. E | -0.3835 | -0.6940 to -0.07304 | Yes | ** | 0.0062 | B-E | | |
| 13 | C vs. D | -0.3805 | -0.6582 to -0.1028 | Yes | ** | 0.0017 | C-D | | |
| 14 | C vs. E | -0.2083 | -0.4860 to 0.06943 | No | ns | 0.3318 | C-E | | |
| 15 | D vs. E | 0.1722 | -0.1055 to 0.4499 | No | ns | 0.7675 | D-E | | |
| 16 | | | | | | | | | |
| 17 | **Test details** | **Mean 1** | **Mean 2** | **Mean Diff.** | **SE of diff.** | **n1** | **n2** | **t** | **DF** |
| 18 | A vs. B | -0.03144 | -0.2817 | 0.2502 | 0.1175 | 12 | 12 | 2.130 | 73 |
| 19 | A vs. C | -0.03144 | -0.1064 | 0.07500 | 0.1073 | 12 | 18 | 0.6993 | 73 |
| 20 | A vs. D | -0.03144 | 0.2740 | -0.3055 | 0.1073 | 12 | 18 | 2.848 | 73 |
| 21 | A vs. E | -0.03144 | 0.1018 | -0.1333 | 0.1073 | 12 | 18 | 1.243 | 73 |
| 22 | B vs. C | -0.2817 | -0.1064 | -0.1753 | 0.1073 | 12 | 18 | 1.634 | 73 |
| 23 | B vs. D | -0.2817 | 0.2740 | -0.5557 | 0.1073 | 12 | 18 | 5.182 | 73 |
| 24 | B vs. E | -0.2817 | 0.1018 | -0.3835 | 0.1073 | 12 | 18 | 3.576 | 73 |
| 25 | C vs. D | -0.1064 | 0.2740 | -0.3805 | 0.09593 | 18 | 18 | 3.966 | 73 |
| 26 | C vs. E | -0.1064 | 0.1018 | -0.2083 | 0.09593 | 18 | 18 | 2.171 | 73 |
| 27 | D vs. E | 0.2740 | 0.1018 | 0.1722 | 0.09593 | 18 | 18 | 1.795 | 73 |

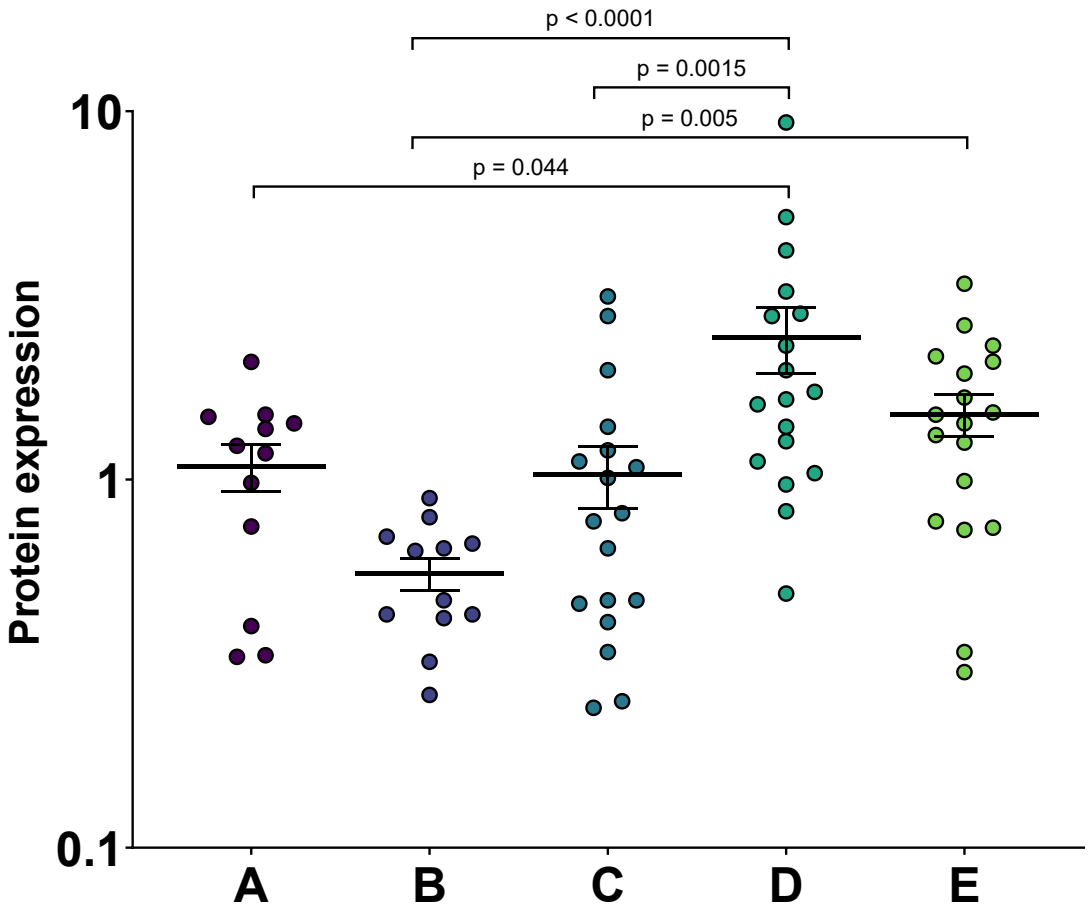
**95% Confidence Intervals (Bonferroni)**



# Analysis of variance

## Presentation of the results



# Exercise 3

Babraham Institute

# Analysis of Quantitative data
# Two-way ANOVA

Hayley Carr & Anne Segonds-Pichon

v2023-09

Babraham Bioinformatics

# Comparison between more than 2 groups

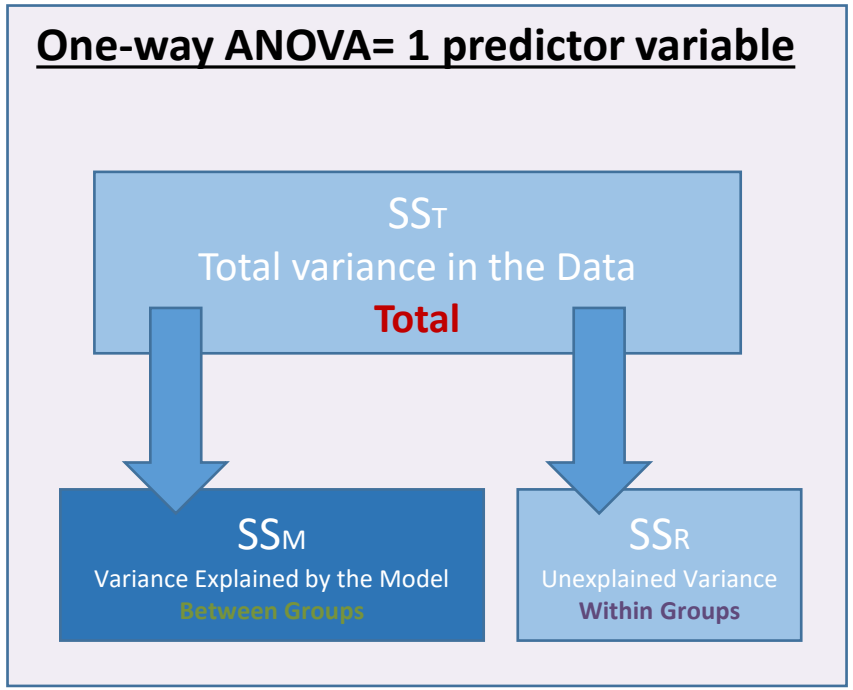
## Two factors = Two predictors

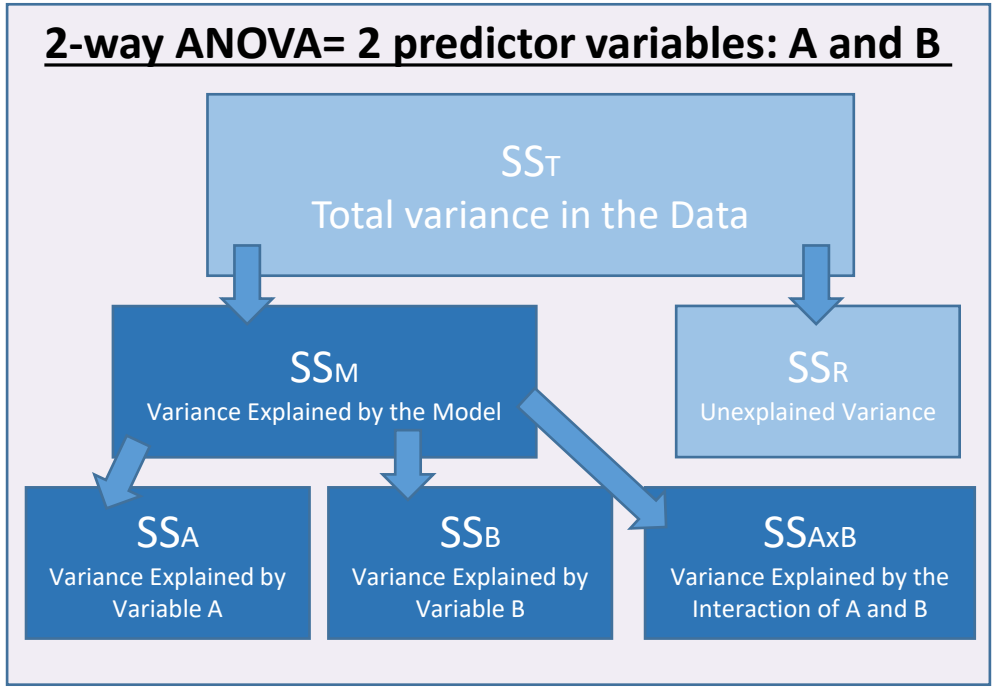
## Two-Way ANOVA

# Two-way Analysis of Variance (Factorial ANOVA)

| Source of variation | Sum of Squares | Df | Mean Square | F | p-value |
|---|---|---|---|---|---|
| Variable A (Between Groups) | 2.665 | 4 | 0.6663 | 8.42 | <0.0001 |
| Within Groups (Residual) | 5.775 | 73 | 0.0791 | | |
| Total | 8.44 | 77 | | | |

| Source of variation | Sum of Squares | Df | Mean Square | F | p-value |
|---|---|---|---|---|---|
| Variable A * Variable B | 1978 | 2 | 989.1 | F (2, 42) = 11.91 | P < 0.0001 |
| Variable B (Between groups) | 3332 | 2 | 1666 | F (2, 42) = 20.07 | P < 0.0001 |
| Variable A (Between groups) | 168.8 | 1 | 168.8 | F (1, 42) = 2.032 | P = 0.1614 |
| Residuals | 3488 | 42 | 83.04 | | |



# Two-way Analysis of Variance

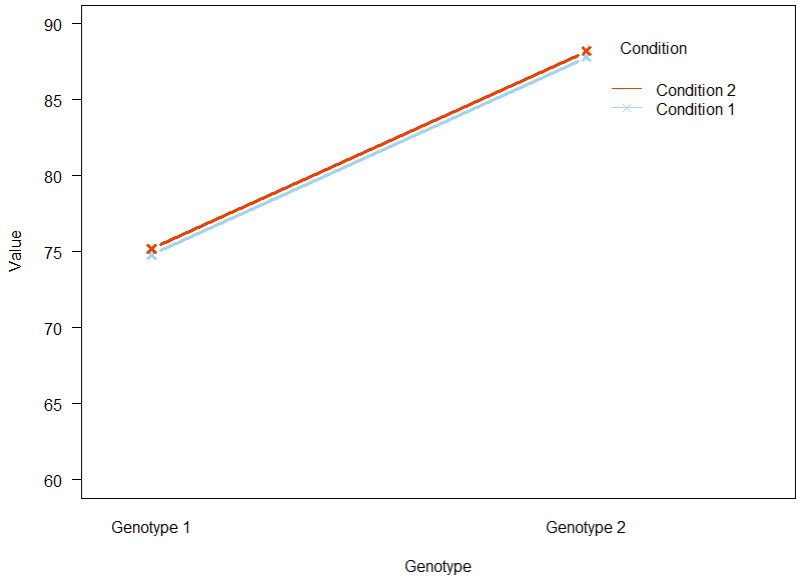
- **Interaction plots: Examples**
  - Fake dataset:
    - 2 factors: Genotype (2 levels) and Condition (2 levels)

| Genotype | Condition | Value |
|---|---|---|
| Genotype 1 | Condition 1 | 74.8 |
| Genotype 1 | Condition 1 | 65 |
| Genotype 1 | Condition 1 | 74.8 |
| Genotype 1 | Condition 2 | 75.2 |
| Genotype 1 | Condition 2 | 75 |
| Genotype 1 | Condition 2 | 75.2 |
| Genotype 2 | Condition 1 | 87.8 |
| Genotype 2 | Condition 1 | 65 |
| Genotype 2 | Condition 1 | 74.8 |
| Genotype 2 | Condition 2 | 88.2 |
| Genotype 2 | Condition 2 | 75 |
| Genotype 2 | Condition 2 | 75.2 |

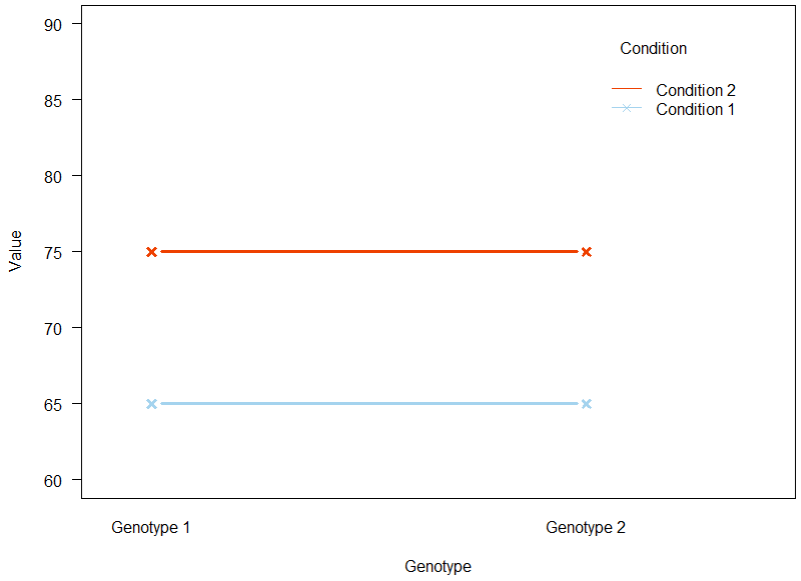
# Two-way Analysis of Variance

- **Interaction plots: Examples**
  - 2 factors: Genotype (2 levels) and Condition (2 levels)

## Single Effect




Genotype Effect

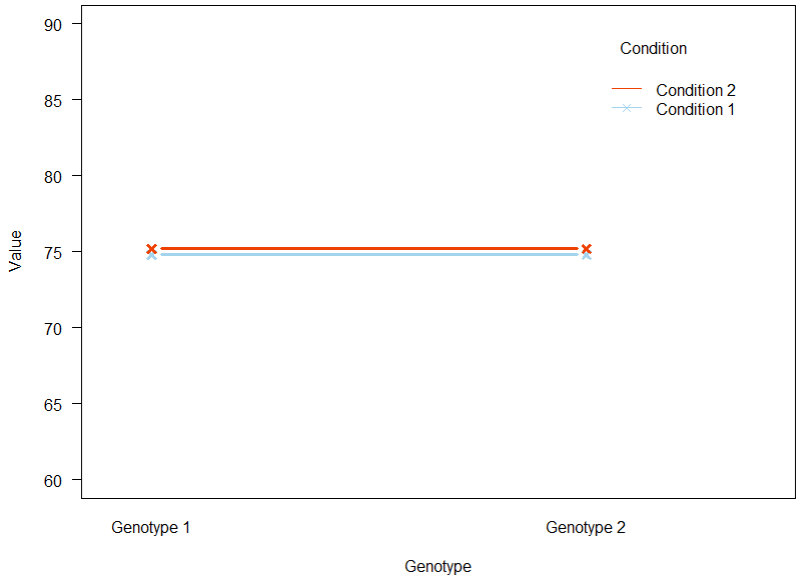



Condition Effect

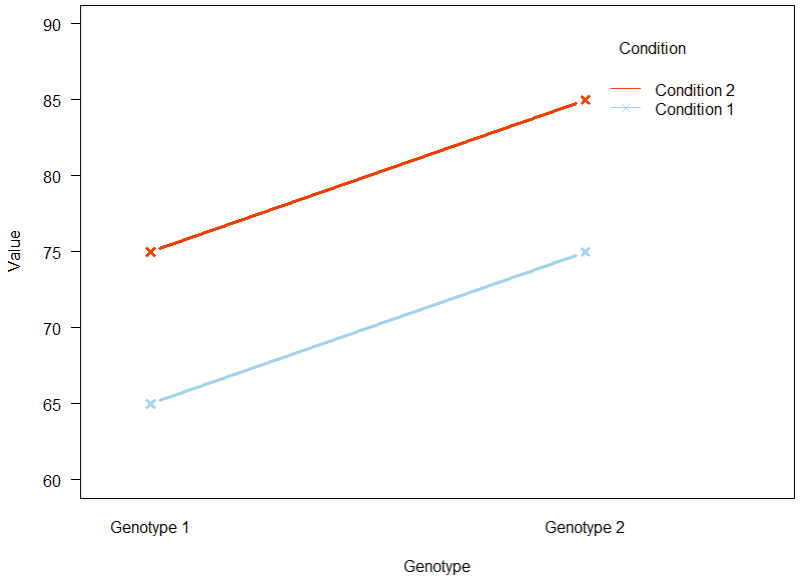
# Two-way Analysis of Variance

- **Interaction plots: Examples**
  - 2 factors: Genotype (2 levels) and Condition (2 levels)

## Zero or Both Effect




Zero Effect

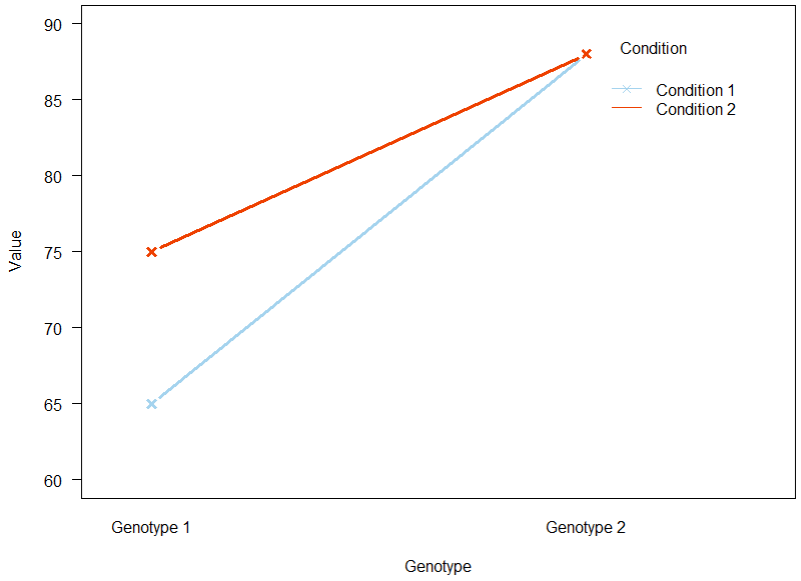


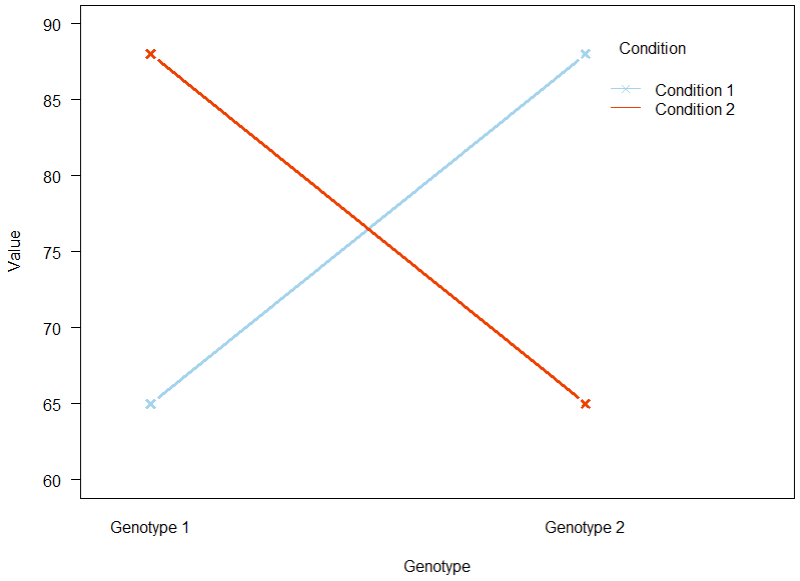

Both Effect

# Two-way Analysis of Variance

- **Interaction plots: Examples**
  - 2 factors: Genotype (2 levels) and Condition (2 levels)

## Interaction



# Two-way Analysis of Variance

**Example: crop.density.csv**

- Want to know if planting density (1=low density, 2=high density) or fertiliser type (1, 2, or 3) have an impact on crop yield

- Three null hypotheses:
  - No difference in yield for any fertiliser type
  - No difference in yield for either planting density
  - Effect of fertiliser type or density on yield does not depend on the effect of the other variable

https://www.scribbr.com/statistics/two-way-anova/

# Two-way Analysis of Variance

## Excel to Prism

| | A | B | C |
|---|---|---|---|
| 1 | density | fertilizer | yield |
| 2 | 1 | 1 | 177.2287 |
| 3 | 1 | 1 | 176.4085 |
| 4 | 1 | 1 | 177.1255 |
| 5 | 1 | 1 | 176.7463 |
| 6 | 1 | 1 | 176.2749 |
| 7 | 1 | 1 | 176.6013 |
| 8 | 1 | 1 | 177.4795 |
| 9 | 1 | 1 | 176.1144 |
| 10 | 1 | 1 | 176.1083 |
| 11 | 1 | 1 | 177.2624 |
| 12 | 1 | 1 | 176.239 |
| 13 | 1 | 1 | 176.0393 |
| 14 | 1 | 1 | 176.1606 |
| 15 | 1 | 1 | 175.9385 |
| 16 | 1 | 1 | 175.3608 |
| 17 | 1 | 1 | 175.9454 |
| 18 | 2 | 1 | 177.55 |
| 19 | 2 | 1 | 177.7036 |
| 20 | 2 | 1 | 176.7783 |
| 21 | 2 | 1 | 177.0612 |
| 22 | 2 | 1 | 177.9672 |
| 23 | 2 | 1 | 177.0305 |
| 24 | 2 | 1 | 176.8741 |
| 25 | 2 | 1 | 176.0084 |
| 26 | 2 | 1 | 178.3574 |
| 27 | 2 | 1 | 176.9188 |
| 28 | 2 | 1 | 176.5731 |
| 29 | 2 | 1 | 176.8179 |

Table format: **Grouped**

| | | Group A | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Fertiliser1 | | | | | | | | | | | | | |
| | | A:1 | A:2 | A:3 | A:4 | A:5 | A:6 | A:7 | A:8 | A:9 | A:10 | A:11 | A:12 | A:13 | A: |
| 1 | Density1 | 177.2287 | 176.4085 | 177.1255 | 176.7463 | 176.2749 | 176.6013 | 177.4795 | 176.1144 | 176.1083 | 177.2624 | 176.2390 | 176.0393 | 176.1606 | 175 |
| 2 | Density2 | 177.5500 | 177.7036 | 176.7783 | 177.0612 | 177.9672 | 177.0305 | 176.8741 | 176.0084 | 178.3574 | 176.9188 | 176.5731 | 176.8179 | 177.2264 | 177 |

Table format: **Grouped**

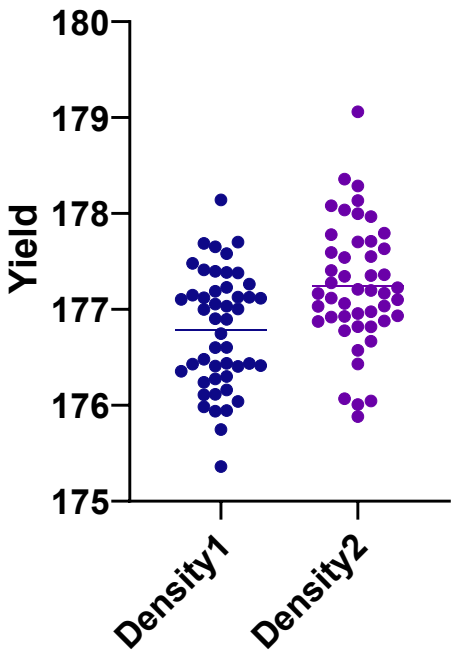
Group A — Fertiliser1

| | | A:1 | A:2 | A:3 | A:4 | A:5 | A:6 | A:7 | A:8 | A:9 | A:10 | A:11 | A:12 | A:13 | A: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Density1 | | | | | | | | | | | | | | |
| 2 | Density2 | | | | | | | | | | | | | | |
| 3 | Title | | | | | | | | | | | | | | |
| 4 | Title | | | | | | | | | | | | | | |
| 5 | Title | | | | | | | | | | | | | | |
| 6 | Title | | | | | | | | | | | | | | |
| 7 | Title | | | | | | | | | | | | | | |
| 8 | Title | | | | | | | | | | | | | | |
| 9 | Title | | | | | | | | | | | | | | |
| 10 | Title | | | | | | | | | | | | | | |
| 11 | Title | | | | | | | | | | | | | | |
| 12 | Title | | | | | | | | | | | | | | |
| 13 | Title | | | | | | | | | | | | | | |
| 14 | Title | | | | | | | | | | | | | | |
| 15 | Title | | | | | | | | | | | | | | |
| 16 | Title | | | | | | | | | | | | | | |
| 17 | Title | | | | | | | | | | | | | | |
| 18 | Title | | | | | | | | | | | | | | |
| 19 | Title | | | | | | | | | | | | | | |
| 20 | Title | | | | | | | | | | | | | | |
| 21 | Title | | | | | | | | | | | | | | |
| 22 | Title | | | | | | | | | | | | | | |
| 23 | Title | | | | | | | | | | | | | | |
| 24 | Title | | | | | | | | | | | | | | |
| 25 | Title | | | | | | | | | | | | | | |

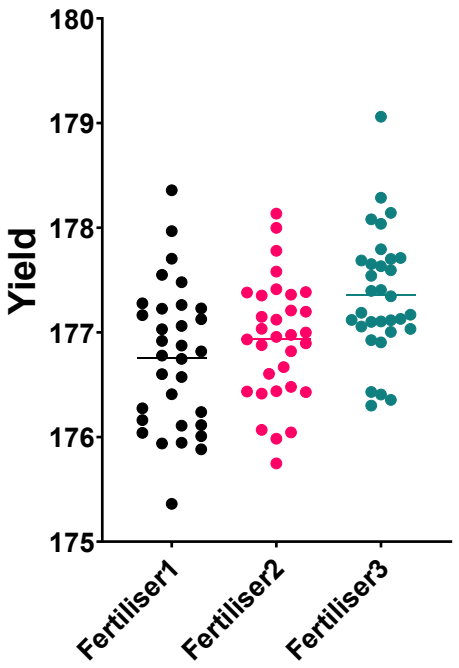
- Cut Ctrl+X
- Copy Ctrl+C
- Paste ›
- Paste Transpose ›
  - Paste Data Ctrl+Shift+T
  - Paste Embed
  - Paste Link
- Paste Special... Ctrl+Shift+V
- Import Data...
- Insert...
- Delete...
- Exclude Values Ctrl+E
- Format Points ›
- Cell Background Color ›
- Decimal Format...
- Column Width
- One More Subcolumn
- Data Object ›
- Select ›
- Use Larger Font Ctrl+=
- Use Smaller Font Ctrl+-

# Two-way Analysis of Variance

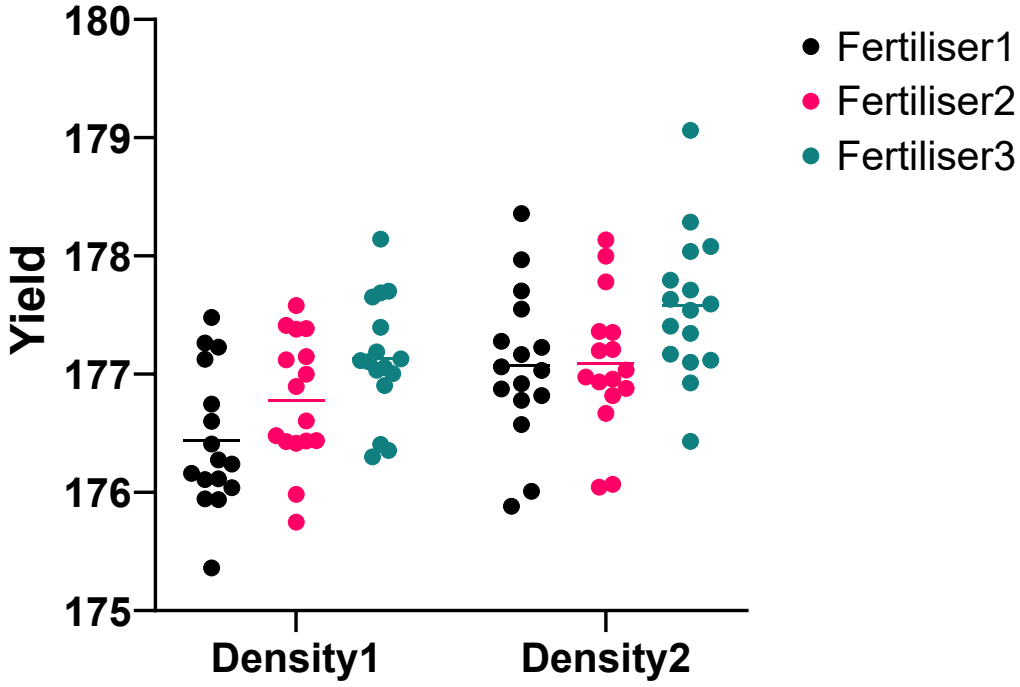
Main effect of Density




Main effect of Fertiliser

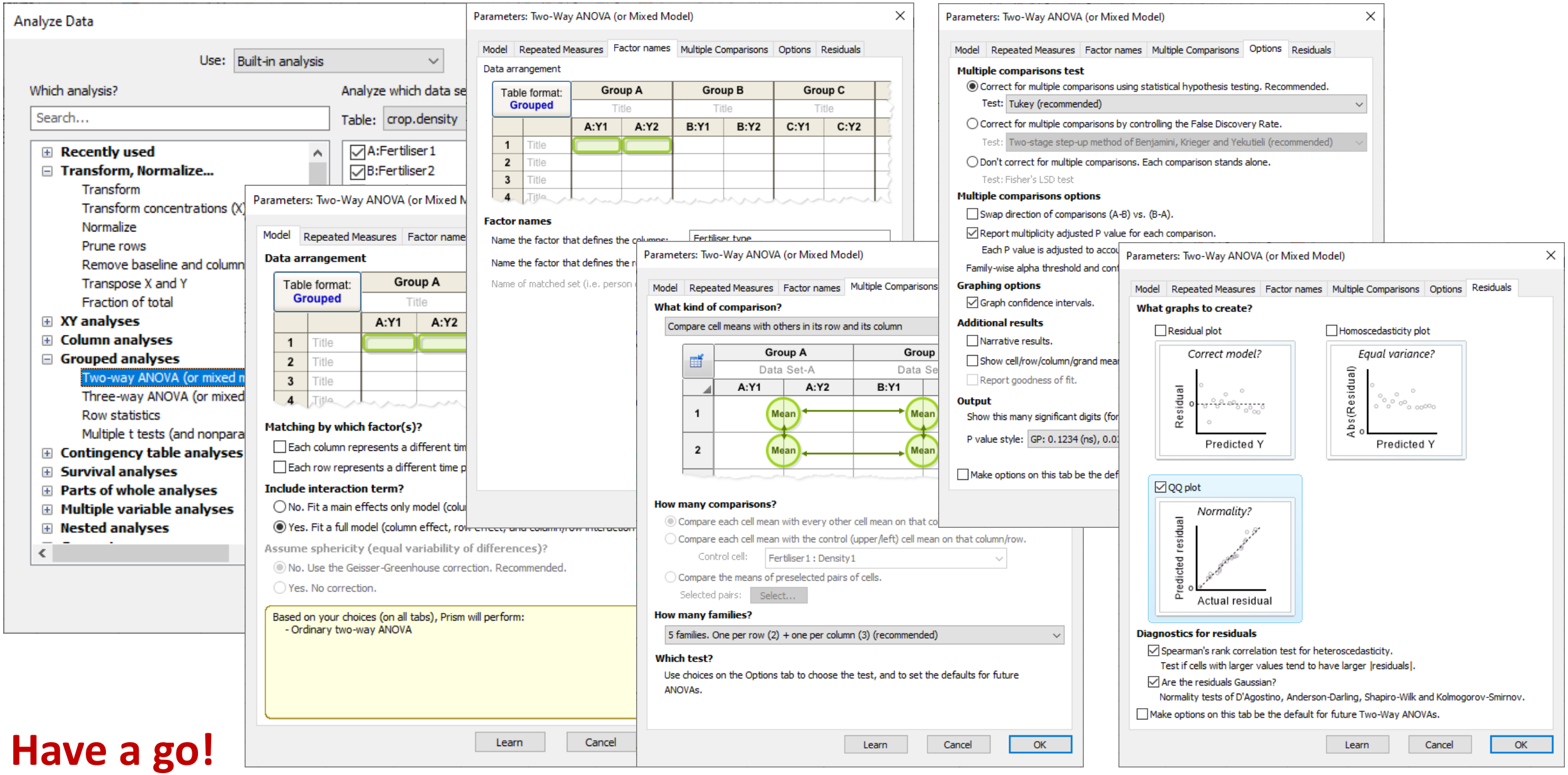



Interaction
between Density and Fertiliser




* Need data in Column format for these graphs

# Two-way Analysis of Variance




**Have a go!**

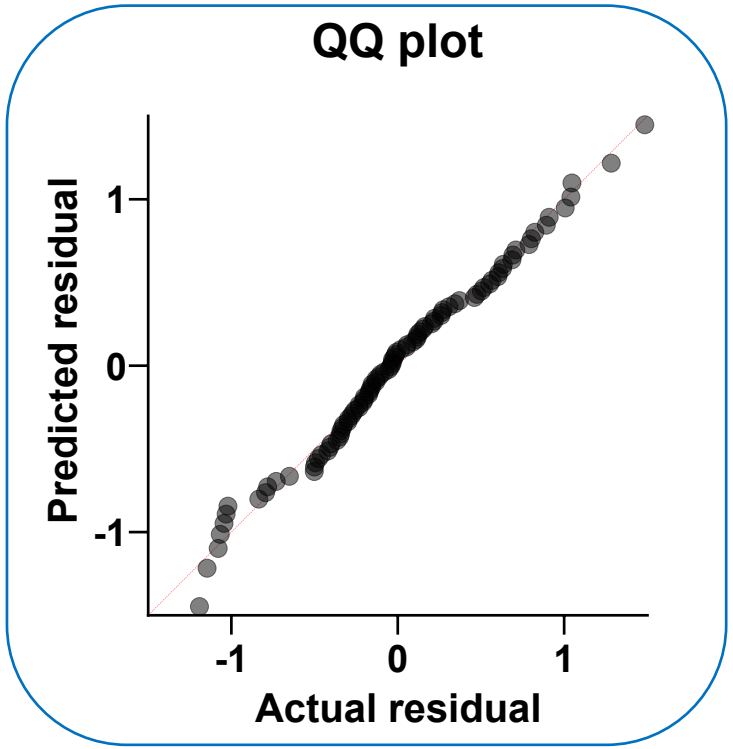
# Two-way Analysis of Variance

ANOVA results | Multiple comparisons

**2way ANOVA**
ANOVA results

| | | | | | |
|---|---|---|---|---|---|
| 1 | Table Analyzed | crop.density | | | | |
| 2 | | | | | | |
| 3 | **Two-way ANOVA** | Ordinary | | | | |
| 4 | Alpha | 0.05 | | | | |
| 5 | | | | | | |
| 6 | **Source of Variation** | **% of total variation** | **P value** | **P value summary** | **Significant?** | |
| 7 | Interaction | 1.020 | 0.5325 | ns | No | |
| 8 | Density | 12.21 | 0.0002 | *** | Yes | |
| 9 | Fertiliser type | 14.46 | 0.0003 | *** | Yes | |
| 10 | | | | | | |
| 11 | **ANOVA table** | **SS** | **DF** | **MS** | **F (DFn, DFd)** | **P value** |
| 12 | Interaction | 0.4278 | 2 | 0.2139 | F (2, 90) = 0.6346 | P=0.5325 |
| 13 | Density | 5.122 | 1 | 5.122 | F (1, 90) = 15.19 | P=0.0002 |
| 14 | Fertiliser type | 6.068 | 2 | 3.034 | F (2, 90) = 9.001 | P=0.0003 |
| 15 | Residual | 30.34 | 90 | 0.3371 | | |

| | | | | |
|---|---|---|---|---|
| **Spearman's test for heteroscedasticity** | | | | |
| Rs of predicted Y vs. \|residual\| | -0.1250 | | | |
| P value (one tailed) | 0.1124 | | | |
| Passed (P > 0.05)? | Yes | | | |
| | | | | |
| **Normality of Residuals** | | | | |
| **Test name** | **Statistics** | **P value** | **Passed normality test** | **P value summary** |
| D'Agostino-Pearson omnibus (K2) | 0.3270 | 0.8492 | Yes | ns |
| Anderson-Darling (A2*) | 0.4851 | 0.2221 | Yes | ns |
| Shapiro-Wilk (W) | 0.9853 | 0.3600 | Yes | ns |
| Kolmogorov-Smirnov (distance) | 0.06758 | 0.1000 | Yes | ns |

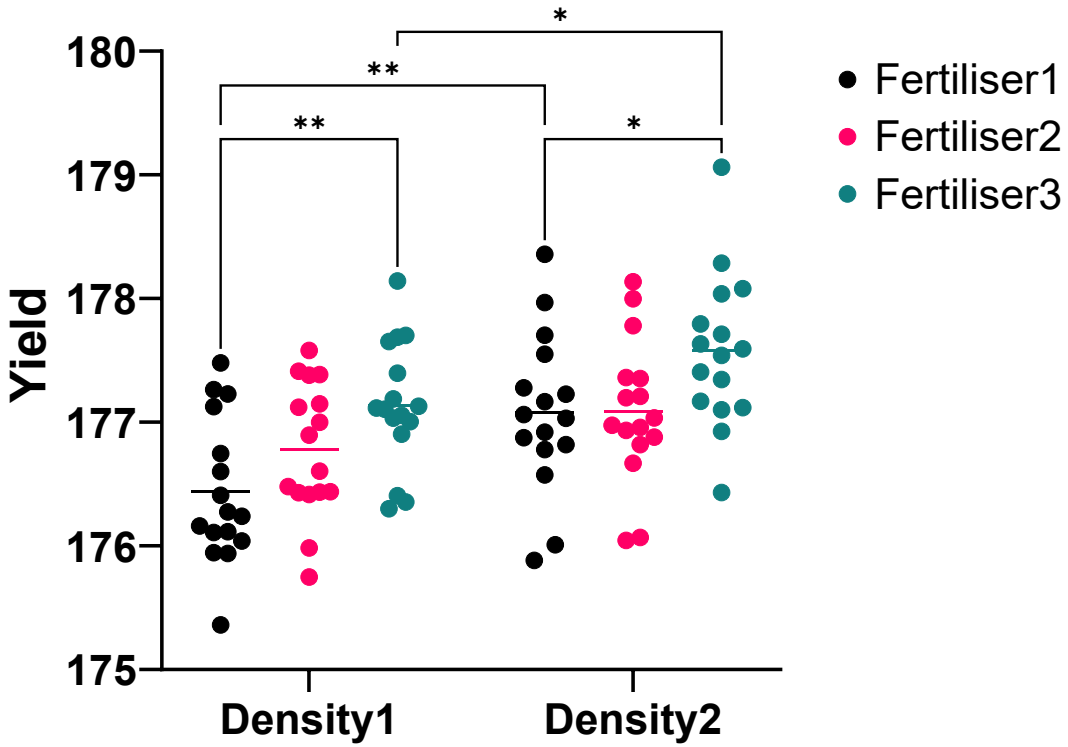


# Two-way Analysis of Variance

**2way ANOVA**
ANOVA results

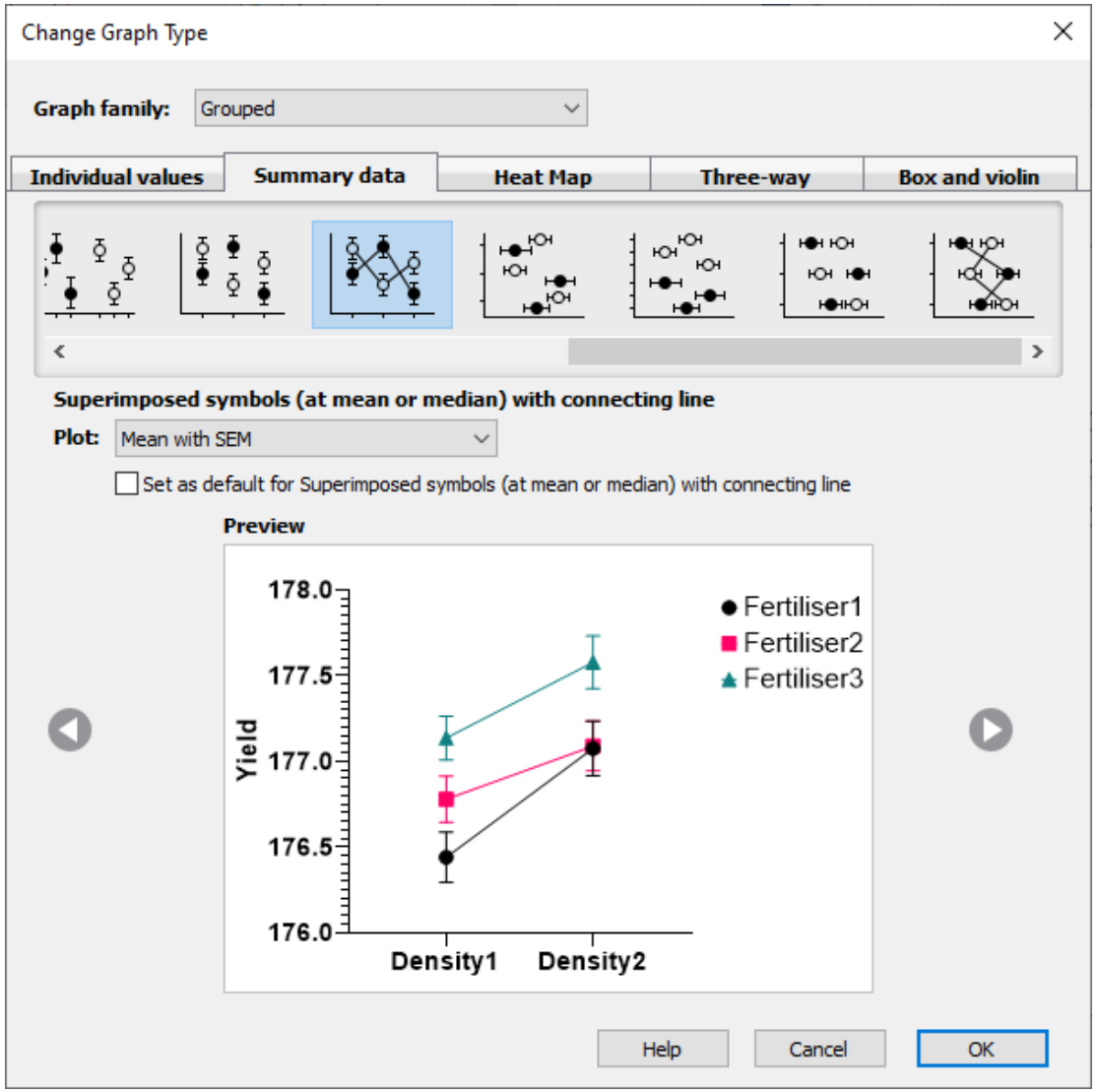
| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | Table Analyzed | crop.density | | | | |
| 2 | | | | | | |
| 3 | **Two-way ANOVA** | Ordinary | | | | |
| 4 | Alpha | 0.05 | | | | |
| 5 | | | | | | |
| 6 | **Source of Variation** | **% of total variation** | **P value** | **P value summary** | **Significant?** | |
| 7 | Interaction | 1.020 | 0.5325 | ns | No | |
| 8 | Density | 12.21 | 0.0002 | *** | Yes | |
| 9 | Fertiliser type | 14.46 | 0.0003 | *** | Yes | |
| 10 | | | | | | |
| 11 | **ANOVA table** | **SS** | **DF** | **MS** | **F (DFn, DFd)** | **P value** |
| 12 | Interaction | 0.4278 | 2 | 0.2139 | F (2, 90) = 0.6346 | P=0.5325 |
| 13 | Density | 5.122 | 1 | 5.122 | F (1, 90) = 15.19 | P=0.0002 |
| 14 | Fertiliser type | 6.068 | 2 | 3.034 | F (2, 90) = 9.001 | P=0.0003 |
| 15 | Residual | 30.34 | 90 | 0.3371 | | |

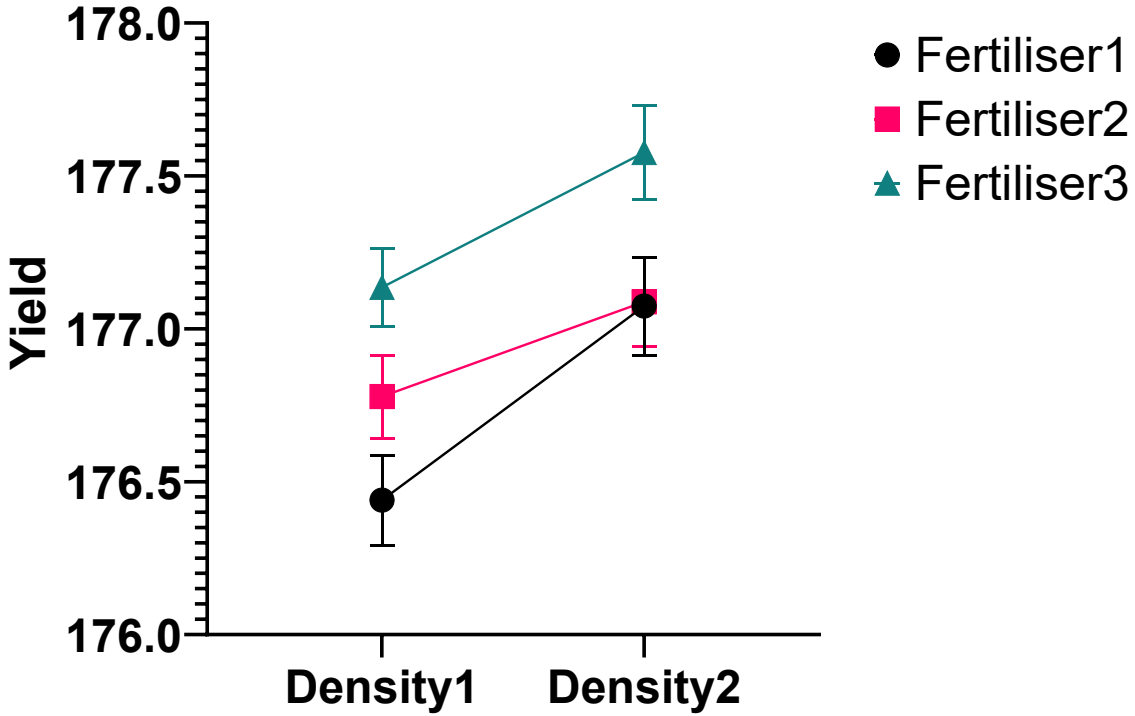
| Tukey's multiple comparisons test | Mean Diff. | 95.00% CI of diff. | Below threshold? | Summary | Adjusted P Value |
|---|---|---|---|---|---|
| Density1 | | | | | |
| Fertiliser1 vs. Fertiliser2 | -0.3387 | -0.8279 to 0.1505 | No | ns | 0.2302 |
| Fertiliser1 vs. Fertiliser3 | -0.6960 | -1.185 to -0.2068 | Yes | ** | 0.0030 |
| Fertiliser2 vs. Fertiliser3 | -0.3573 | -0.8465 to 0.1319 | No | ns | 0.1958 |
| Density2 | | | | | |
| Fertiliser1 vs. Fertiliser2 | -0.01366 | -0.5028 to 0.4755 | No | ns | 0.9976 |
| Fertiliser1 vs. Fertiliser3 | -0.5023 | -0.9914 to -0.01309 | Yes | * | 0.0428 |
| Fertiliser2 vs. Fertiliser3 | -0.4886 | -0.9778 to 0.0005727 | No | ns | 0.0503 |
| Fertiliser1 | | | | | |
| Density1 vs. Density2 | -0.6349 | -1.043 to -0.2271 | Yes | ** | 0.0026 |
| Fertiliser2 | | | | | |
| Density1 vs. Density2 | -0.3098 | -0.7176 to 0.09796 | No | ns | 0.1347 |
| Fertiliser3 | | | | | |
| Density1 vs. Density2 | -0.4411 | -0.8489 to -0.03334 | Yes | * | 0.0343 |



# Two-way Analysis of Variance

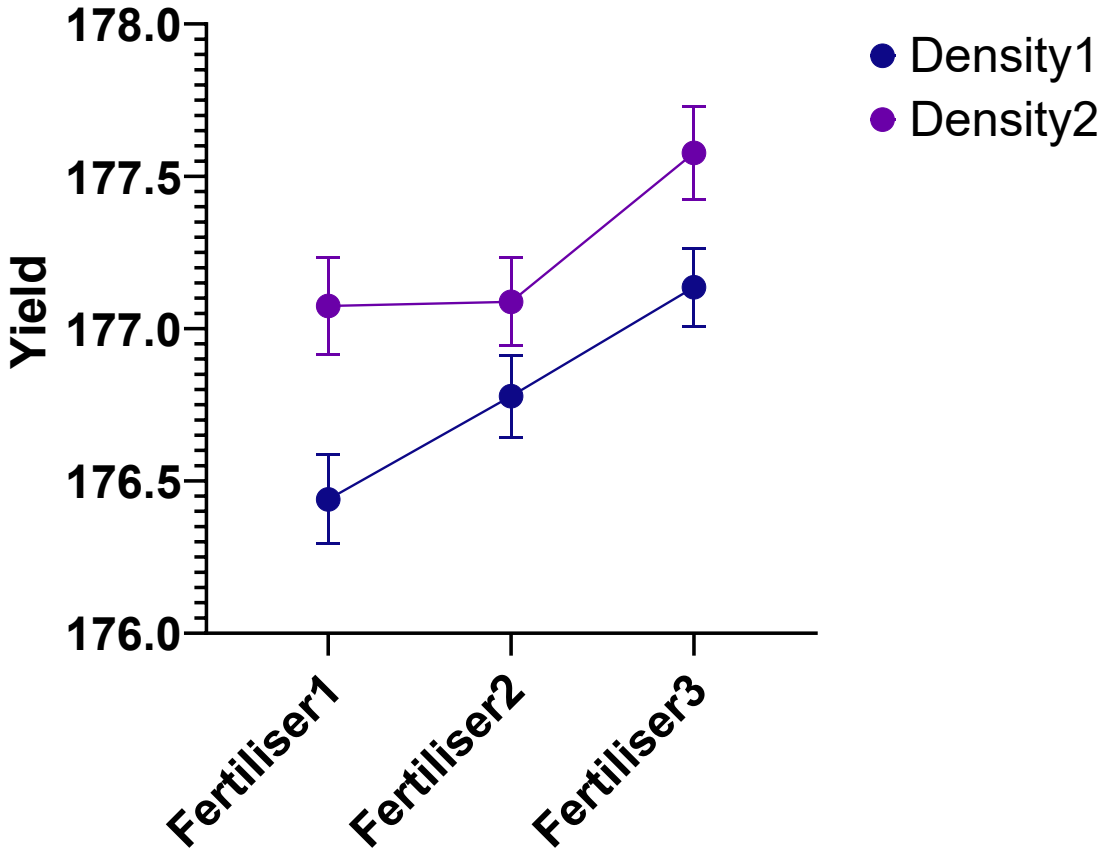
Quick way to have a look at the interaction

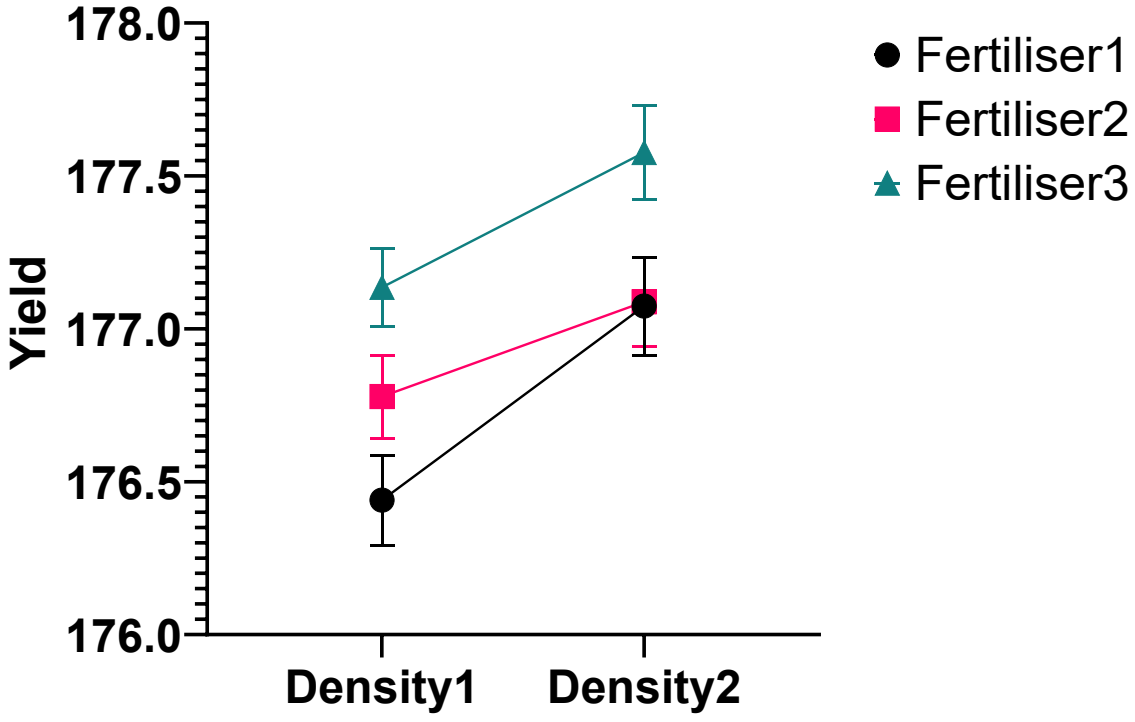


# Two-way Analysis of Variance

Quick way to have a look at the interaction



Babraham Institute

# Analysis of Quantitative data
# Linear and non-linear relationships

Hayley Carr & Anne Segonds-Pichon

v2024-06

Babraham Bioinformatics

# Association between 2 continuous variables

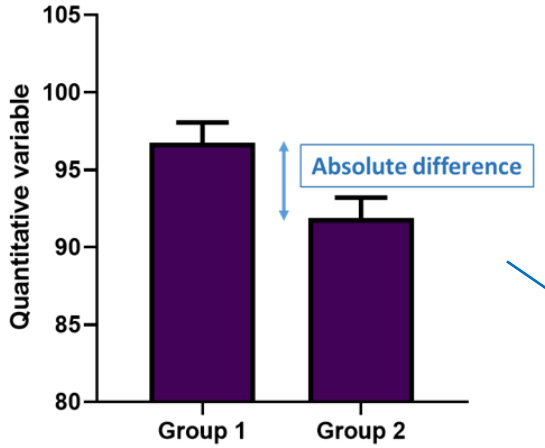
## One variable X and One variable Y

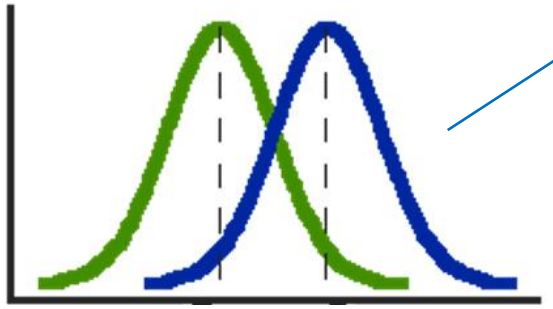
## One predictor

## Linear relationship

## Correlation and Regression

# Signal-to-noise ratio



$$\frac{\text{Difference (signal)}}{\text{Variation (noise)}}$$

$$\frac{\text{Signal}}{\text{Noise}} = \textbf{statistical significance}$$

$$\frac{\text{Signal}}{\text{Noise}} = \textbf{no statistical significance}$$

# Signal-to-noise ratio and Correlation

$$\frac{\text{Difference (signal)}}{\text{Variation (noise)}}$$

- For correlation, signal is **similarity** of behaviour between variable x and variable y
- **Coefficient of correlation**: $\mathbf{r} = \dfrac{\text{Similarity}}{\text{Variability}} = \dfrac{\text{Signal}}{\text{Noise}}$

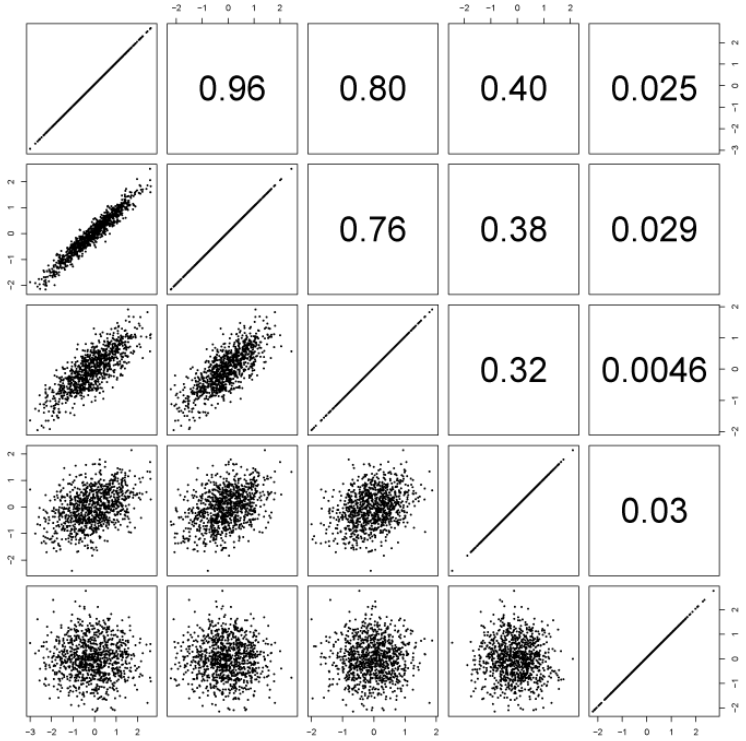
$$\mathbf{r} = \frac{\text{Similarity}}{\text{Variability}} = \frac{COV_{xy}}{SD_x SD_y}$$

**Covariance**

**Standard Deviation**

# Correlation

- Most widely-used correlation coefficient:
  - **Pearson product-moment correlation coefficient: r**
    - The **magnitude** and the **direction** of the relation between 2 variables
    - Designed to range in value between **-1 and +1**
    - Often look for **>|0.6|**
  - **Coefficient of determination: $r^2$**
    - Gives the proportion of variance in Y that can be explained by X
      - Helps with the interpretation of r
      - Basically the **effect size**




| Coefficient (+ve or –ve) | Strength of relationship |
|---|---|
| 0.0 to 0.2 | Negligible |
| 0.2 to 0.4 | Weak |
| 0.4 to 0.7 | Moderate |
| 0.7 to 0.9 | Strong |
| 0.9 to 1.0 | Very strong |

# Correlation

$p = 0.0002$

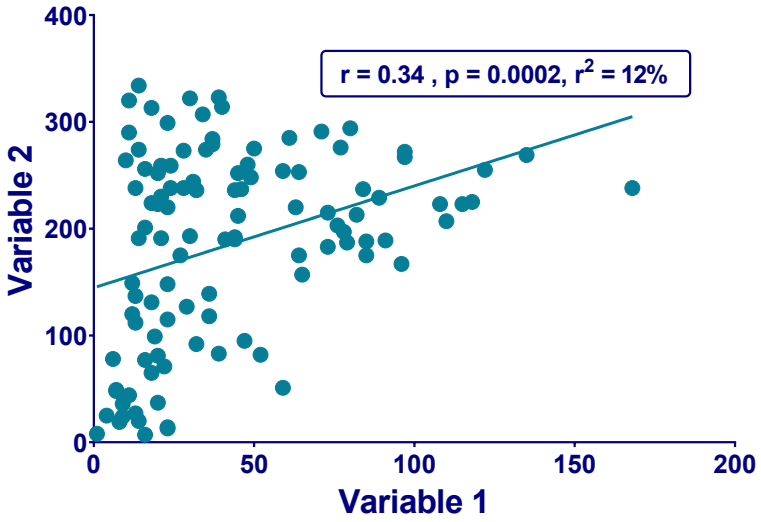
$r = 0.34$

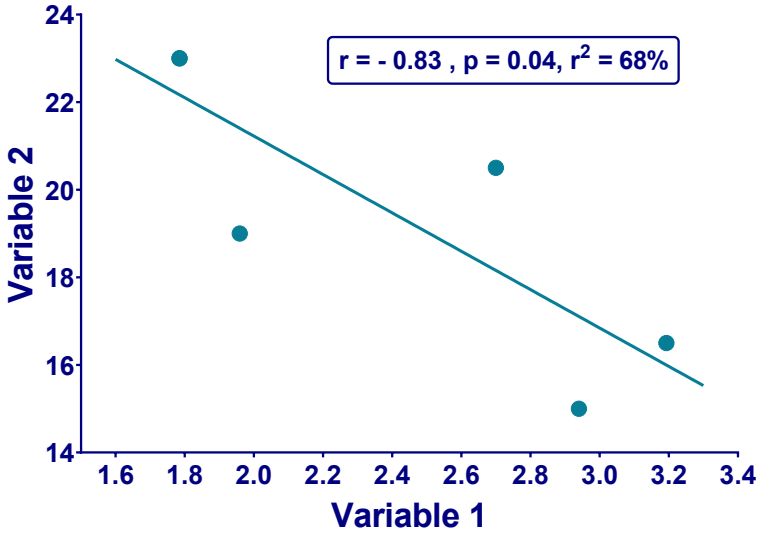
$r^2 = 0.12$

$p = 0.04$

$r = -0.83$

$r^2 = 0.68$




**Power!!**

# Coefficient of determination

| Unpaired t test | |
|---|---|
| Table Analyzed | coyotes |
| | |
| Column B | Males |
| vs. | vs. |
| Column A | Females |
| | |
| **Unpaired t test** | |
| P value | 0.1045 |
| P value summary | ns |
| Significantly different (P < 0.05)? | No |
| One- or two-tailed P value? | Two-tailed |
| t, df | t=1.641, df=84 |
| | |
| **How big is the difference?** | |
| Mean of column A | 89.71 |
| Mean of column B | 92.06 |
| Difference between means (B - A) ± SEM | 2.344 ± 1.428 |
| 95% confidence interval | -0.4964 to 5.185 |
| R squared (eta squared) | 0.03107 |

3%

## 5 cell lines on protein expression

| Ordinary one-way ANOVA<br>ANOVA results | |
|---|---|
| Table Analyzed | Transform of Protein expression |
| Data sets analyzed | A-E |
| | |
| **ANOVA summary** | |
| F | 8.127 |
| P value | <0.0001 |
| P value summary | **** |
| Significant diff. among means (P < 0.05)? | Yes |
| R square | 0.3081 |

31%

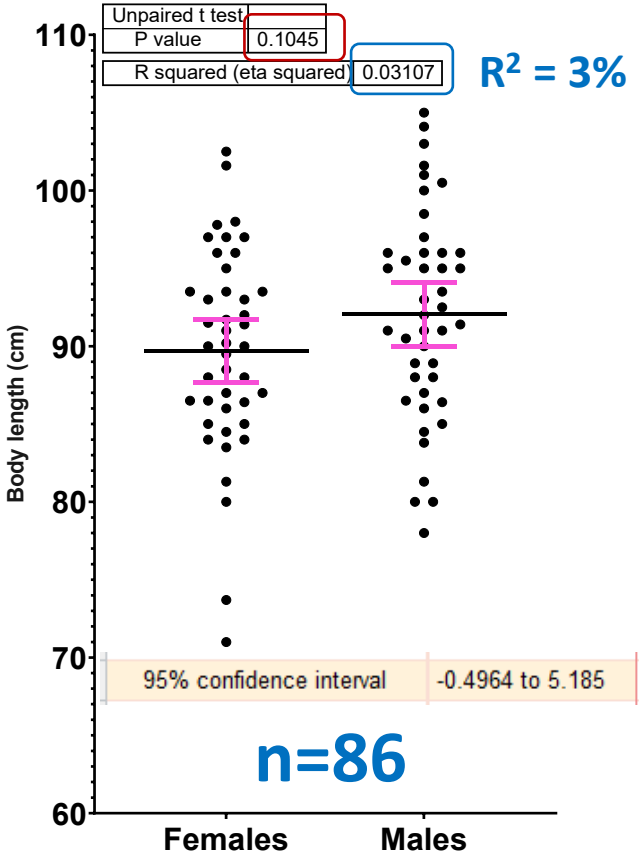
| RM one-way ANOVA<br>ANOVA results | |
|---|---|
| Table Analyzed | Neutrophils |
| | |
| **Repeated measures ANOVA summary** | |
| Assume sphericity? | No |
| F | 28.57 |
| P value | 0.0002 |
| P value summary | *** |
| Statistically significant (P < 0.05)? | Yes |
| Geisser-Greenhouse's epsilon | 0.6916 |
| R square | 0.8772 |

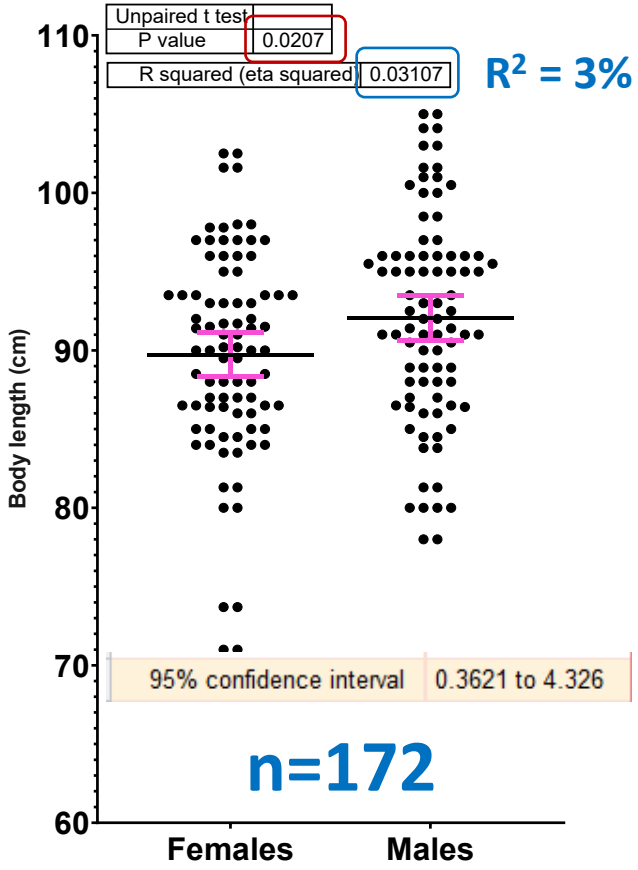
88%

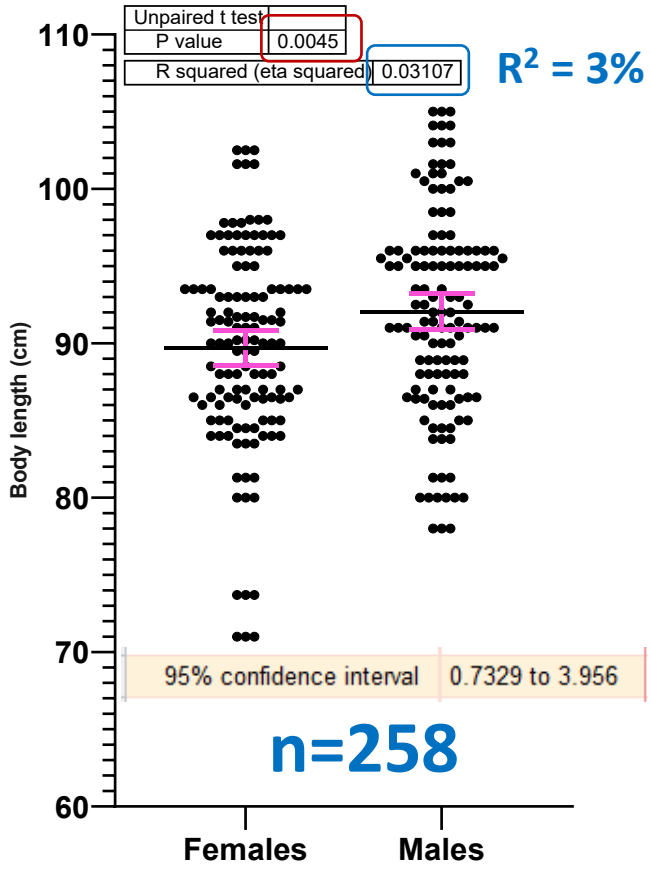
# Coefficient of determination

## An illustration: change in sample size

Sample ↑ = Power ↑ → $R^2$ does not change but p-value ↓



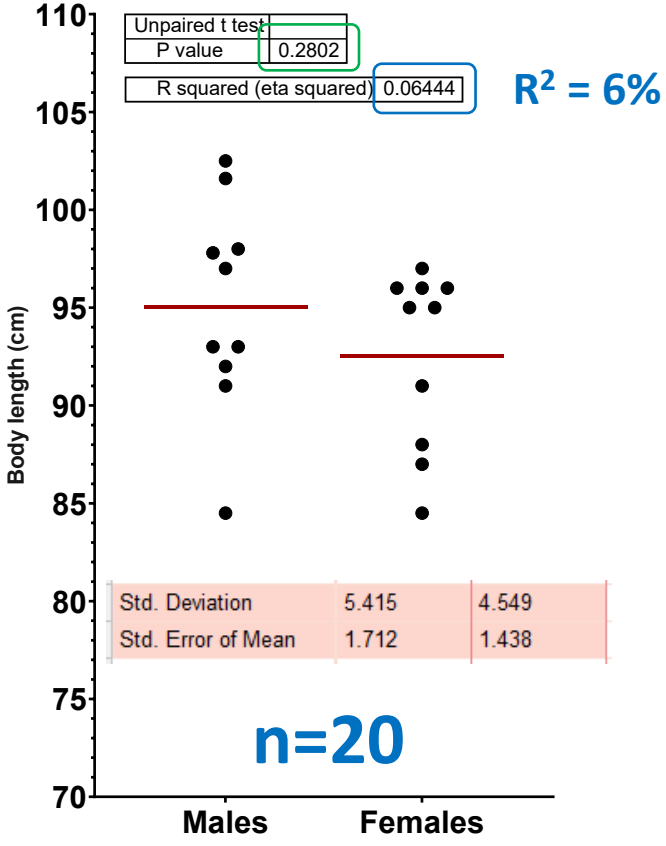



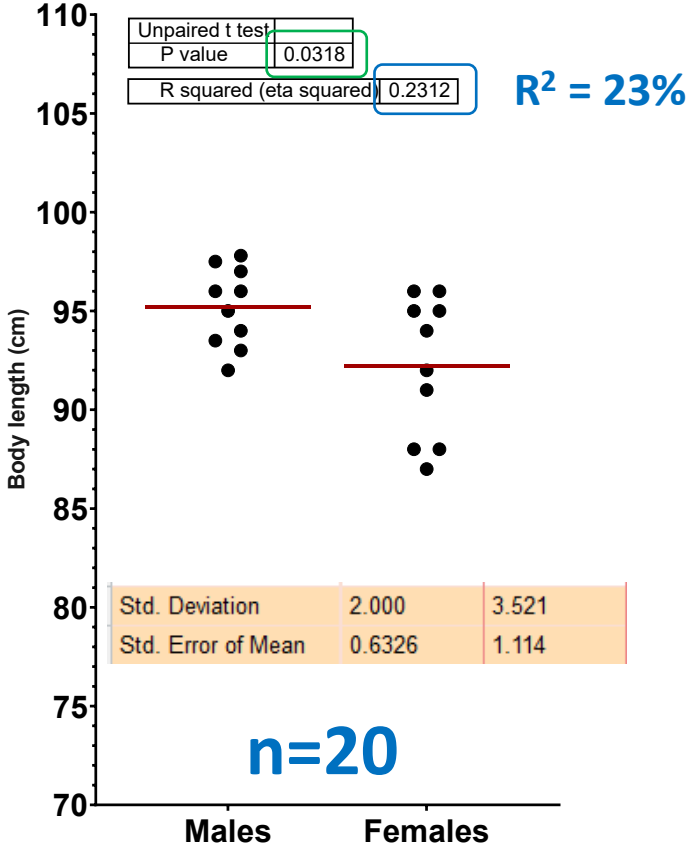



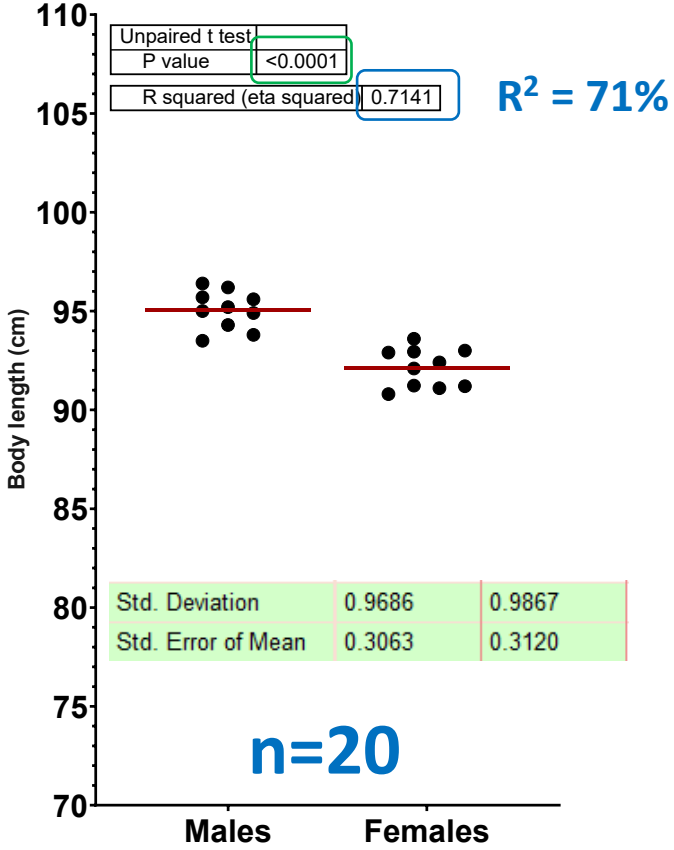
# Coefficient of determination

## An illustration: change in variability

Same sample = same power → variability ↓ $R^2$ ↑ p-value ↓



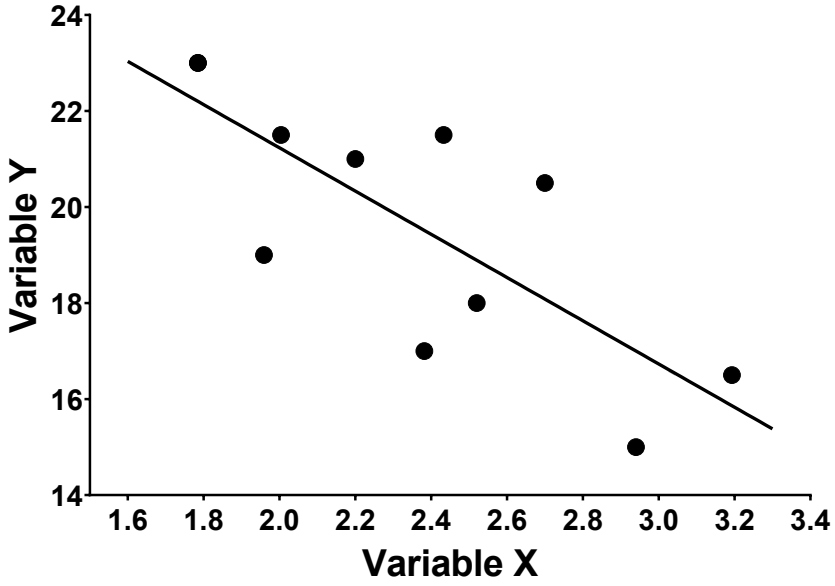



# Correlation: Assumptions

Pearson correlation is a parametric test for linear relationships

First assumption for parametric test: Normality

**Correlation: bivariate Gaussian distribution**



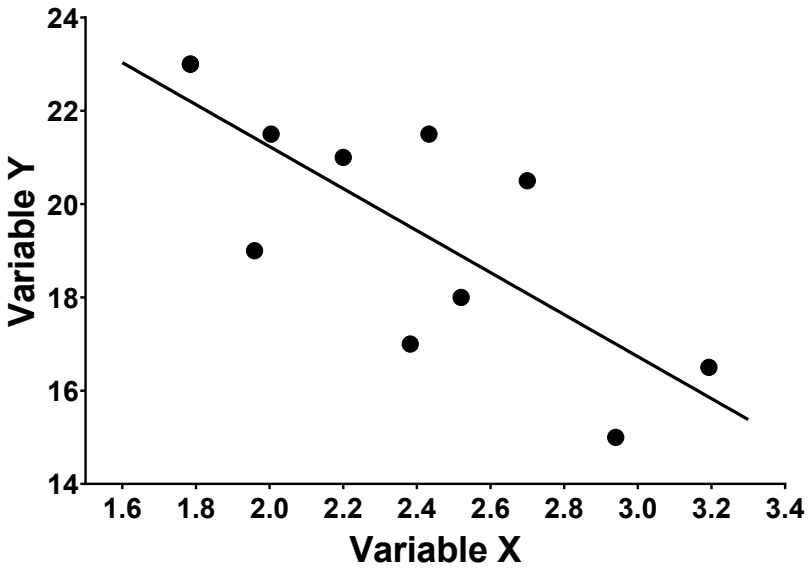

**Symmetry of the values on either side of the line of best fit.**

# Correlation and regression

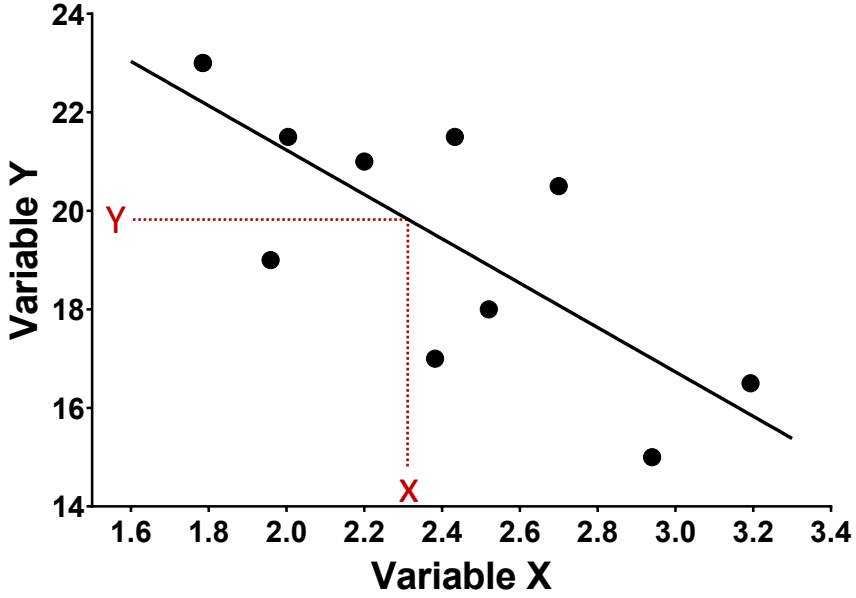
Line of best fit comes from a regression

**Correlation:** nature and strength of the association

**Regression:** nature and strength of the association <u>and</u> prediction




**Correlation = Association**




**Regression = Prediction**

**Y = A*X + B**

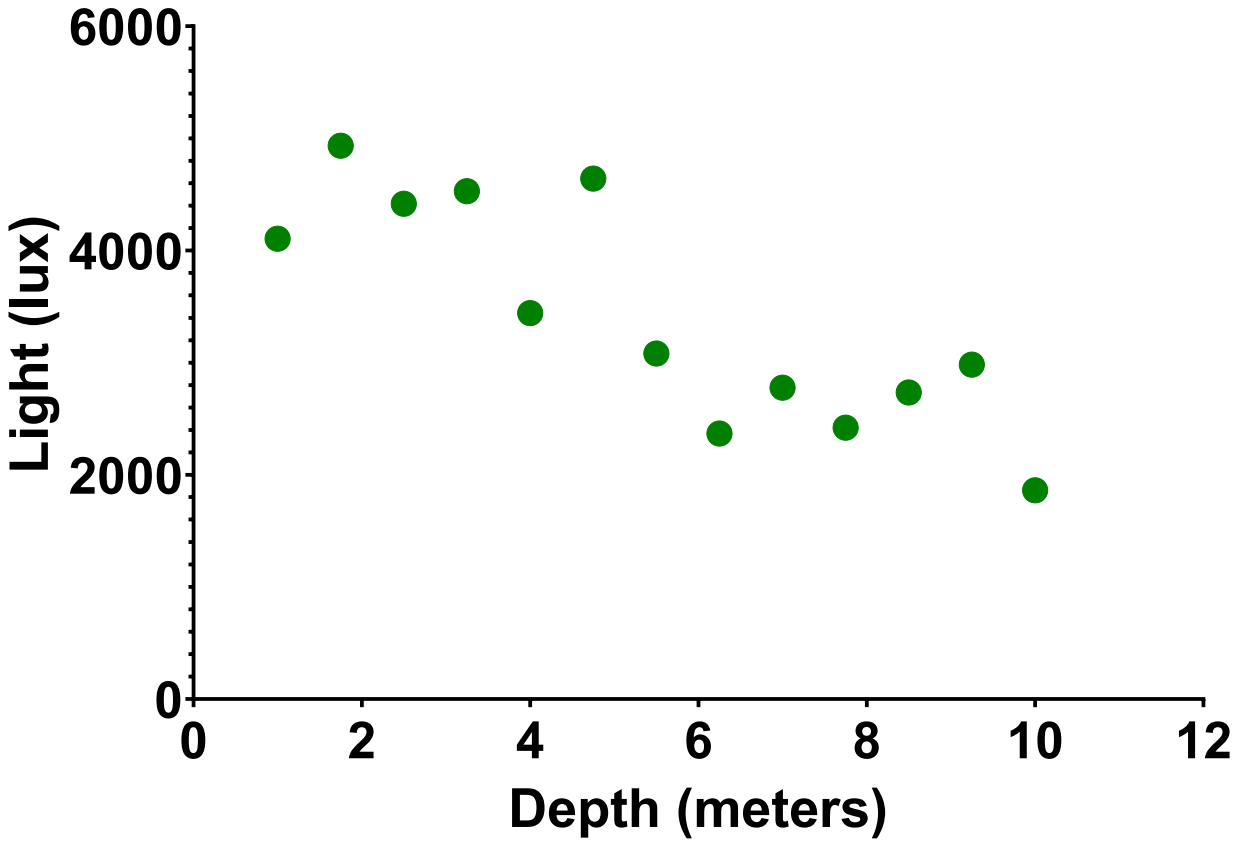
# Correlation

## **treelight.xlsx** (conifer)

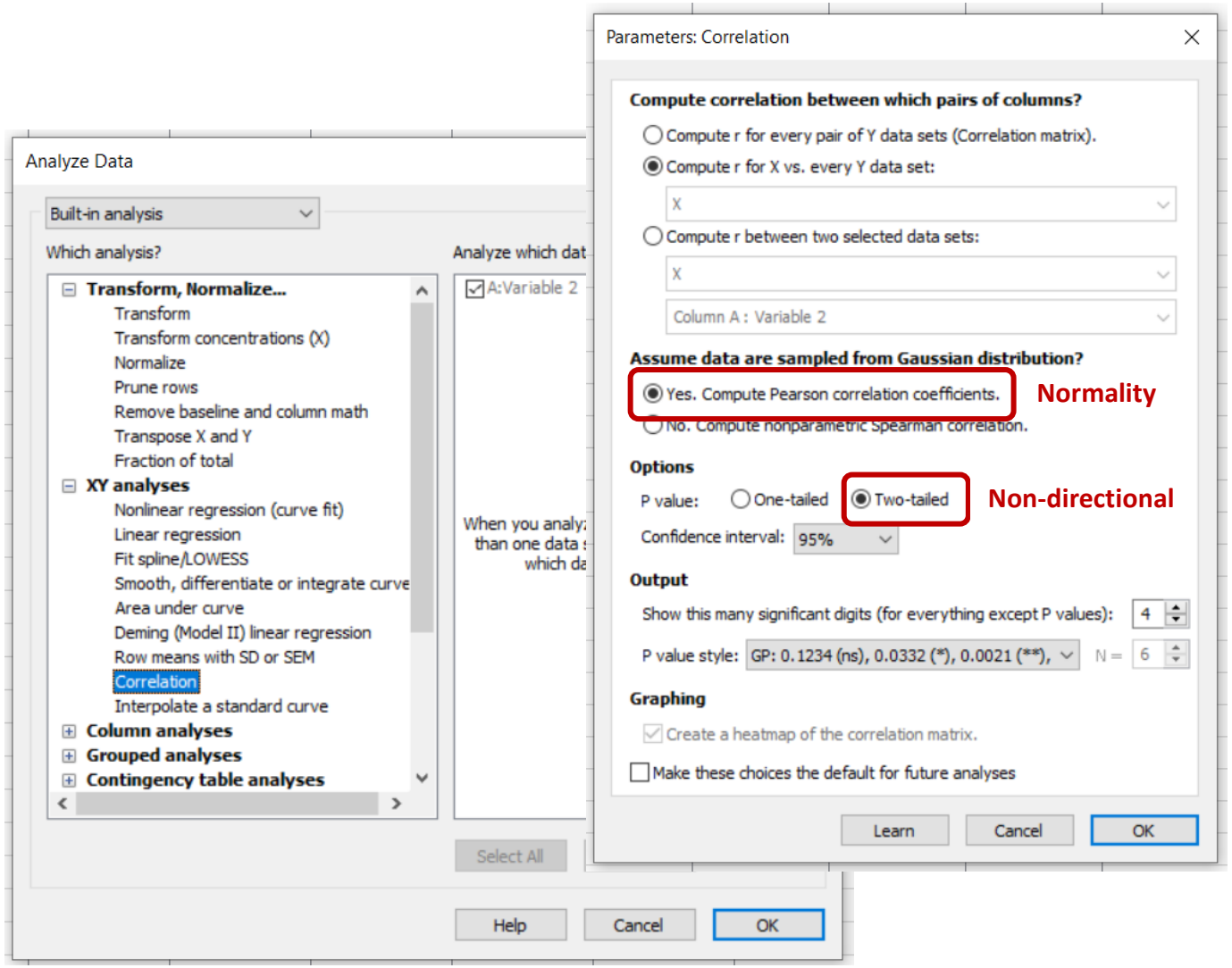
Amount of light in a tree

# Correlation with Prism

## Amount of light in a tree



# Correlation with Prism




## Results

| | Correlation | Depth vs. Light |
|---|---|---|
| 1 | **Pearson r** | |
| 2 | r | -0.8465 |
| 3 | 95% confidence interval | -0.9530 to -0.5537 |
| 4 | R squared | 0.7165 |
| 5 | | |
| 6 | **P value** | |
| 7 | P (two-tailed) | 0.0003 |
| 8 | P value summary | *** |
| 9 | Significant? (alpha = 0.05) | Yes |
| 10 | | |
| 11 | **Number of XY Pairs** | 13 |

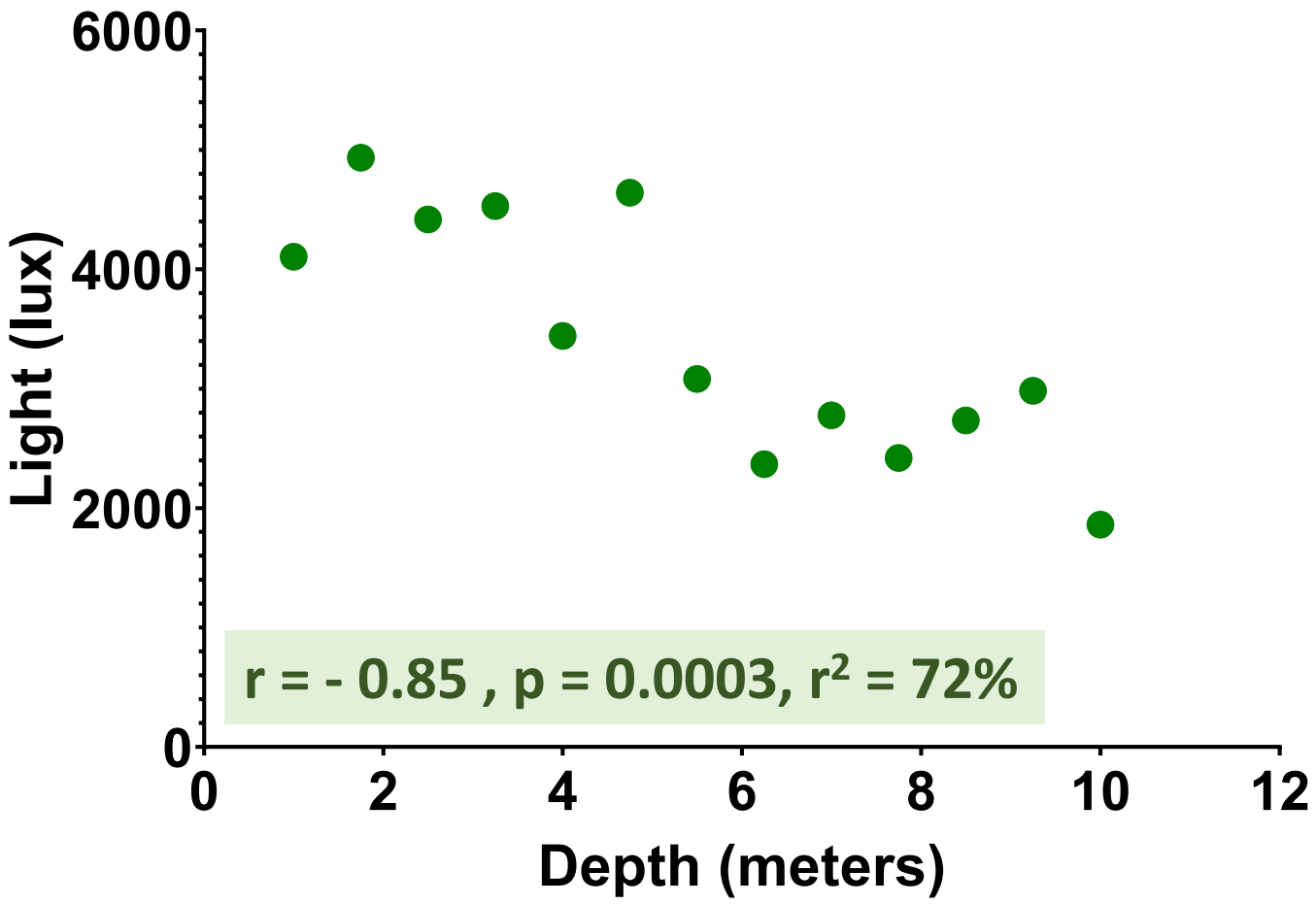
Association

Strong Negative

Significant

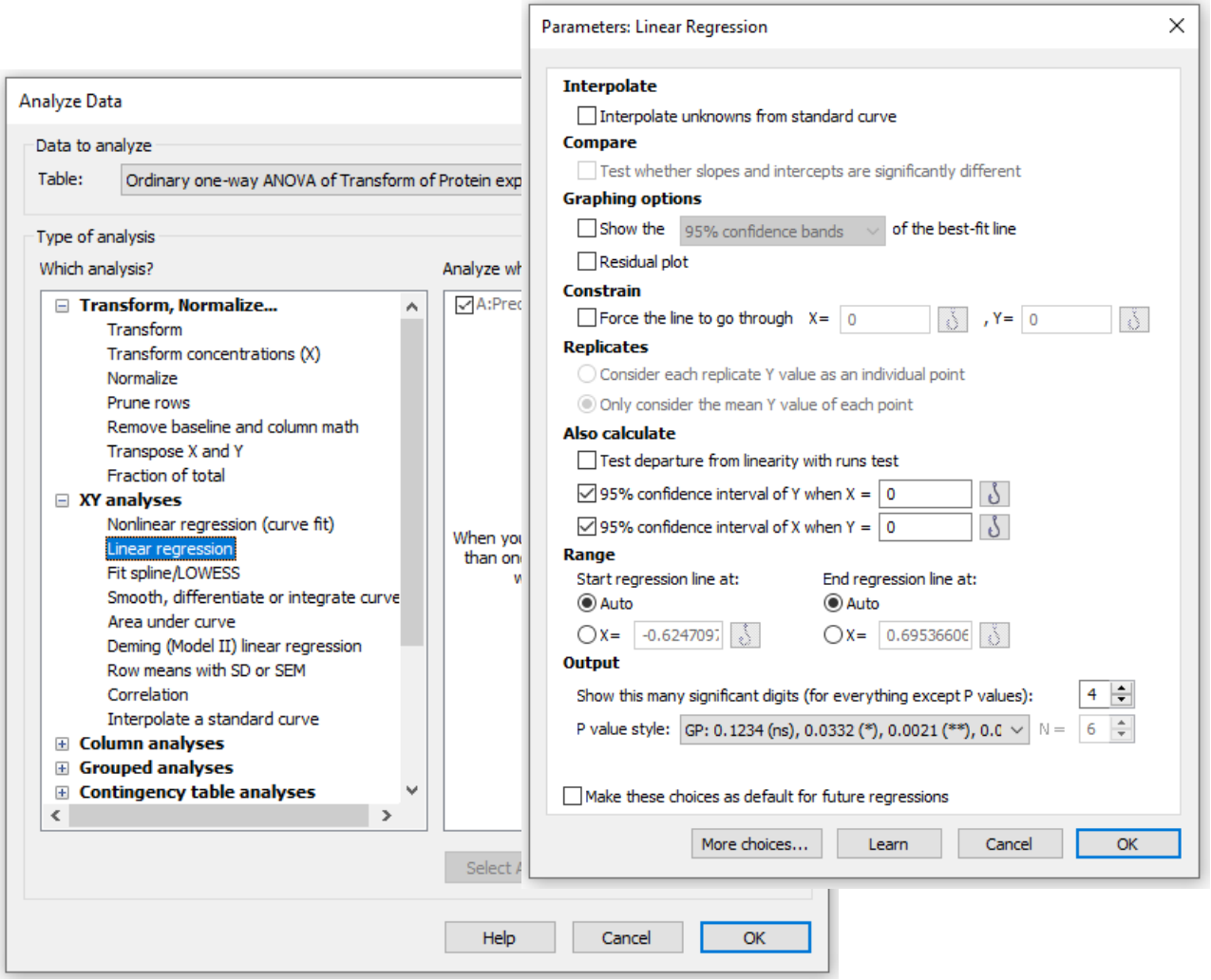
# Correlation with Prism

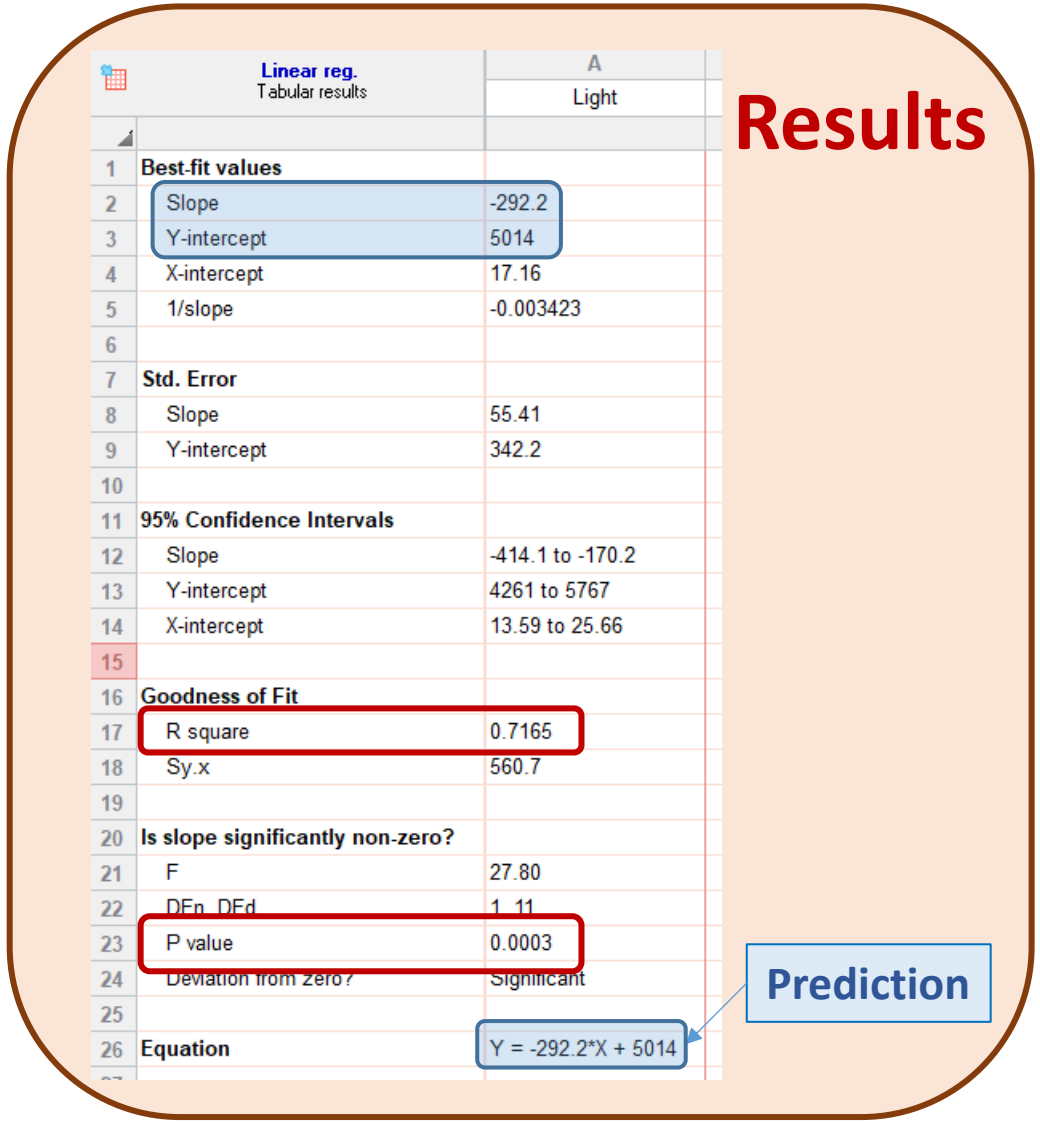
## Amount of light in a tree



# Correlation with Prism

## (if you also want a line of best fit)



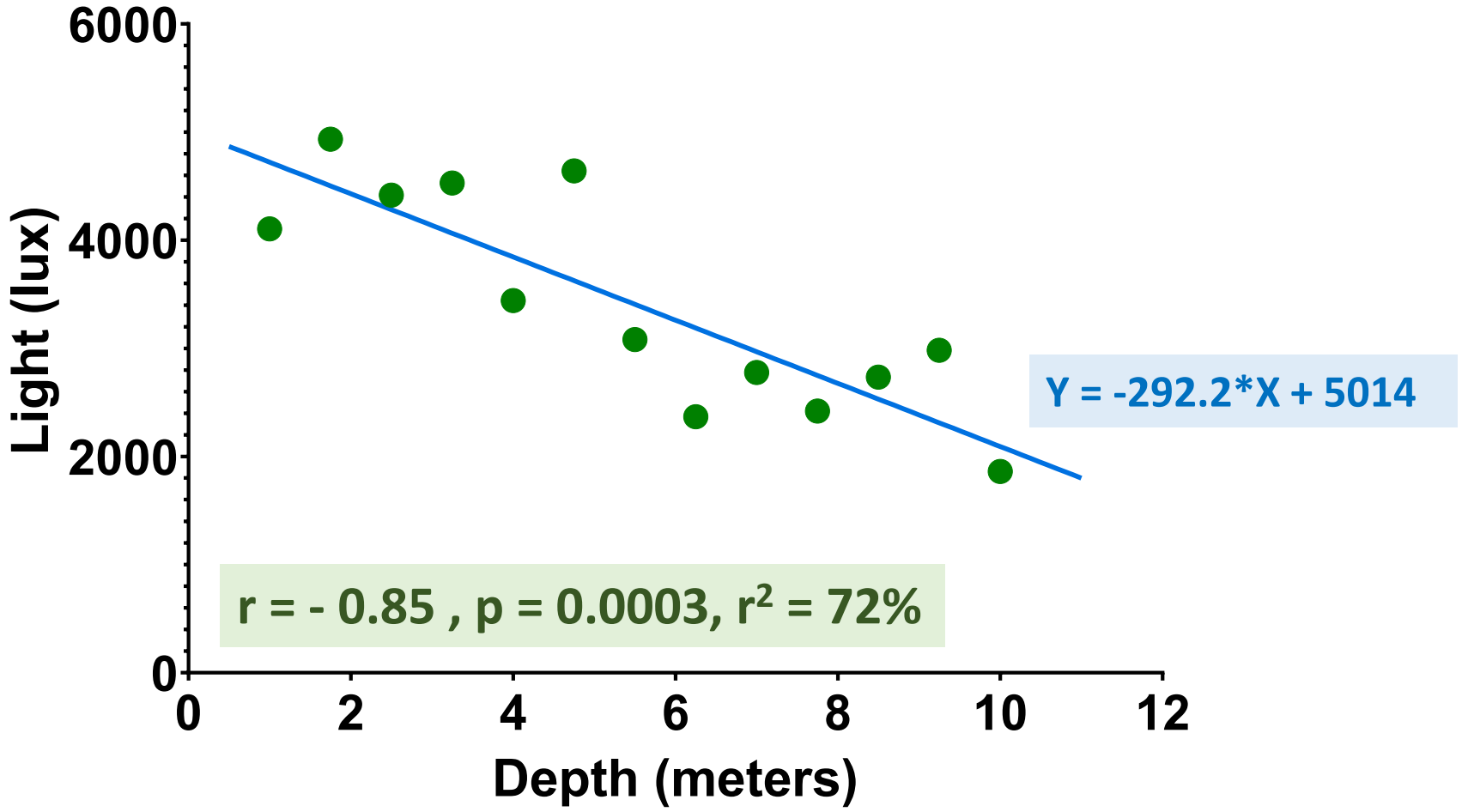
Results

| | Linear reg. Tabular results | A Light |
|---|---|---|
| 1 | **Best-fit values** | |
| 2 | Slope | -292.2 |
| 3 | Y-intercept | 5014 |
| 4 | X-intercept | 17.16 |
| 5 | 1/slope | -0.003423 |
| 6 | | |
| 7 | **Std. Error** | |
| 8 | Slope | 55.41 |
| 9 | Y-intercept | 342.2 |
| 10 | | |
| 11 | **95% Confidence Intervals** | |
| 12 | Slope | -414.1 to -170.2 |
| 13 | Y-intercept | 4261 to 5767 |
| 14 | X-intercept | 13.59 to 25.66 |
| 15 | | |
| 16 | **Goodness of Fit** | |
| 17 | R square | 0.7165 |
| 18 | Sy.x | 560.7 |
| 19 | | |
| 20 | **Is slope significantly non-zero?** | |
| 21 | F | 27.80 |
| 22 | DFn, DFd | 1, 11 |
| 23 | P value | 0.0003 |
| 24 | Deviation from zero? | Significant |
| 25 | | |
| 26 | **Equation** | Y = -292.2*X + 5014 |

Prediction

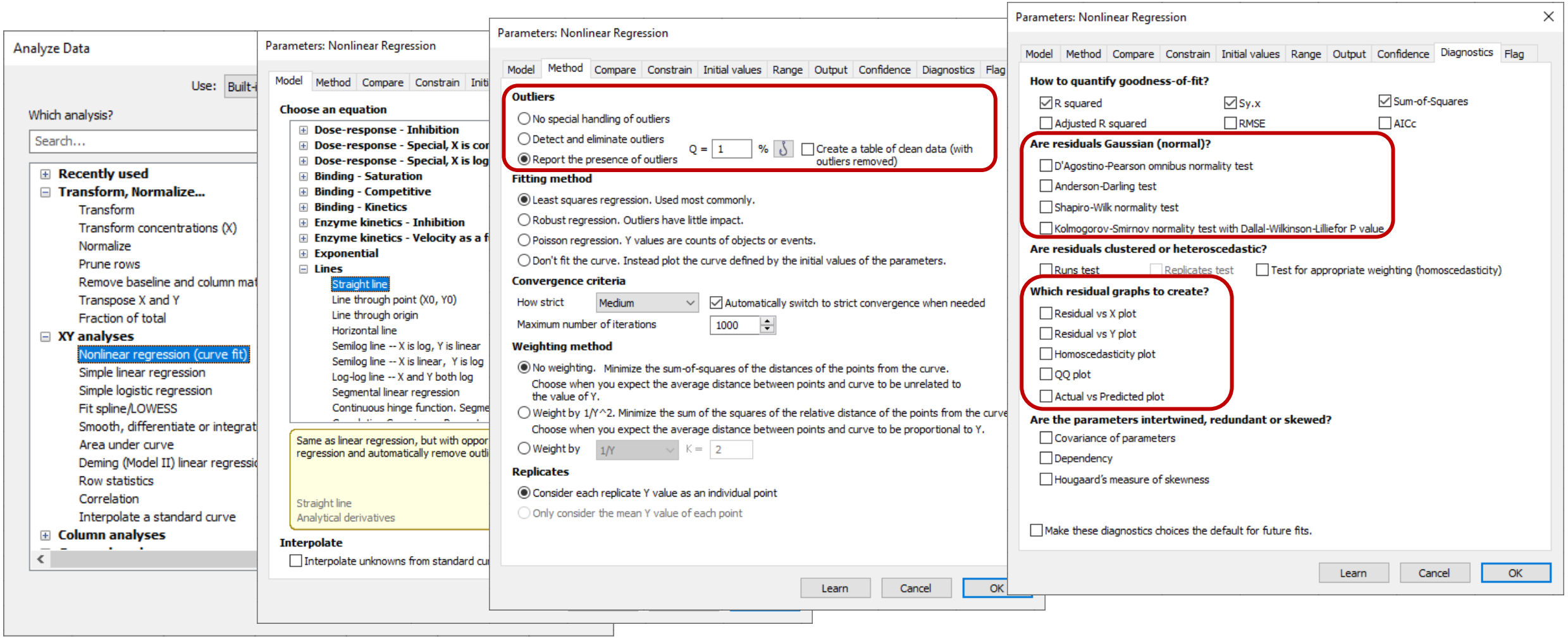
# Correlation with Prism

## Amount of light in a tree



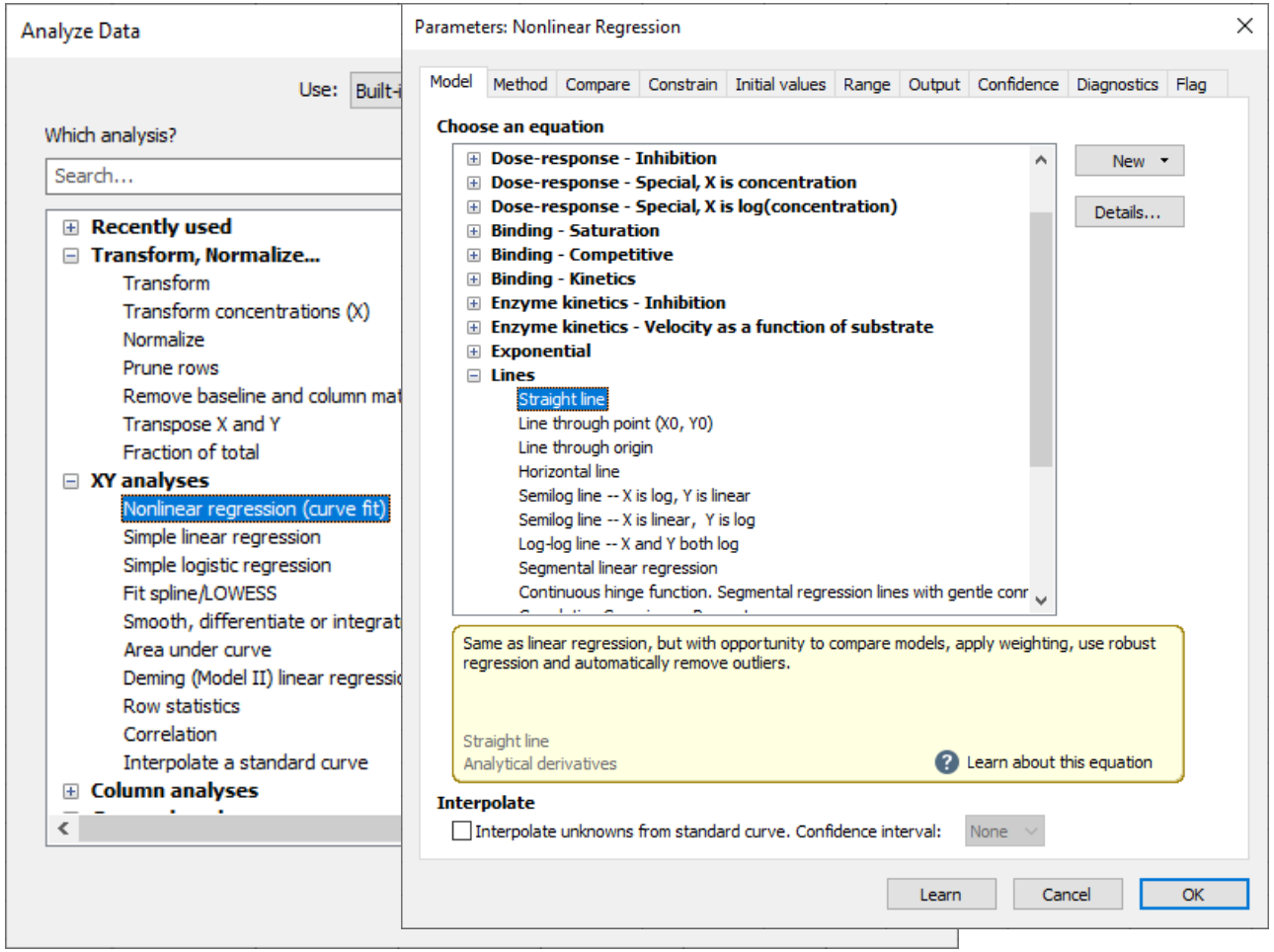
# Correlation with Prism

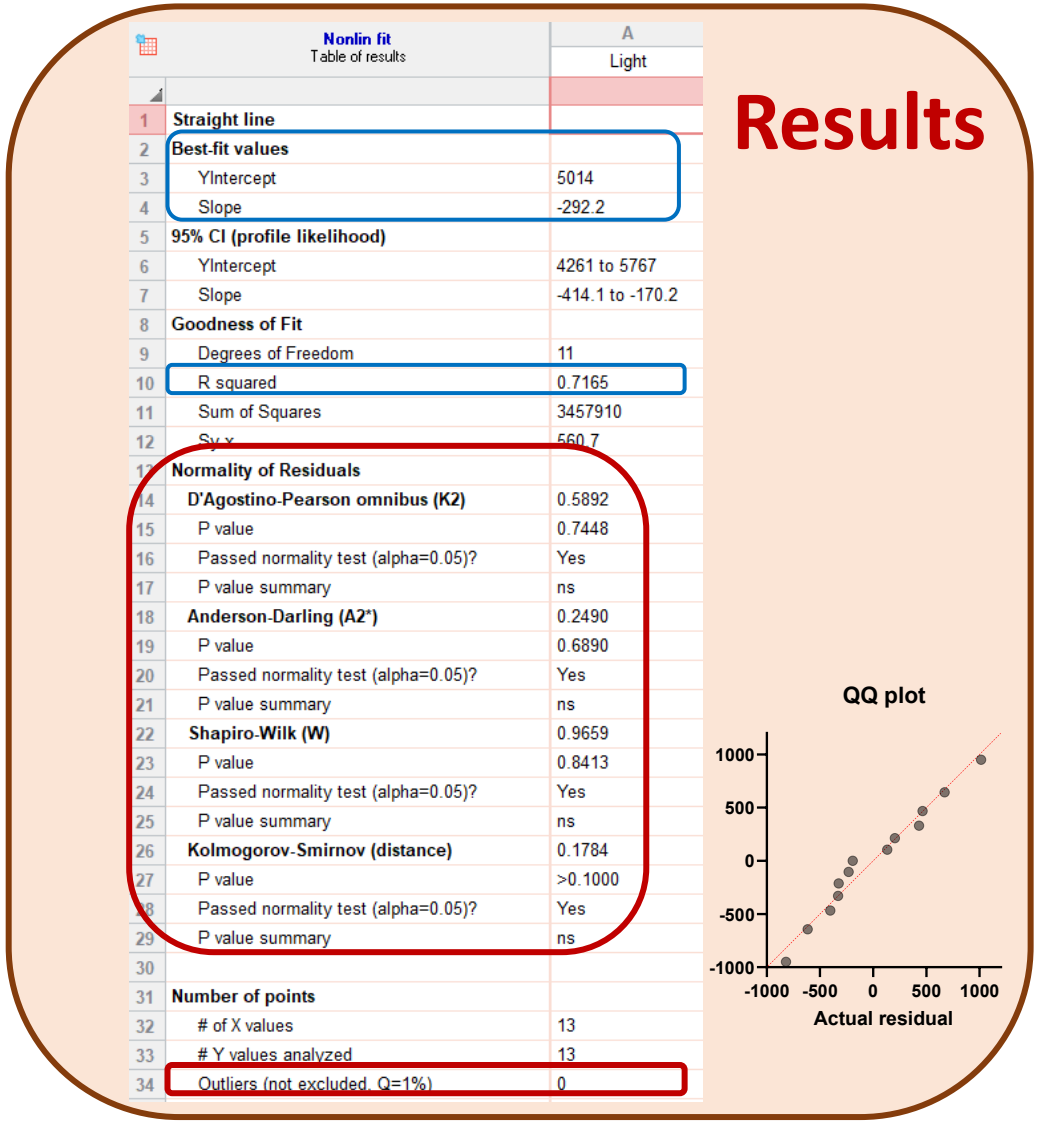
## (if you also want more options)



# Correlation with Prism

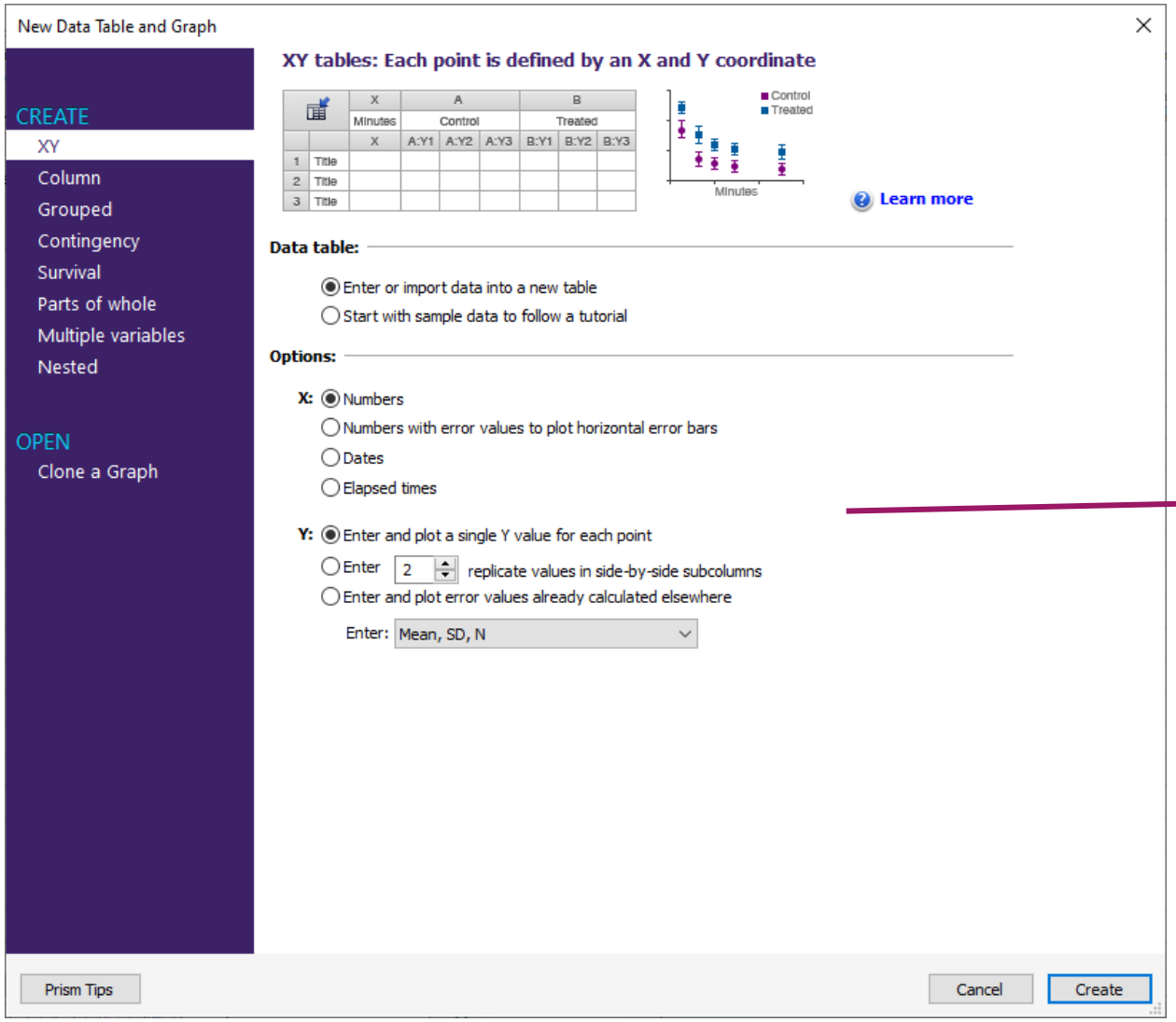
## (if you also want more options)



# Correlation with Prism

## If you have more than one group




| Table format: XY | X | Group A | Group B |
|---|---|---|---|
| | Parasite burden | Female | Male |
| | X | Y | Y |
| 1 Title | 2.025306 | | 21.0 |
| 2 Title | 2.699838 | | 20.5 |
| 3 Title | 1.959041 | | 21.0 |
| 4 Title | 2.382017 | | 17.0 |
| 5 Title | 2.940018 | | 15.0 |
| 6 Title | 3.193403 | | 16.5 |
| 7 Title | 1.959041 | | 19.0 |
| 8 Title | 2.519828 | | 15.5 |
| 9 Title | 1.785330 | | 23.0 |
| 10 Title | 2.004321 | | 21.5 |
| 11 Title | 1.785330 | | 23.0 |
| 12 Title | 2.432969 | | 21.5 |
| 13 Title | 2.456366 | 17.0 | |
| 14 Title | 2.432969 | 21.0 | |
| 15 Title | 1.491362 | 22.5 | |
| 16 Title | 2.220108 | 23.0 | |
| 17 Title | 2.178977 | 24.5 | |
| 18 Title | 2.257679 | 16.5 | |
| 19 Title | 2.432969 | 20.0 | |
| 20 Title | 2.082785 | 24.5 | |
| 21 Title | 1.491362 | 23.0 | |
| 22 Title | 2.257679 | 21.0 | |
| 23 Title | 1.491362 | 20.5 | |

# Exercise 4

# Association between 2 continuous variables

## One variable X and One variable Y

## One predictor

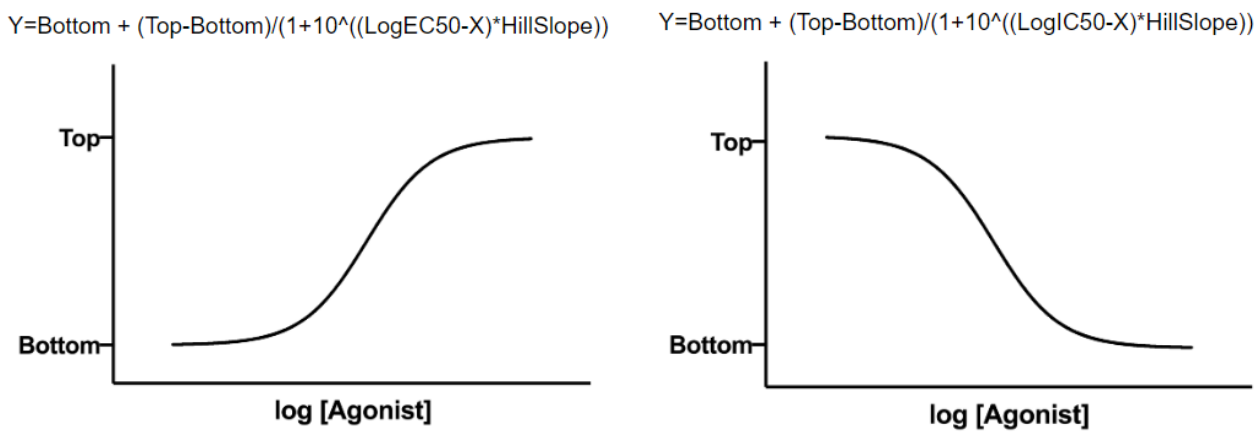
## Non-linear relationship

## <u>Curve fitting</u>

# Curve fitting

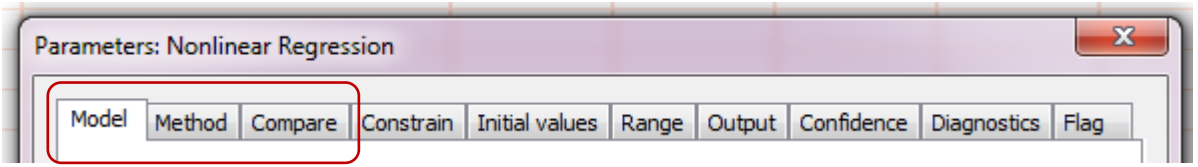
- **Dose-response curves**
  - Nonlinear regression
  - Dose-response experiments typically use 5-10 doses of agonist, equally spaced on a logarithmic scale
  - Y values are responses

- The aim is often to determine the **IC50** or the **EC50**
  - **IC50 (I=Inhibition)**: concentration of an agonist that provokes a response half way between the maximal (Top) response and the maximally inhibited (Bottom) response
  - **EC50 (E=Effective):** concentration that gives half-maximal response



# Curve fitting

## Example: Inhibition data.xlsx

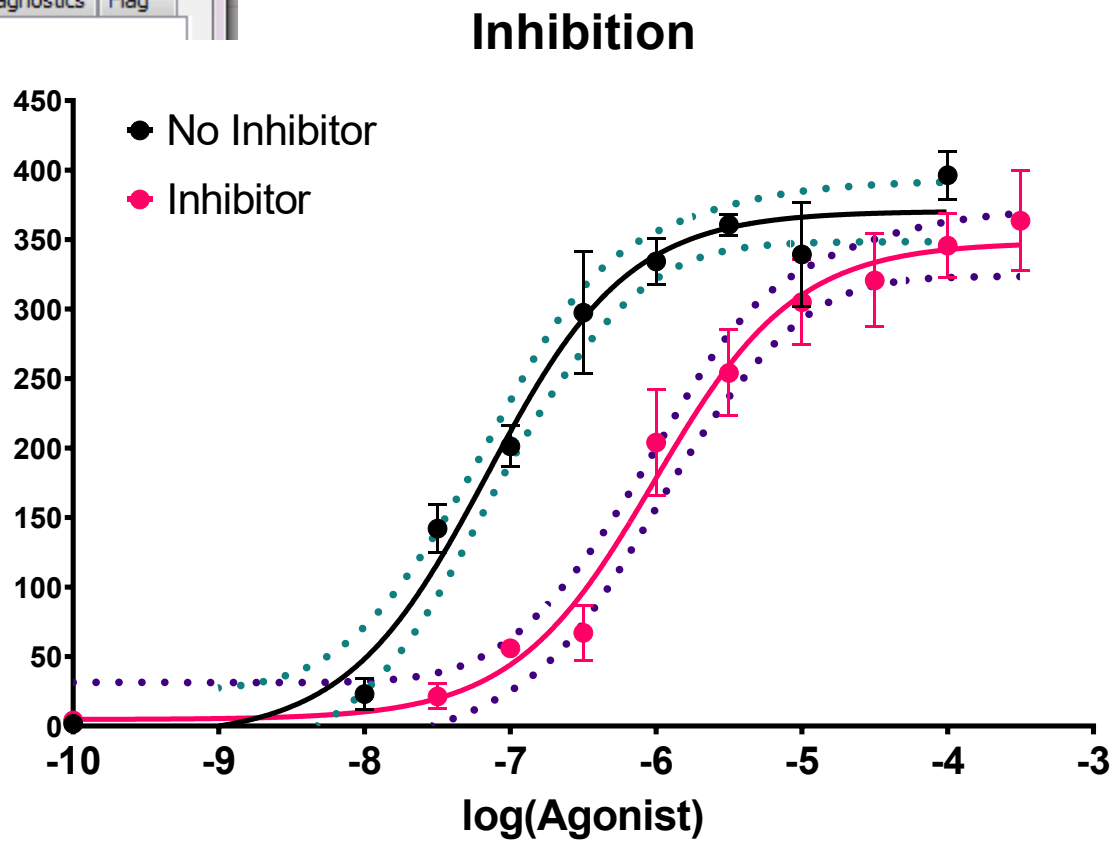



Step by step analysis and considerations:

1. Choose a **Model**

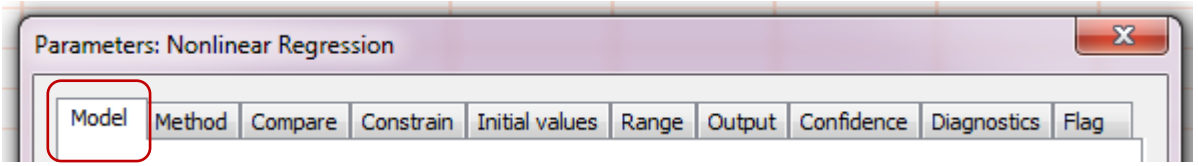
2. Choose a **Method**

3. **Compare** different conditions



# Curve fitting

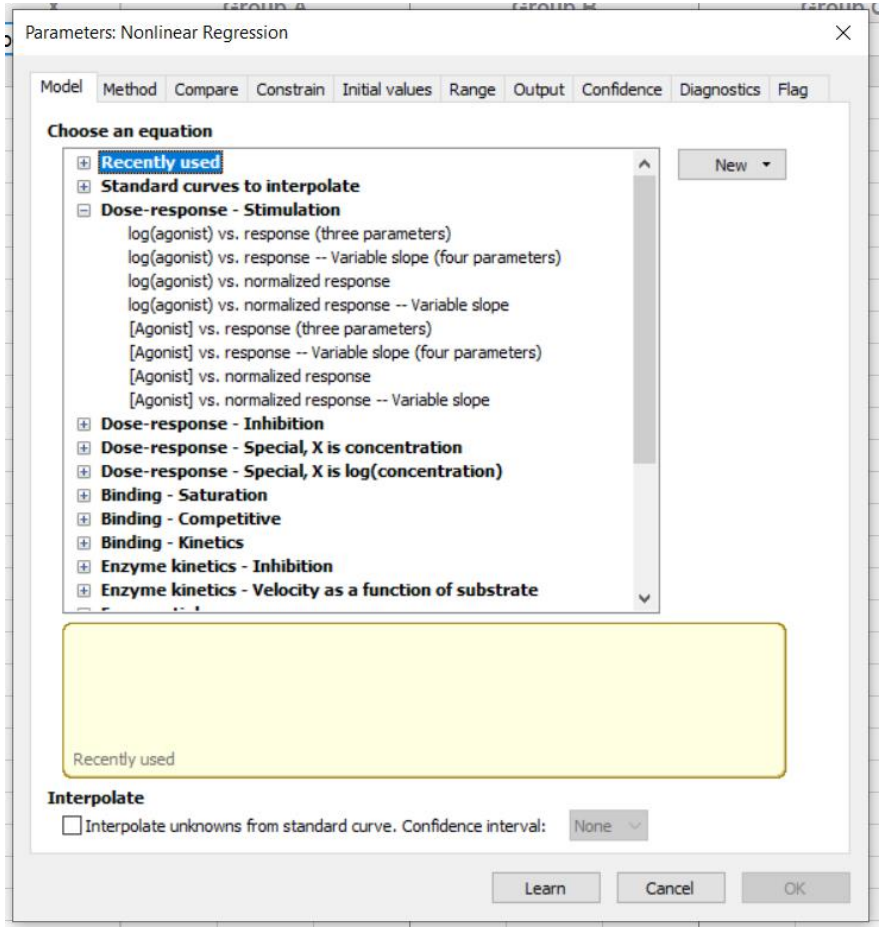
## Example: Inhibition data.xlsx




Step by step analysis and considerations:

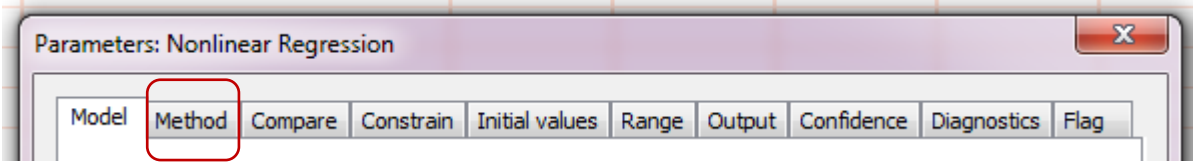
1. Choose a **Model**

- not necessary to normalise
- should choose it when values defining 0 and 100 are precise
- variable slope better if plenty of data points (variable slope or 4 parameters)



# Curve fitting

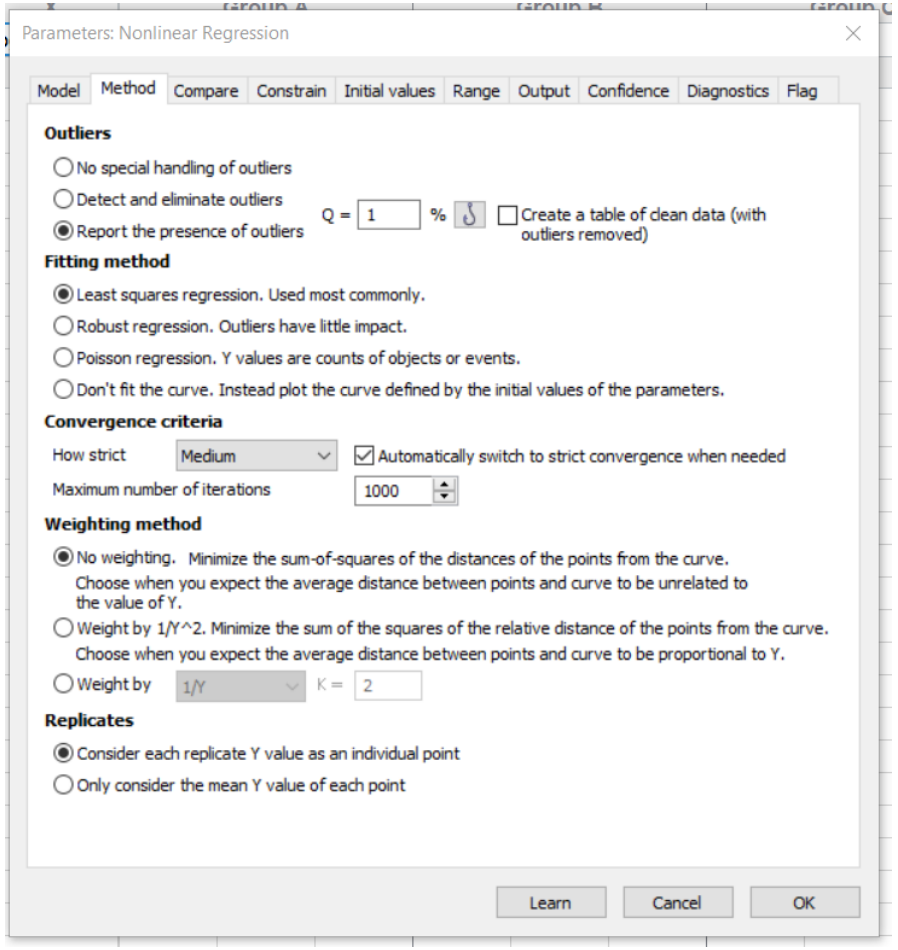
## Example: Inhibition data.xlsx




Step by step analysis and considerations:

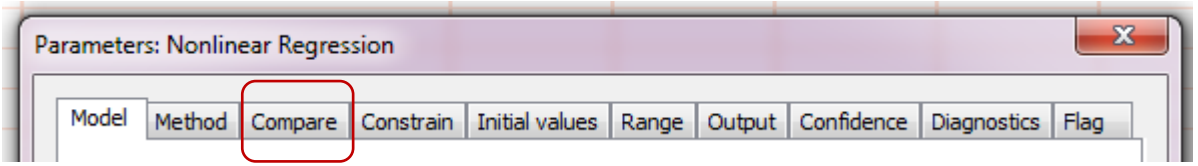
2. Choose a **Method**

- outliers, fitting method, weighting method and replicates



# Curve fitting

## Example: Inhibition data.xlsx




Step by step analysis and considerations:

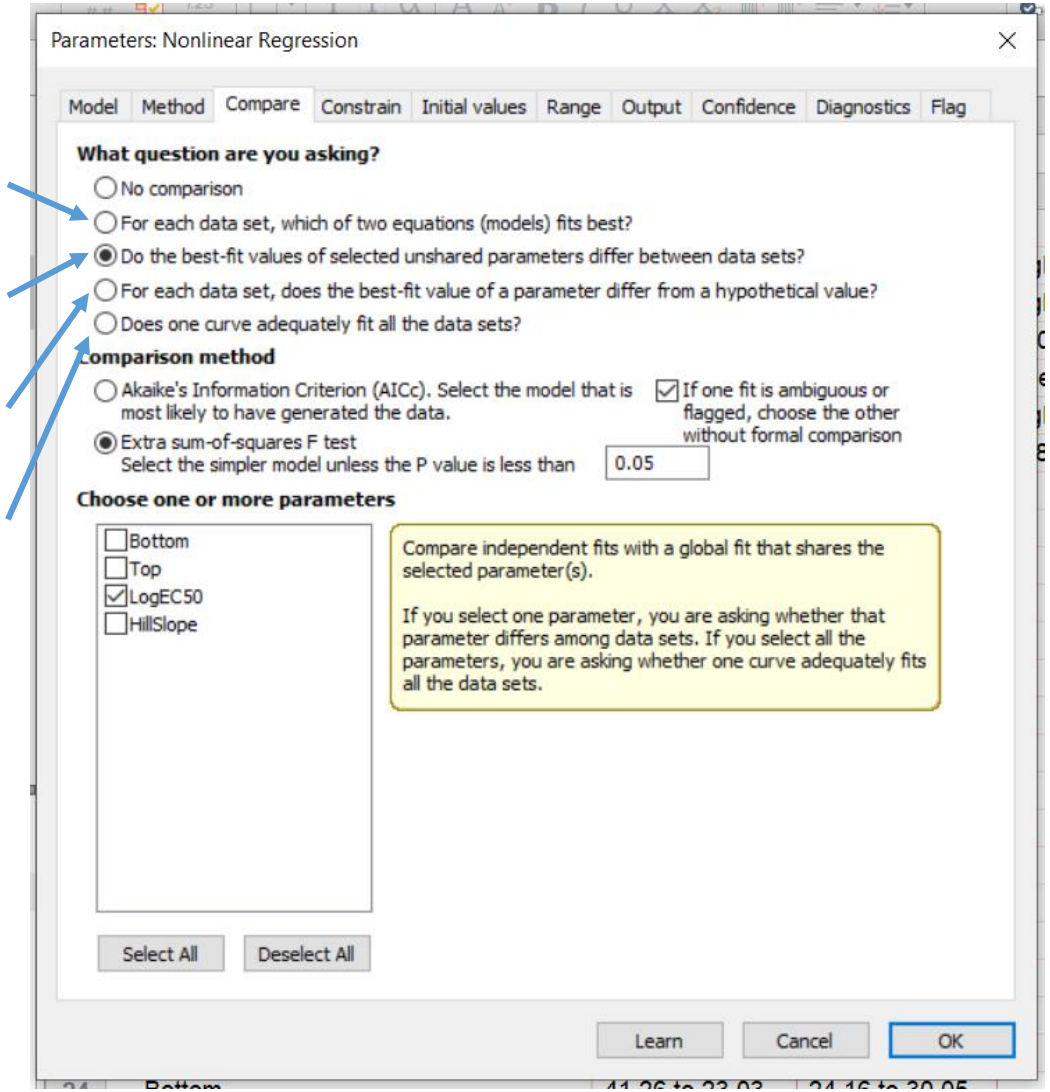
3. **Compare** different conditions

Diff in parameters

Diff between conditions for one or more parameters

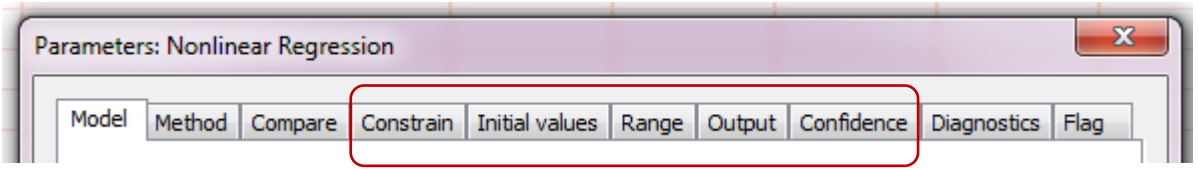
Constraint vs no constraint

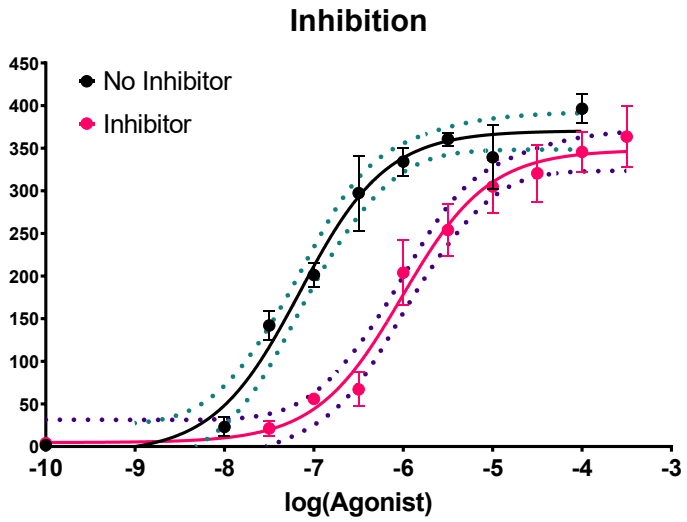
Diff between conditions for one or more parameters



# Curve fitting

## Example: Inhibition data




4. **Constrain**:

   depends on your experiment

   depends if your data don't define the top or the bottom of the curve

5. **Initial values**:

   defaults usually OK unless the fit looks funny

6. **Range**:

   defaults usually OK unless you are not interested in the x-variable full range (i.e. time)

7. **Output**:

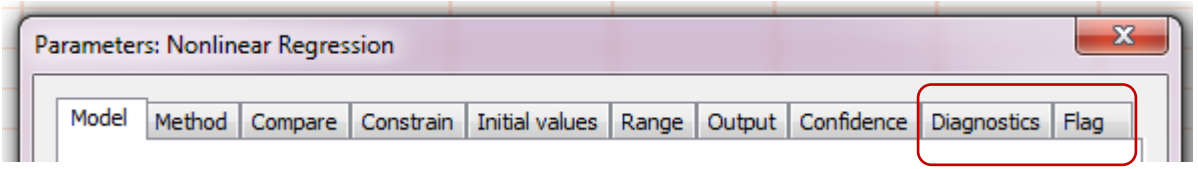
   summary table presents same results in a summarised way.

8. **Confidence**:

   calculate and plot confidence intervals

# Curve fitting

## Example: Inhibition data




9. **Diagnostics**:

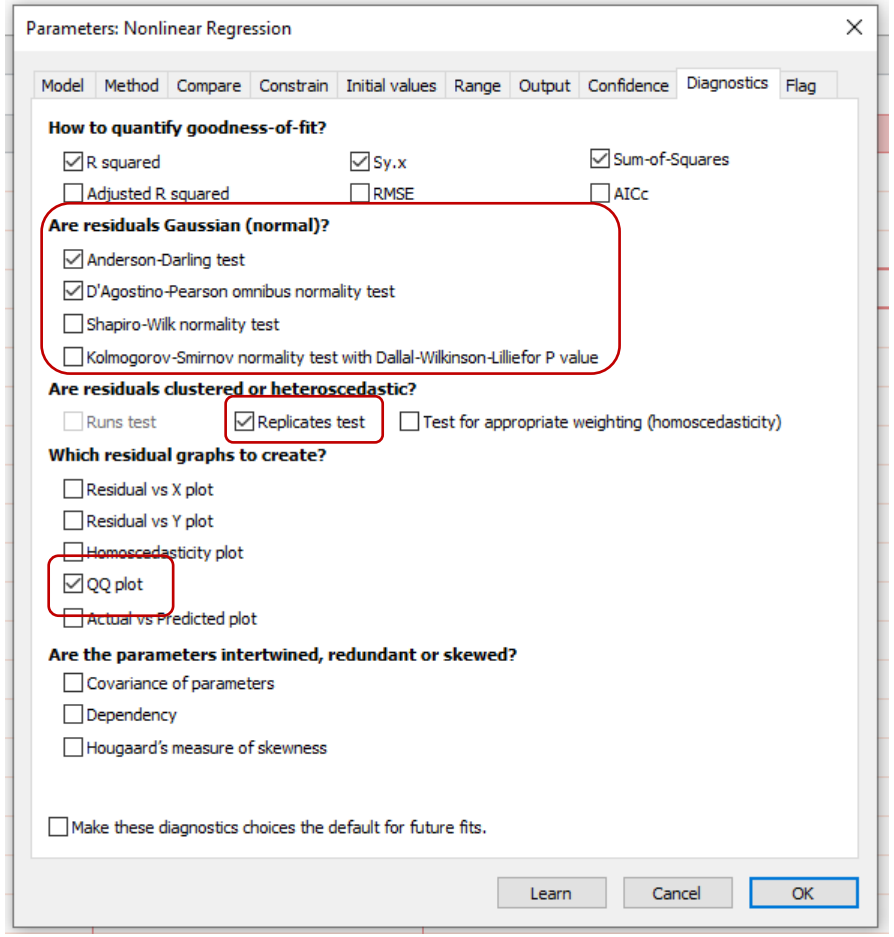
   Check for normality

   Check Replicates test (heteroscedasticity)

   QQ plots

10. **Flag**:

   Only useful if doing lots of curve fits

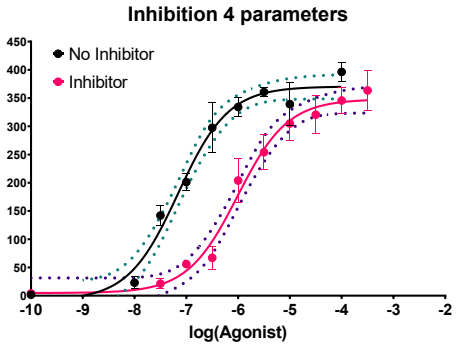


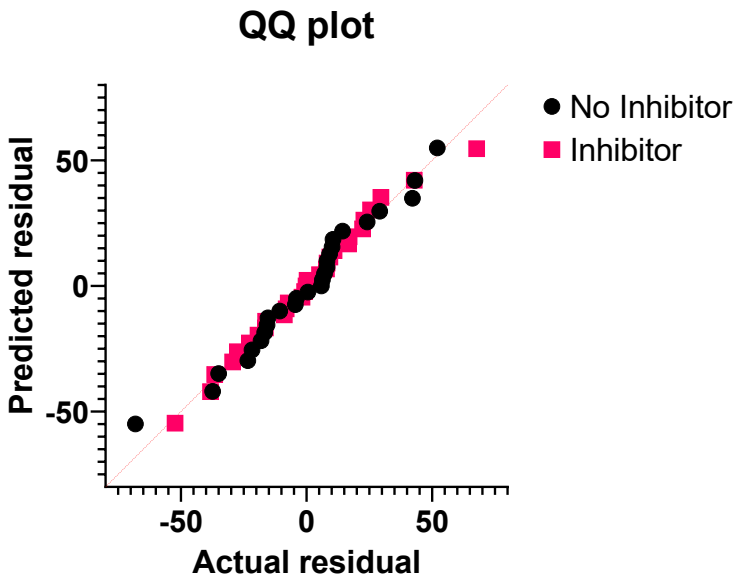
# Curve fitting: Diagnostics

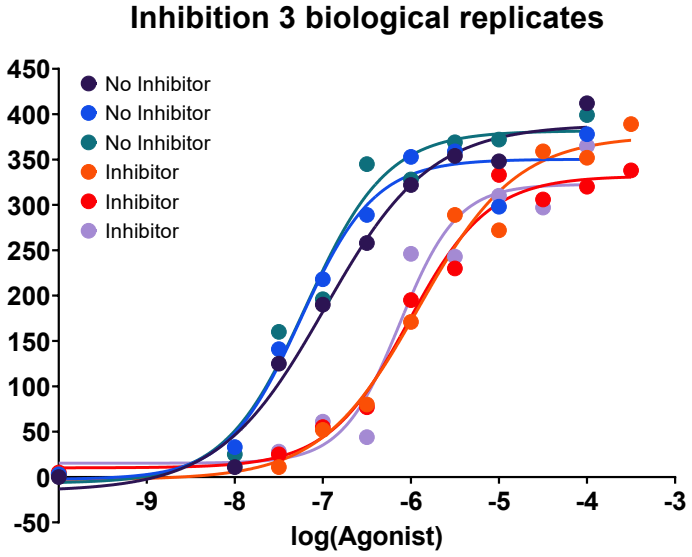
Table of results | Outliers

| | Nonlin fit<br>Table of results | A<br>No Inhibitor | B<br>Inhibitor |
|---|---|---|---|
| 28 | **Normality of Residuals** | | |
| 29 | **Anderson-Darling (A2*)** | 0.3406 | 0.1465 |
| 30 | P value | 0.4702 | 0.9622 |
| 31 | Passed normality test (alpha=0.05)? | Yes | Yes |
| 32 | P value summary | ns | ns |
| 33 | **D'Agostino-Pearson omnibus (K2)** | 1.343 | 1.374 |
| 34 | P value | 0.5110 | 0.5031 |
| 35 | Passed normality test (alpha=0.05)? | Yes | Yes |
| 36 | P value summary | ns | ns |
| 37 | **Replicates test for lack of fit** | | |
| 38 | SD replicates | 22.71 | 25.52 |
| 39 | SD lack of fit | 41.84 | 32.38 |
| 40 | Discrepancy (F) | 3.393 | 1.610 |
| 41 | P value | 0.0247 | 0.1989 |
| 42 | Evidence of inadequate model? | Yes | No |

| Table format: XY | X | Group A | Group B | Group C | Group D | Group E | Group F |
|---|---|---|---|---|---|---|---|
| | log(Agonist) | No Inhibitor | No Inhibitor | No Inhibitor | Inhibitor | Inhibitor | Inhibitor |
| | X | Y | Y | Y | Y | Y | Y |
| Title | -7.0 | 190 | 218 | 196 | 52 | 55 | 61 |
| Title | -6.5 | 258 | 289 | 345 | 80 | 77 | 44 |
| Title | -6.0 | 322 | 353 | 328 | 171 | 195 | 246 |
| Title | -5.5 | 354 | 359 | 369 | 289 | 230 | 243 |
| Title | -5.0 | 348 | 298 | 372 | 272 | 333 | 310 |
| Title | -4.5 | | | | 359 | 306 | 297 |
| Title | -4.0 | 412 | 378 | 399 | 352 | 320 | 365 |
| Title | -3.5 | | | | 389 | 338 | |



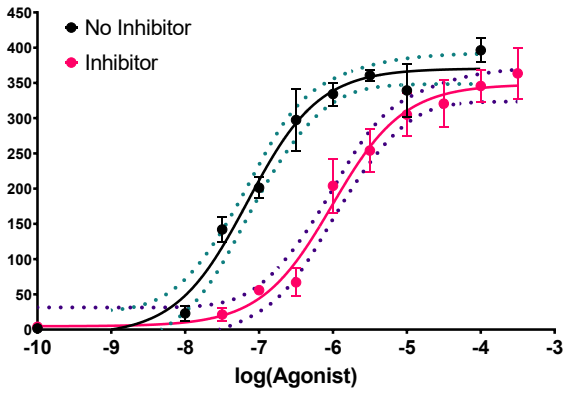




**Have a go!**

# Curve fitting: Results

## Inhibition 4 parameters

| |
|---|
| LogEC50 same for all data sets |
| LogEC50 different for each data set |
| < 0.0001 |
| Reject null hypothesis |
| LogEC50 different for each data set |
| 64.86 (1,48) |

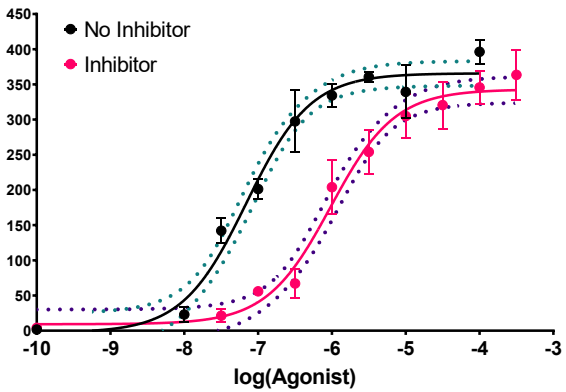



| 95% Confidence Intervals | | |
|---|---|---|
| Bottom | -41.39 to 24.94 | -22.15 to 31.56 |
| Top | 348.3 to 392.6 | 323.1 to 373.0 |
| LogEC50 | -7.324 to -6.991 | -6.185 to -5.837 |
| HillSlope | 0.6347 to 1.159 | 0.6095 to 1.186 |
| EC50 | 4.739e-008 to 1.020e-007 | 6.538e-007 to 1.455e-006 |

| | | |
|---|---|---|
| R square | 0.9663 | 0.9653 |

## Inhibition 3 parameters

| |
|---|
| LogEC50 same for all data sets |
| LogEC50 different for each data set |
| < 0.0001 |
| Reject null hypothesis |
| LogEC50 different for each data set |
| 101.0 (1,50) |

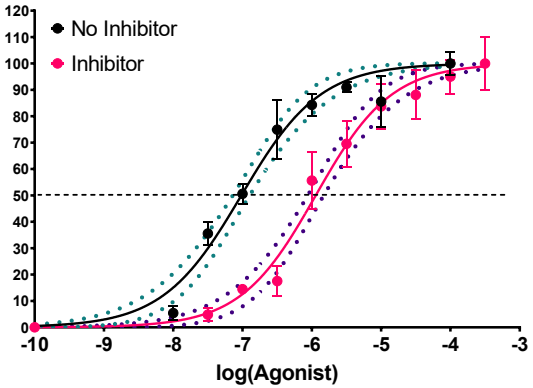



| 95% Confidence Intervals | | |
|---|---|---|
| Bottom | -30.74 to 24.78 | -11.65 to 30.07 |
| Top | 348.2 to 383.2 | 324.3 to 361.4 |
| LogEC50 | -7.312 to -7.006 | -6.175 to -5.859 |
| EC50 | 4.875e-008 to 9.858e-008 | 6.677e-007 to 1.385e-006 |

| | | |
|---|---|---|
| R square | 0.9655 | 0.9648 |

## Normalize 4 parameters

| |
|---|
| LogEC50 same for all data sets |
| LogEC50 different for each data set |
| < 0.0001 |
| Reject null hypothesis |
| LogEC50 different for each data set |
| 162.8 (1,52) |

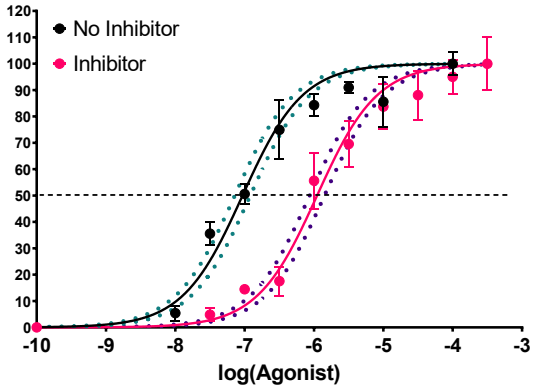



| 95% Confidence Intervals | | |
|---|---|---|
| LogEC50 | -7.137 to -6.897 | -6.057 to -5.830 |
| HillSlope | 0.6094 to 0.9184 | 0.6467 to 0.9460 |
| EC50 | 7.295e-008 to 1.268e-007 | 8.763e-007 to 1.481e-006 |

| | | |
|---|---|---|
| R square | 0.9580 | 0.9635 |

## Normalize 3 parameters

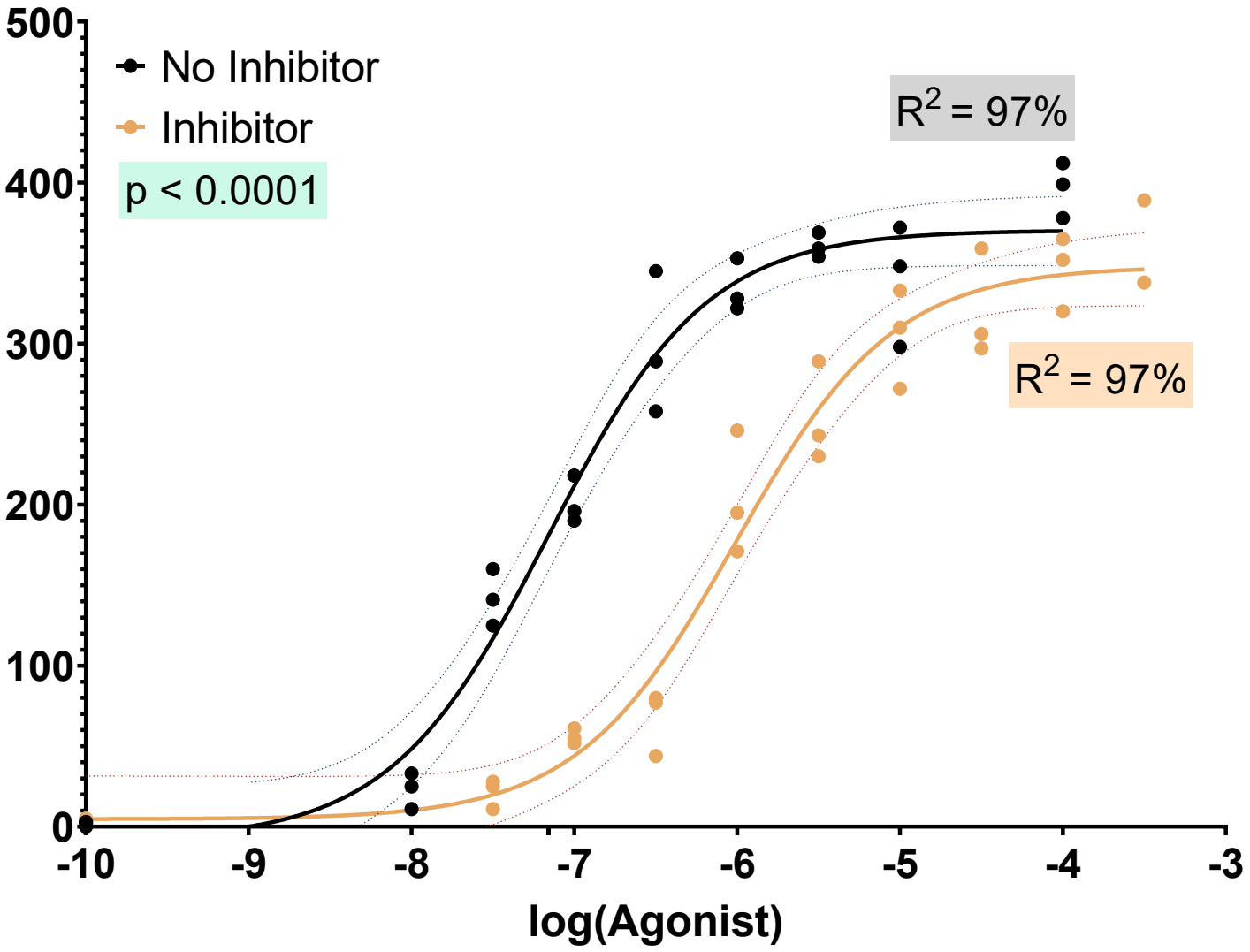
| |
|---|
| One curve for all data sets |
| Different curve for each data set |
| < 0.0001 |
| Reject null hypothesis |
| Different curve for each data set |
| 175.0 (1,54) |




| 95% Confidence Intervals | | |
|---|---|---|
| LogEC50 | -7.144 to -6.917 | -6.064 to -5.848 |
| EC50 | 7.179e-008 to 1.209e-007 | 8.633e-007 to 1.420e-006 |

| | | |
|---|---|---|
| R square | 0.9476 | 0.9568 |

# Curve fitting



Babraham Institute

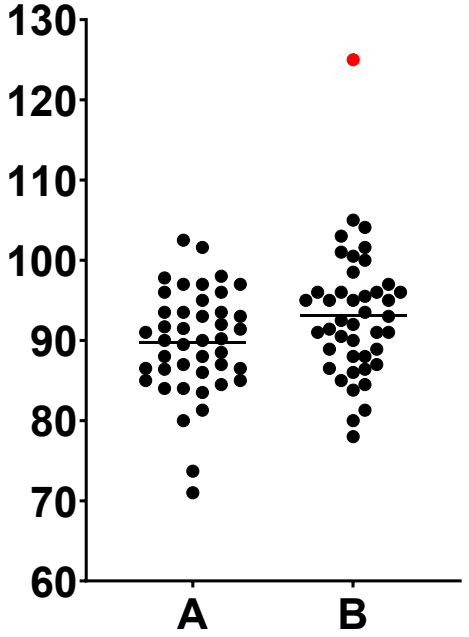
# Analysis of Quantitative data
# Non parametric statistics

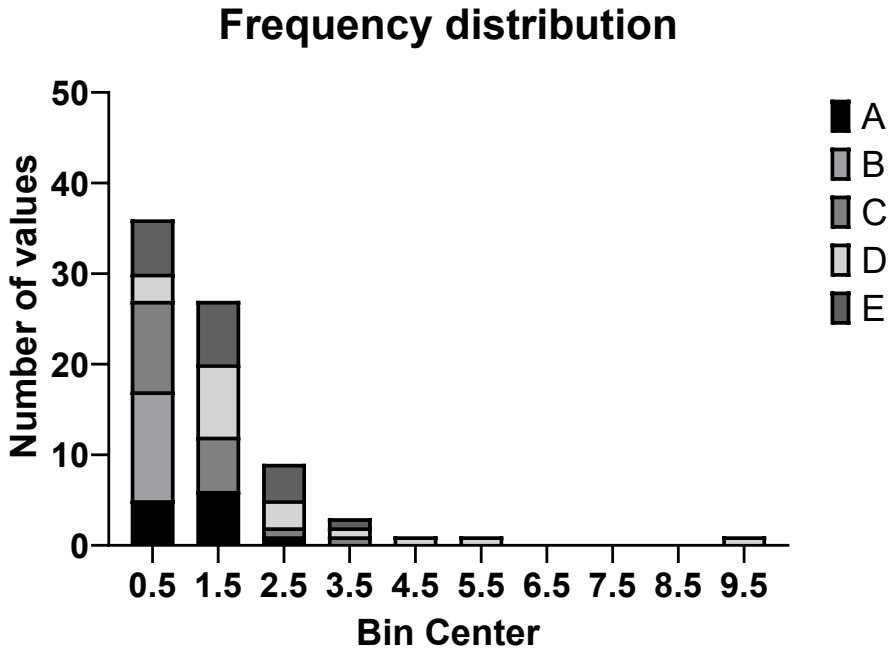
Hayley Carr & Anne Segonds-Pichon
v2024-05

Babraham Bioinformatics

# Non-parametric tests

- General principle: original data are transformed into **ranks**
- **Not meeting the assumptions for parametric tests is not enough** to switch to a non-parametric approach
- Data exploration is key:
  - Outliers?
  - Possible transformation?
  - Parametric with corrections?
- If outcome is a rank or a score with limited possible values: often non-parametric approach



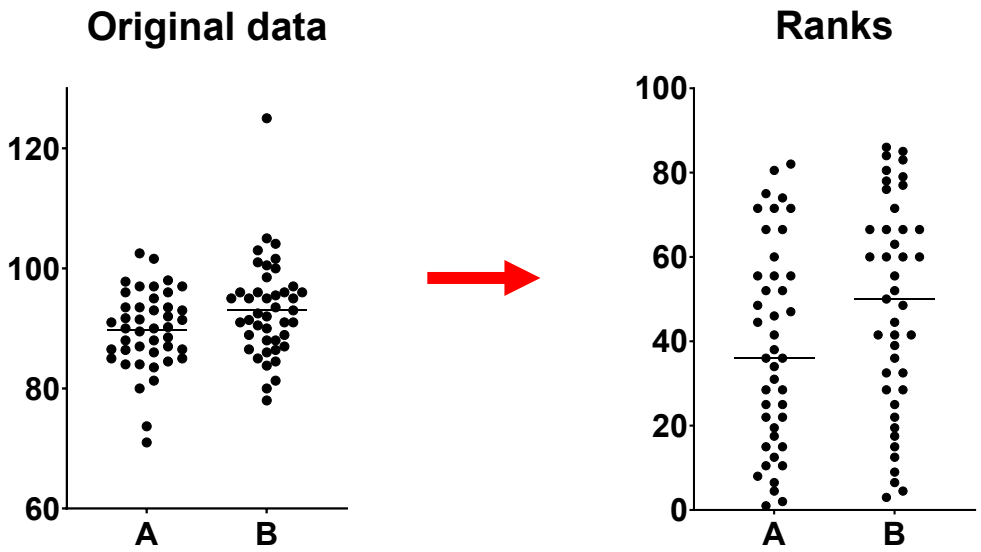



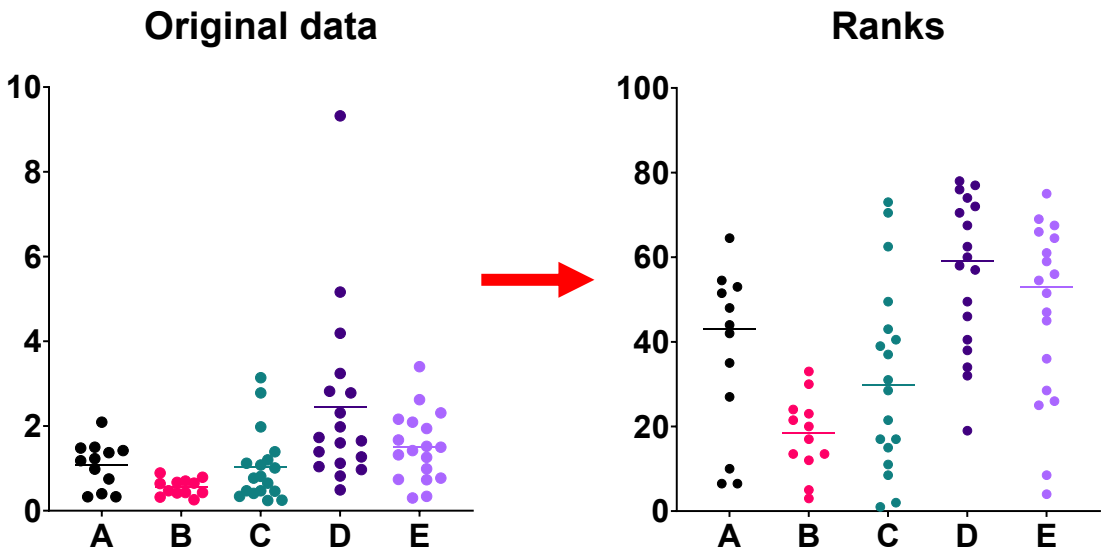
# Non-parametric tests

- General principle: **original data** are transformed into **ranks**

- Beware of misinterpretation: distribution of the data
  - Distributions = symmetrical and similar → compares means
  - Distributions = similar → compares medians
  - Distributions = not similar → compares distributions (though not always)

- A correction is applied when there are ties



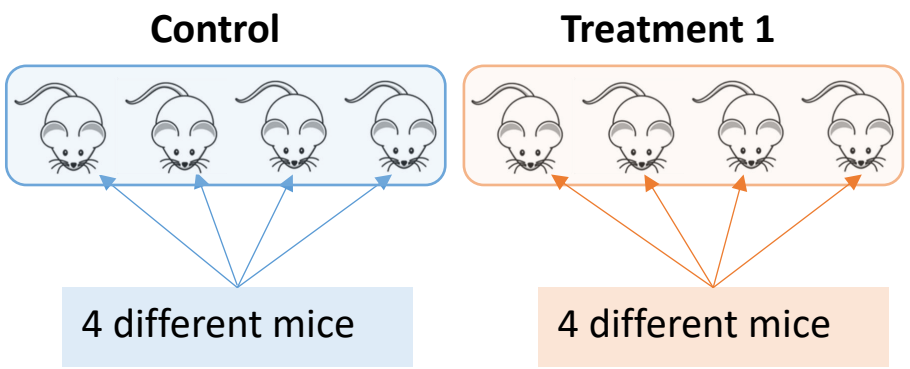
# Comparison between 2 groups

## Non-Parametric data

# Comparison between 2 independent groups

## Mann-Whitney *U* test

- Non-parametric equivalent of the *t*-test (and not)
- In the case of **inequality of variance** (violation of the homoscedasticity assumption), the 'unequal' version of the t-test is a possibility: **Welch's t-test**
- For a correct interpretation of the test: **Data exploration!**
- **Mann-Whitney *U* test** (Mann–Whitney–Wilcoxon, Wilcoxon rank-sum test or Wilcoxon–Mann–Whitney)
  - Wilcoxon: equal sample size
  - Mann and Whitney: different sample size



# Comparison between 2 independent groups

## Mann-Whitney *U* test

- **How does the Mann-Whitney *U* test work?**

| Group 1 | Group 2 |
|---|---|
| 5 | 8 |
| 7 | 9 |
| 3 | 6 |

| Real values | Ranks |
|---|---|
| 3 | 1 |
| 5 | 2 |
| 6 | 3 |
| 7 | 4 |
| 8 | 5 |
| 9 | 6 |

| | Group 1 | Group 2 |
|---|---|---|
| | 2 | 5 |
| | 4 | 6 |
| | 1 | 3 |
| Sum | $R_1$=7 | $R_2$=14 |

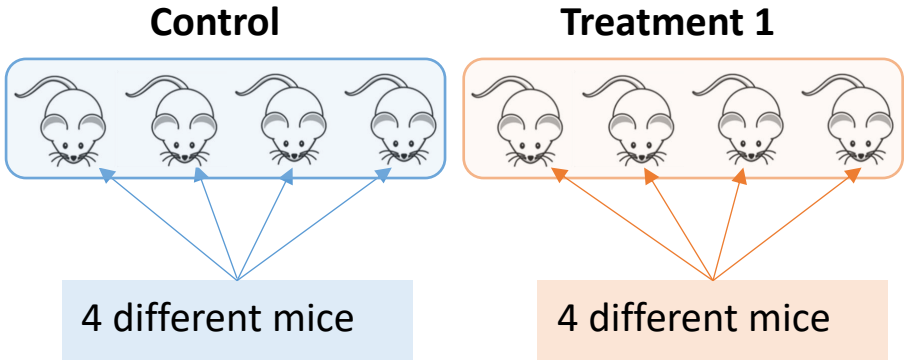
$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}$$

$$U_2 = R_2 - \frac{n_2(n_2+1)}{2}$$

**Where**:
- R = sum of ranks
- n = sample size

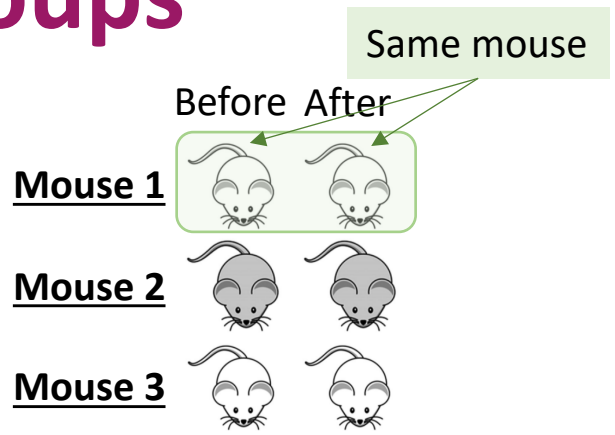
- Statistic of the Mann-Whitney test: **U (W)**

  **$U_1$** = 7-6 = 1 and **$U_2$** = 14-6 = 8
  - Smallest of the 2 Us: $U_1$
  - U1 comparison to critical value + sample size → **p-value**



# Comparison between 2 paired groups

## Wilcoxon's signed-rank test




- Non-parametric equivalent of the paired *t*-test (ish)
- Information about the Mann-Whitney test also applies
- **How does the Wilcoxon's signed-rank test work?**

2+3=5/2=2.5: average rank

| Before | After | Differences |
|---|---|---|
| 9 | 3 | -6 |
| 7 | 4 | -3 |
| 10 | 4 | -6 |
| 8 | 5 | -3 |
| 5 | 6 | 1 |
| 8 | 2 | -6 |
| 7 | 7 | 0 |
| 9 | 4 | -5 |
| 10 | 5 | -5 |

| Abs. Diff. | Ranking | Ranks |
|---|---|---|
| 0 | | |
| 1 | 1 | 1 |
| 3 | 2 | 2.5 |
| 3 | 3 | 2.5 |
| 5 | 4 | 4.5 |
| 5 | 5 | 4.5 |
| 6 | 6 | 7 |
| 6 | 7 | 7 |
| 6 | 8 | 7 |

| | Negative ranks | Positives ranks |
|---|---|---|
| | | 1 |
| | -2.5 | |
| | -2.5 | |
| | -4.5 | |
| | -4.5 | |
| | -7 | |
| | -7 | |
| | -7 | |
| Sum | -35 | 1 |

- Statistic of the Wilcoxon's signed-rank test: Sum of signed ranks = **W**
  - Here: W = -35 + 1 = -34
  - Statistic W + sample size → **p-value**

# Comparison between more than 2 groups

## One factor

## Non-Parametric data

# Non-parametric tests

## Kruskal-Wallis and Friedman tests

- Non-parametric equivalent of the One-Way ANOVA (ish)
  - Data replaced by ranks
  - **Data exploration**
    - If data represent different distributions: comparison of said distributions
    - If original data come from similar distributions: comparison of the medians

- **Kruskal-Wallis**: independent measures
  - Statistic = **H**
- **Friedman**: repeated measures
  - Statistic = **Q** or **T1** or **FM**

- Post-hoc test associated with Kruskal-Wallis and Friedman: **Dunn's test**
  - Works pretty much like the Mann-Whitney test

# Comparison between more than 2 groups

## Independent: Kruskal-Wallis test

Actual values: n=15

| No | Once | Twice |
|---|---|---|
| 63 | 0 | 2239 |
| -261 | -652 | 171 |
| -153 | 4724 | 40 |
| -13 | -2 | 1395 |
| 965 | 0 | |
| | -86 | |

Ranks: 15

| No | Once | Twice |
|---|---|---|
| 10 | 7.5 | 14 |
| 2 | 1 | 11 |
| 3 | 15 | 9 |
| 5 | 6 | 13 |
| 12 | 7.5 | |
| | 4 | |
| **32** | **41** | **47** |

$$\mathbf{H} = \left[\frac{12}{n(n+1)}\sum_{j=1}^{c}\frac{T_j^2}{n_j}\right] - 3(n+1)$$

$$\mathbf{H} = \left[\frac{12}{15(15+1)}\left(\frac{32^2}{5}+\frac{41^2}{6}+\frac{47^2}{4}\right)\right] - 3(15+1) = 3.868$$

**Where**:

- n = total sample size across all groups
- c = number of groups
- $T_j$ = sum of ranks in the ($j^{th}$) group
- $n_j$ = size of the ($j^{th}$) group

Interpretation of the test: H + degrees of freedom = p-value

# Comparison between more than 2 groups

## Matched/repeated: Friedman test

Actual values

| Violinists | Violin A | Violin B | Violin C |
|---|---|---|---|
| 1 | 9 | 7 | 6 |
| 2 | 9.5 | 6.5 | 8 |
| 3 | 5 | 7 | 4 |
| 4 | 7.5 | 7.5 | 6 |
| 5 | 9.5 | 5 | 7 |
| 6 | 7.5 | 8 | 6.5 |
| 7 | 8 | 6 | 6 |
| 8 | 7 | 6.5 | 4 |
| 9 | 8.5 | 7 | 6.5 |
| 10 | 6 | 7 | 3 |

Ranks

Matched set of values

| Violinists | Violin A | Violin B | Violin C |
|---|---|---|---|
| 1 | 3 | 2 | 1 |
| 2 | 3 | 1 | 2 |
| 3 | 2 | 3 | 1 |
| 4 | 2.5 | 2.5 | 1 |
| 5 | 3 | 1 | 2 |
| 6 | 2 | 3 | 1 |
| 7 | 3 | 1.5 | 1.5 |
| 8 | 3 | 2 | 1 |
| 9 | 3 | 2 | 1 |
| 10 | 2 | 3 | 1 |
| Sum | $R_A$ = 26.5 | $R_B$=21 | $R_C$=12.5 |

- Basic idea: if the sums are very different (here $R_A$ $R_B$ and $R_C$) the p-value will be small.

$$Q\ or\ T1\ or\ FM\ or\ F = \left[\frac{12}{N \times k \times (k+1)}\right] \times \sum R^2 - [3 \times N \times (k+1)]$$

$$F = \left[\frac{12}{10 \times 3 \times (3+1)}\right] \times [26.5^2 + 21^2 + 12.5^2] - [3 \times 10 \times (3+1)]$$

$$F = \left[\frac{12}{120}\right] \times [702.25 + 441 + 156.25] - 120 = 9.95$$

**Where**:

- N = the number of subjects (violinists)
- k = number of groups (violins)
- R = sum of ranks in the group (e.g. $R_A$)

Interpretation of the test: Q or T1 or FM + df= p-value

# Association between 2 continuous variables

## Linear relationship

## Non-Parametric data

# Non-parametric tests

## Spearman Correlation Coefficient

- Similar concepts as for the other non-parametric tests
- **ρ (rho)** is the equivalent of r and calculated in a similar way
- **Spearman's ρ** is Pearson's **r** applied on ranks

$$\rho = r_s = \frac{\text{Similarity}}{\text{Variability}} = \frac{COV_{R(x)R(y)}}{SD_{R(x)}SD_{R(y)}}$$

# Exercise 5

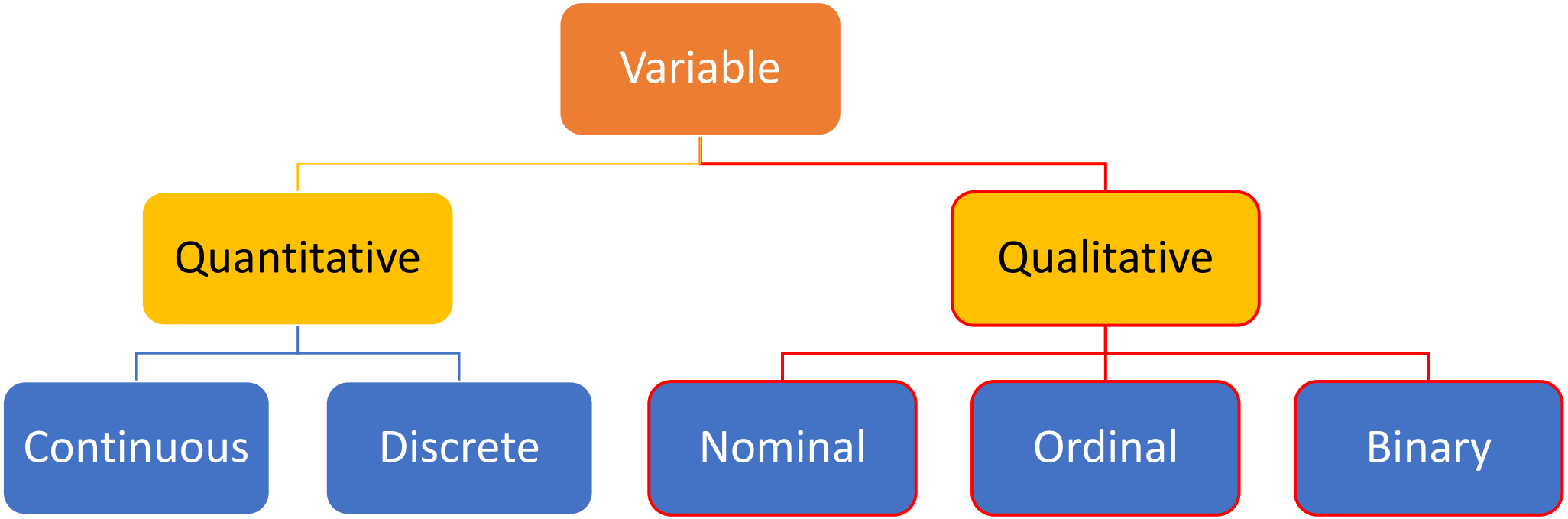
Babraham Institute

# Analysis of Qualitative data

Hayley Carr & Anne Segonds-Pichon

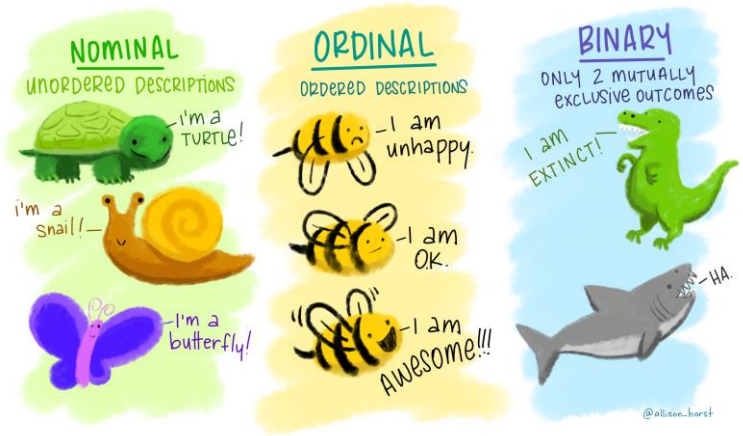
v2025-02

Babraham Bioinformatics



https://github.com/allisonhorst/stats-illustrations#other-stats-artwork

# Qualitative data




https://github.com/allisonhorst/stats-illustrations#other-stats-artwork

- Values taken = usually names (also nominal)
    - e.g. genotypes
- Can be numbers but not numerical
    - e.g. group number = numerical label but not unit of measurement
- Qualitative variable with intrinsic order in their categories = ordinal
    - e.g. low/medium/high
- Particular case: qualitative variable with 2 categories: binary or dichotomous
    - e.g. alive/dead or presence/absence

# Comparison between 2 groups

## Comparison between 2 proportions

## Binary outcome

# Chi-square and Fisher's tests

- $Chi^2$ is an approximation
- $Chi^2$ test very easy to calculate by hand but Fisher's very hard
- Often software will not perform a Fisher's test on tables > 2x2

- **Fisher's test** more accurate than $Chi^2$ test on **small samples**
- **$Chi^2$ test** generally preferable on **large samples**

- $Chi^2$ test assumptions:
  - 2x2 table: no expected count <5
  - Bigger tables: all expected > 1 and no more than 20% < 5

# Chi-square test

- In a chi-square test, the observed frequencies for two or more groups are compared with expected frequencies by chance

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

- O = Observed frequencies
- E = Expected frequencies

# Fisher's exact and Chi$^2$ tests

## Example: cats and dogs.xlsx

- Cats and dogs trained to line dance
- 2 different rewards: food or affection
- **Question**: Is there a difference between the rewards?
  - Is there a significant relationship between the 2 variables?
  - Does the reward significantly affect the likelihood of dancing?

- To answer this type of question:
  - Contingency table
  - Fisher's exact or Chi$^2$ tests

| | Food | Affection |
|---|---|---|
| Dance | ? | ? |
| No dance | ? | ? |

- But first: **how many animals** do we need?
  - Power analysis

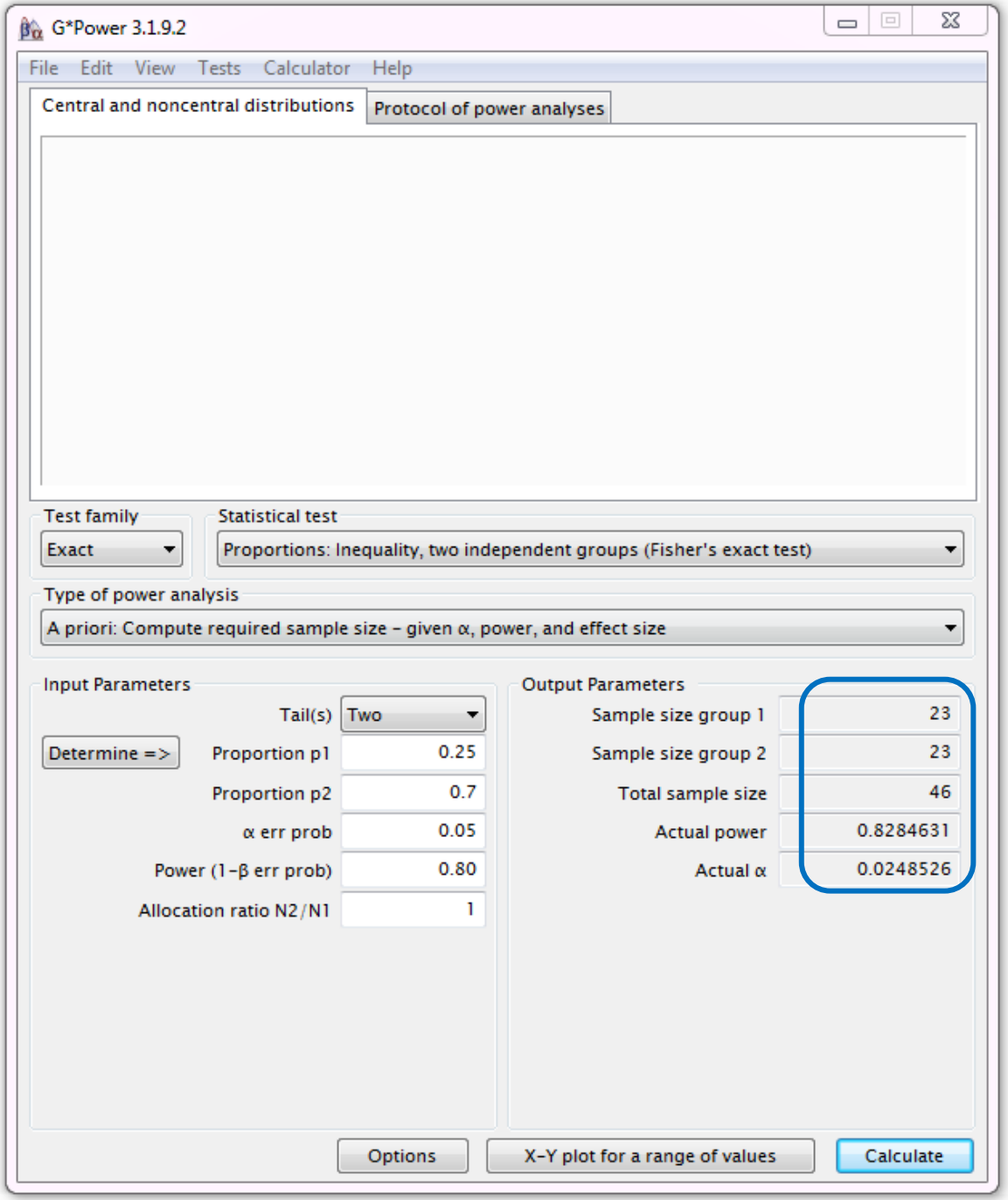
# Exercise: Power calculation

- Preliminary results from a pilot study: **25%** of cats line-danced after having received affection as a reward vs. **70%** after having received food
  - **How many cats** do we need?

# Exercise: Power calculation

**Pilot study:**
**25%** of cats line-danced after having received affection as a reward vs. **70%** after having received food

Output:
If the values from the pilot study are good predictors and if we use a sample of **n=23 for each group**, we will achieve a power of 83%



# How are the expected frequencies calculated?

**Example**: expected frequency of cats line dancing after having received food as a reward

**Direct counts approach**:

Expected frequency

= (row total)*(column total)/grand total

= **32*32/68** = **15.1**

**Probability approach**: The Multiplicative Rule

Probability of line dancing: **32/68**

Probability of receiving food: **32/68**

Expected frequency:(**32/68**)*(**32/68**)=0.22: **22% of 68 = 15.1**

Observed frequencies

| | Food | Affection | Total |
|---|---|---|---|
| Dance | **26** | 6 | **32** |
| No dance | 6 | 30 | 36 |
| Total | **32** | 36 | **68** |

Expected frequencies

| | Food | Affection |
|---|---|---|
| Dance | **15.1** | 16.9 |
| No dance | 16.9 | 19.1 |

# Chi$^2$ test

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Observed frequencies

| | Food | Affection |
|---|---|---|
| Dance | 26 | 6 |
| No dance | 6 | 30 |

Expected frequencies

| | Food | Affection |
|---|---|---|
| Dance | 15.1 | 16.9 |
| No dance | 16.9 | 19.1 |

Chi$^2$ = $(26-15.1)^2/15.1 + (6-16.9)^2/16.9 + (6-16.9)^2/16.9 + (30-19.1)^2/19.1 = \mathbf{28.4}$

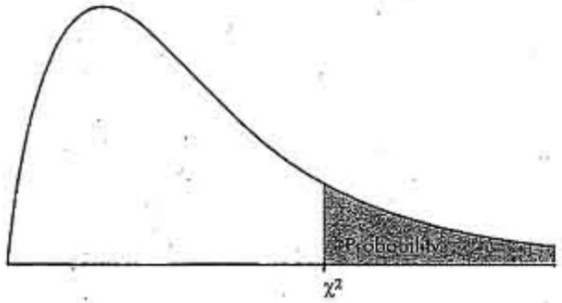
**Is 28.4 big enough for the test to be significant?**

# Is 28.4 big enough for the test to be significant?

*The old fashion way*

Degree of freedom: df

df = (row-1)(col-1)=1




Critical value

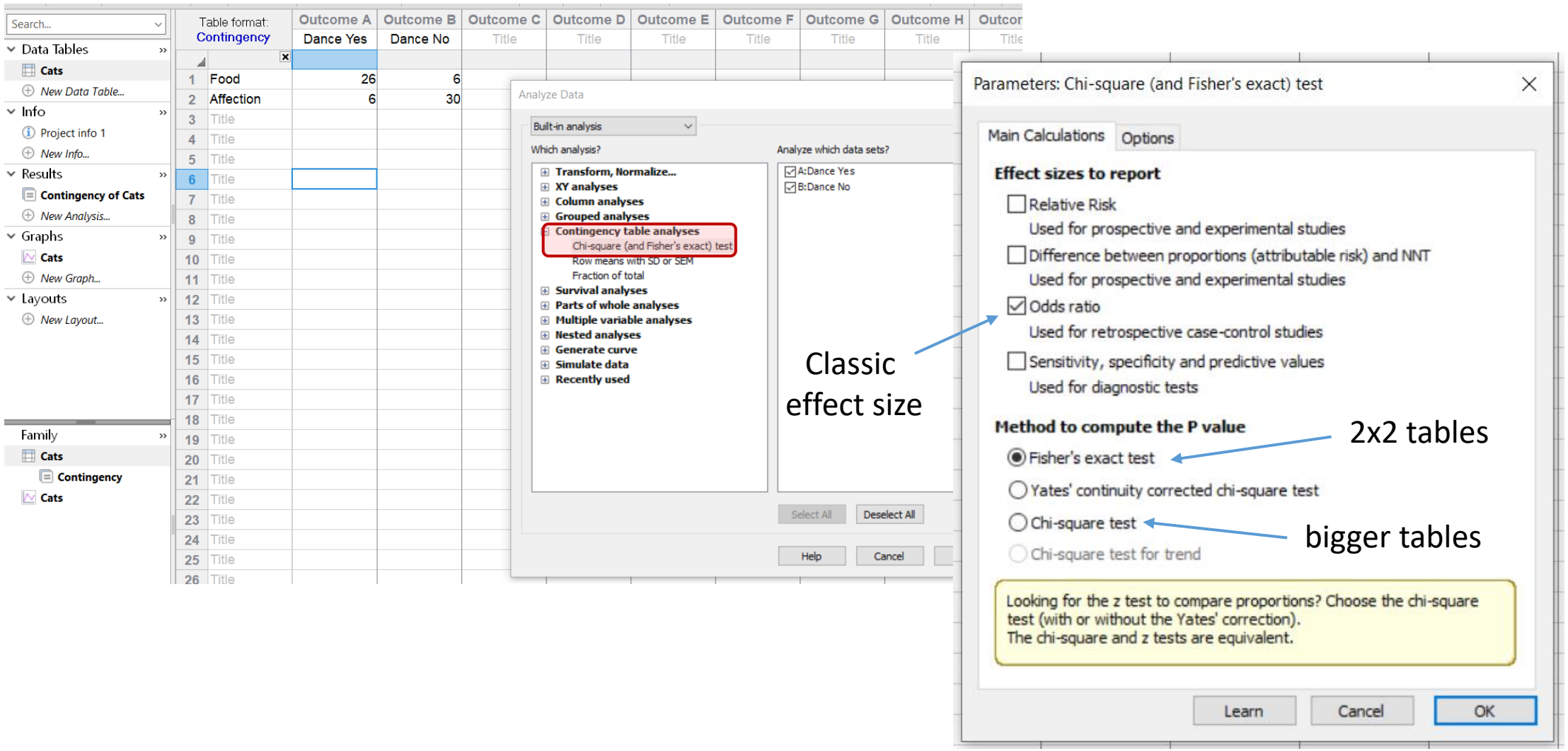
| | Food | Affection |
|---|---|---|
| Dance | 26 | 6 |
| No dance | 6 | 30 |

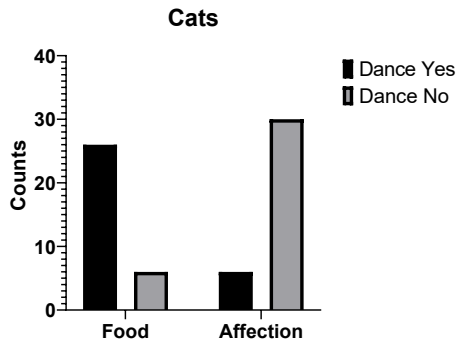
**TABLE C: $\chi^2$ CRITICAL VALUES**

| | Tail probability $p$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| df | .25 | .20 | .15 | .10 | .05 | .025 | .02 | .01 | .005 |
| 1 | 1.32 | 1.64 | 2.07 | 2.71 | 3.84 | 5.02 | 5.41 | 6.63 | 7.88 |
| 2 | 2.77 | 3.22 | 3.79 | 4.61 | 5.99 | 7.38 | 7.82 | 9.21 | 10.60 |
| 3 | 4.11 | 4.64 | 5.32 | 6.25 | 7.81 | 9.35 | 9.84 | 11.34 | 12.84 |
| 4 | 5.39 | 5.99 | 6.74 | 7.78 | 9.49 | 11.14 | 11.67 | 13.28 | 14.86 |
| 5 | 6.63 | 7.29 | 8.12 | 9.24 | 11.07 | 12.83 | 13.39 | 15.09 | 16.75 |
| 6 | 7.84 | 8.56 | 9.45 | 10.64 | 12.59 | 14.45 | 15.03 | 16.81 | 18.55 |
| 7 | 9.04 | 9.80 | 10.75 | 12.02 | 14.07 | 16.01 | 16.62 | 18.48 | 20.28 |
| 8 | 10.22 | 11.03 | 12.03 | 13.36 | 15.51 | 17.53 | 18.17 | 20.09 | 21.95 |
| 9 | 11.39 | 12.24 | 13.29 | 14.68 | 16.92 | 19.02 | 19.68 | 21.67 | 23.59 |
| 10 | 12.55 | 13.44 | 14.53 | 15.99 | 18.31 | 20.48 | 21.16 | 23.21 | 25.19 |

**$\chi^2$ = 28.4 > 3.84 so Yes!**

# Fisher's exact and Chi² tests with Prism

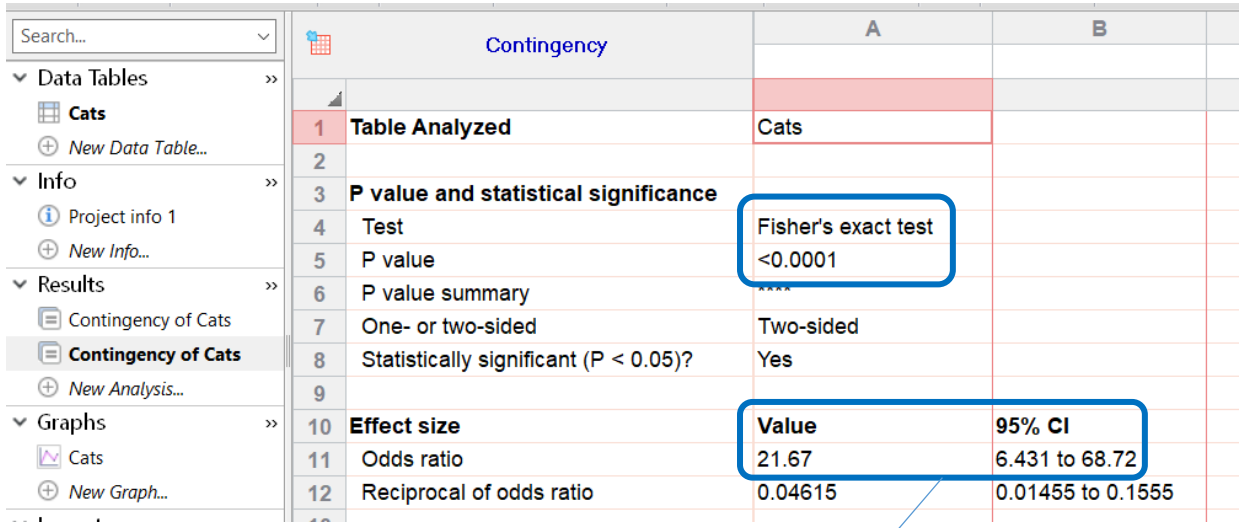





# Fisher's exact and Chi² tests

## Results: Cats

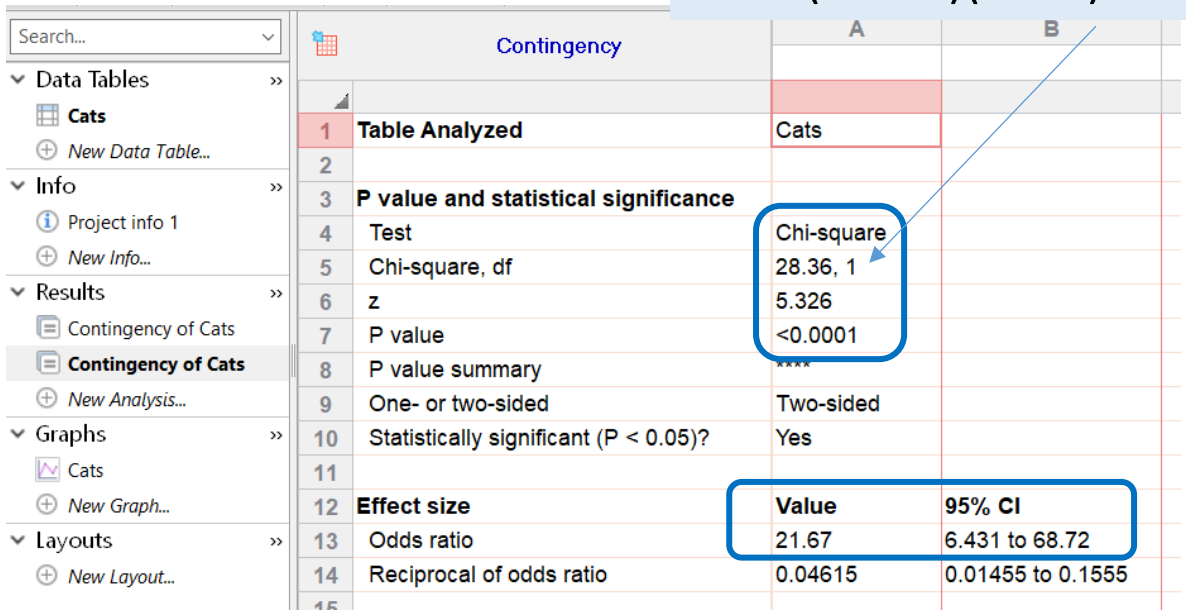
Degree of freedom: df
df = (row-1)(col-1)=1



Search...
Data Tables
Cats
New Data Table...
Info
Project info 1
New Info...
Results
Contingency of Cats
Contingency of Cats
New Analysis...
Graphs
Cats
New Graph...

Contingency

| | A | B |
|---|---|---|
| 1 Table Analyzed | Cats | |
| 2 | | |
| 3 P value and statistical significance | | |
| 4 Test | Fisher's exact test | |
| 5 P value | <0.0001 | |
| 6 P value summary | **** | |
| 7 One- or two-sided | Two-sided | |
| 8 Statistically significant (P < 0.05)? | Yes | |
| 9 | | |
| 10 Effect size | Value | 95% CI |
| 11 Odds ratio | 21.67 | 6.431 to 68.72 |
| 12 Reciprocal of odds ratio | 0.04615 | 0.01455 to 0.1555 |



Search...
Data Tables
Cats
New Data Table...
Info
Project info 1
New Info...
Results
Contingency of Cats
Contingency of Cats
New Analysis...
Graphs
Cats
New Graph...
Layouts
New Layout...

Contingency

| | A | B |
|---|---|---|
| 1 Table Analyzed | Cats | |
| 2 | | |
| 3 P value and statistical significance | | |
| 4 Test | Chi-square | |
| 5 Chi-square, df | 28.36, 1 | |
| 6 z | 5.326 | |
| 7 P value | <0.0001 | |
| 8 P value summary | **** | |
| 9 One- or two-sided | Two-sided | |
| 10 Statistically significant (P < 0.05)? | Yes | |
| 11 | | |
| 12 Effect size | Value | 95% CI |
| 13 Odds ratio | 21.67 | 6.431 to 68.72 |
| 14 Reciprocal of odds ratio | 0.04615 | 0.01455 to 0.1555 |

| | Food | Affection |
|---|---|---|
| Dance | 26 | 6 |
| No dance | 6 | 30 |

Odds of dancing on Food group = **26/6**

Odds of dancing on Affection group = **6/30**

O.R. = ratio of the odds = 26/6/6/30 = **21.7**

**Odds Ratio = 21.7**

If you are a dancing cat, you are almost 22 times more likely to have received food than affection as a reward (p<0.0001).

# Exercise 6

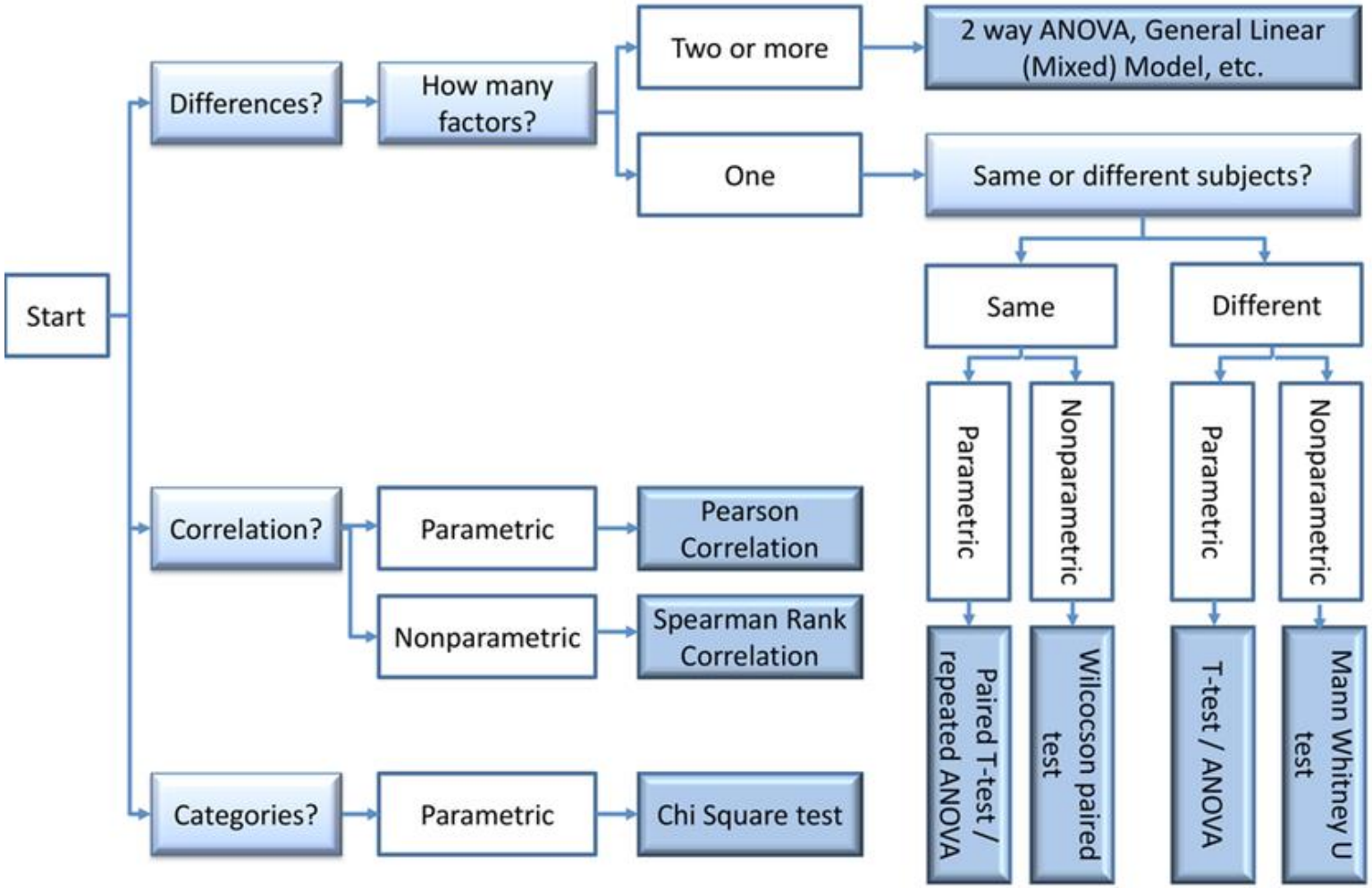
Babraham Institute

# Decision trees & resources

Hayley Carr

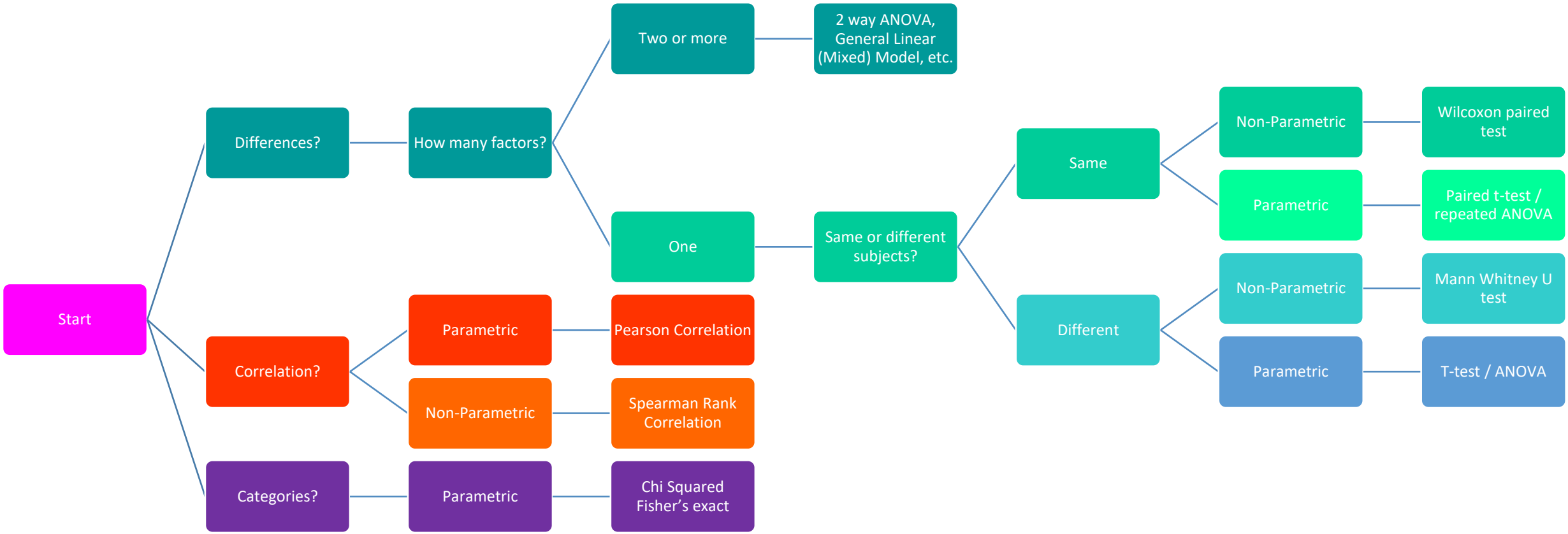
Babraham Bioinformatics

# Choosing a test: Flow charts



# Statistics Decision tree

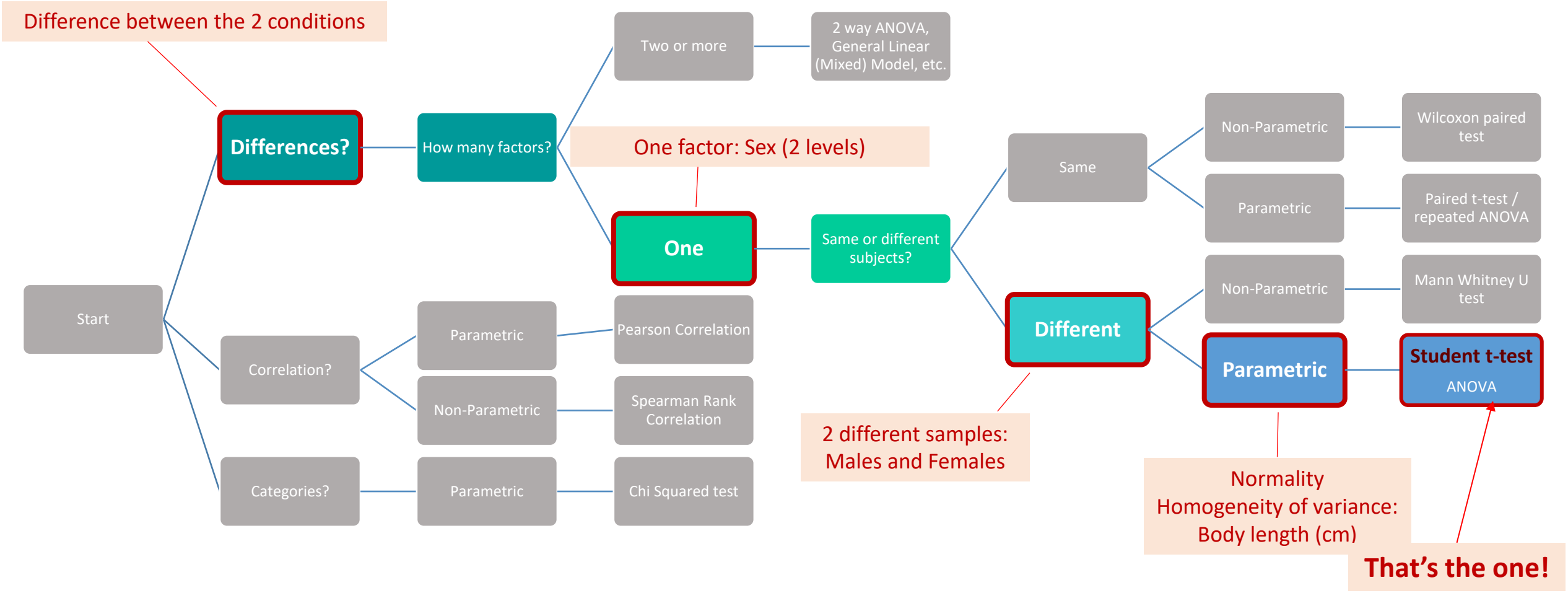
Anne Segonds-Pichon

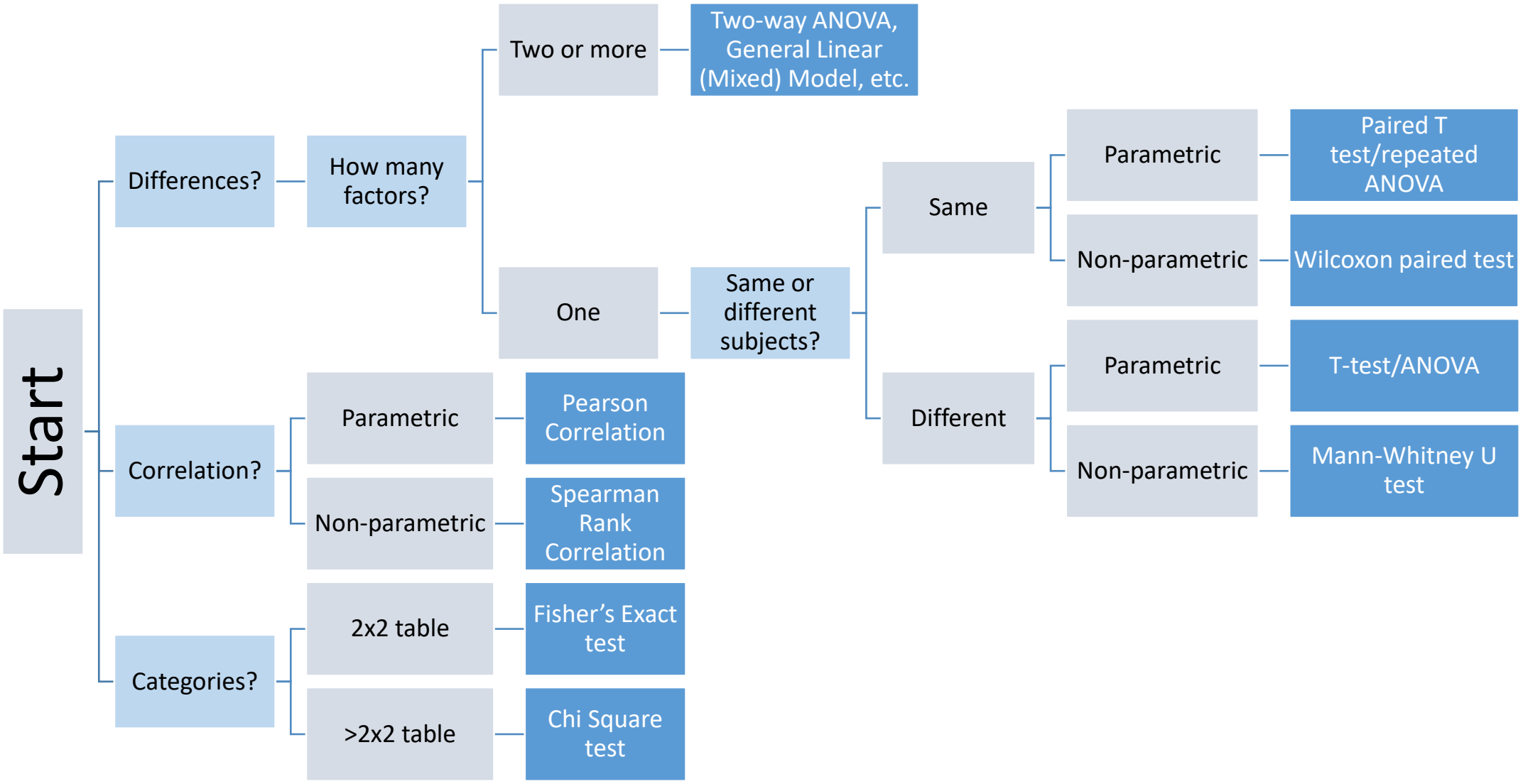


# Statistics Decision tree

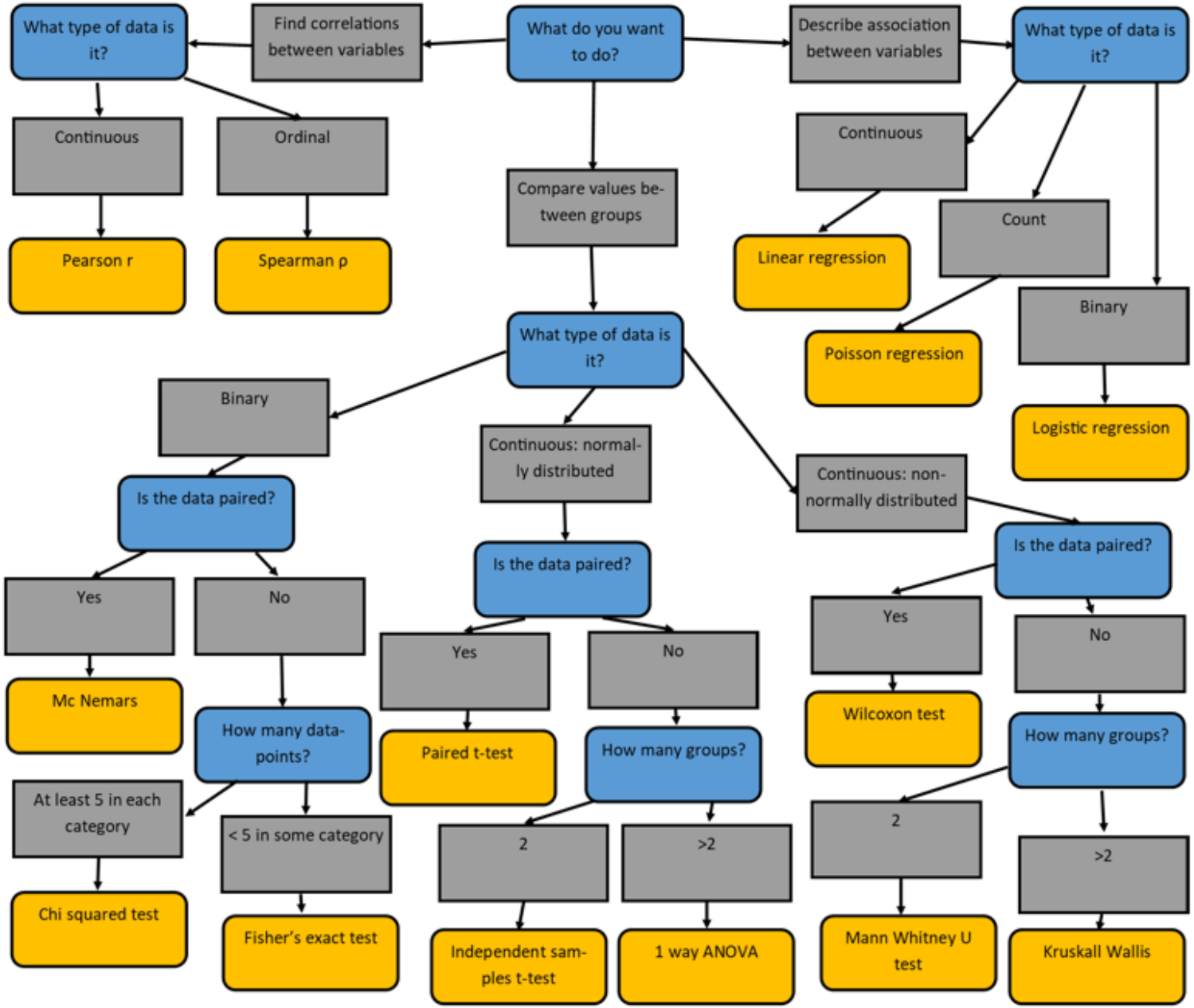
Anne Segonds-Pichon

Is there a difference between males and females coyotes in the body length?

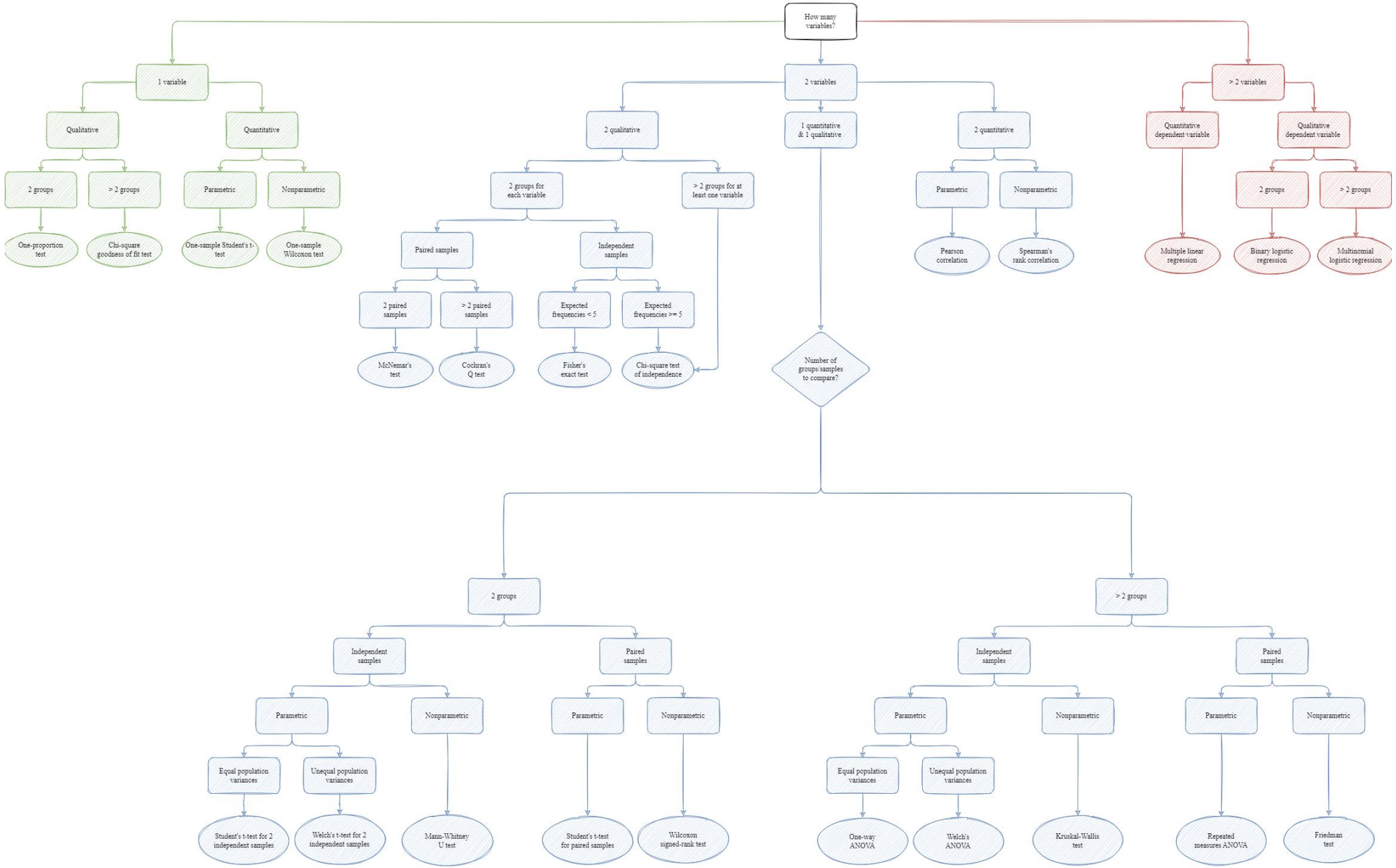



Start
Differences?
How many factors?
Two or more
Two-way ANOVA, General Linear (Mixed) Model, etc.
One
Same or different subjects?
Same
Parametric
Paired T test/repeated ANOVA
Non-parametric
Wilcoxon paired test
Different
Parametric
T-test/ANOVA
Non-parametric
Mann-Whitney U test
Correlation?
Parametric
Pearson Correlation
Non-parametric
Spearman Rank Correlation
Categories?
2x2 table
Fisher's Exact test
>2x2 table
Chi Square test

# Which statistical test should I use?



# What statistical test should I do?




@statsandr

https://statsandr.com/blog/what-statistical-test-should-i-do/

# Statistics resources

Hayley Carr (i.e. me!):
hayley.carr@babraham.ac.uk

GraphPad

Statistics Guide | Curve Fitting Guide

Welcome to the Prism 9 Statistics Guide

- Welcome to Prism 9 Statistics Guide
- Learn about analyses with Prism
- View the Prism User or Regression Guide
- How to cite these pages
- PRINCIPLES OF STATISTICS
- STATISTICS WITH PRISM 9

This guide examines general principles of statistical analysis, looks at how to conduct those analyses in Prism, and how to interpret results of these analyses.

**New to Statistical Analysis?**

The Principles of Statistics section of this guide will help you learn about a number of fundamental topics in statistics!

**Statistics in Prism**

Check out our Statistical analyses with Prism section of this guide to start analyzing your data and interpreting your results!

**More Guides!**

You're currently browsing the Prism Statistics Guide. We also have our User Guide, focused on getting to know Prism, and our Curve Fitting Guide! Both of these guides do more than important concepts about data analysis!

- Switch to the Prism 9 User Guide
- Switch to the Prism 9 Curve Fitting Guide

© 1995-2021 GraphPad Software, LLC. All rights reserved.

https://www.graphpad.com/guides/prism/latest/statistics/index.htm

https://www.nature.com/collections/qghhqm

*Not always the friendliest, but covers lots of relevant topics*

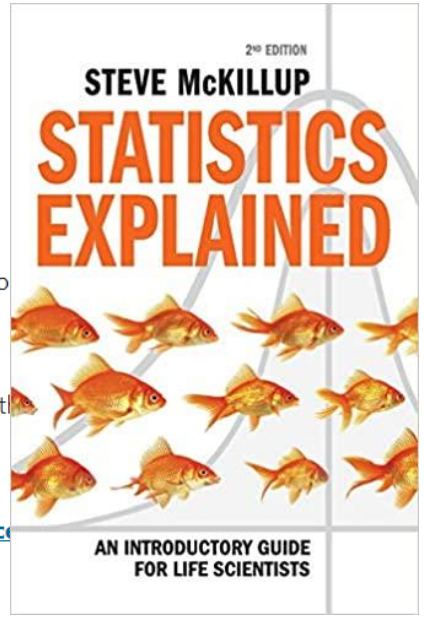
**Collection** | **09 May 2017**

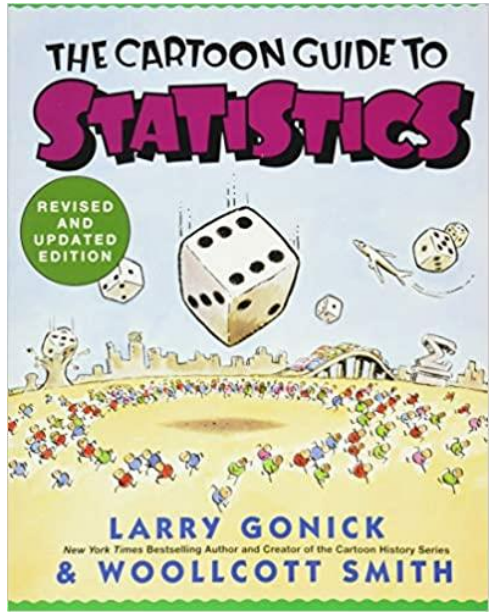
## Statistics for Biologists

There is no disputing the importance of statistical analysis in biological research, but too often it is considered only after an experiment is co
may be too late.

This collection highlights important statistical issues that biologists should be aware of and provides practical advice to help them improve t
work.

*Nature Methods'* **Points of Significance** column on statistics explains many key statistical and experimental design concepts. **Other resourc**
online plotting tool and links to statistics guides from other publishers.



# Exercise 7

# Help us improve!

# FEEDBACK FORM

## What did you think of this training? Thanks for letting us know in this short 10 question evaluation form

https://forms.office.com/e/tK16t77Jbd

Babraham Bioinformatics

Babraham Institute

# COURSE DATES

Babraham Bioinformatics

Babraham Institute

## Did You Know...

**We develop and deliver a wide range of Bioinformatics, Statistics and Scientific Research training courses**

bioinformatics.babraham.ac.uk/training.html