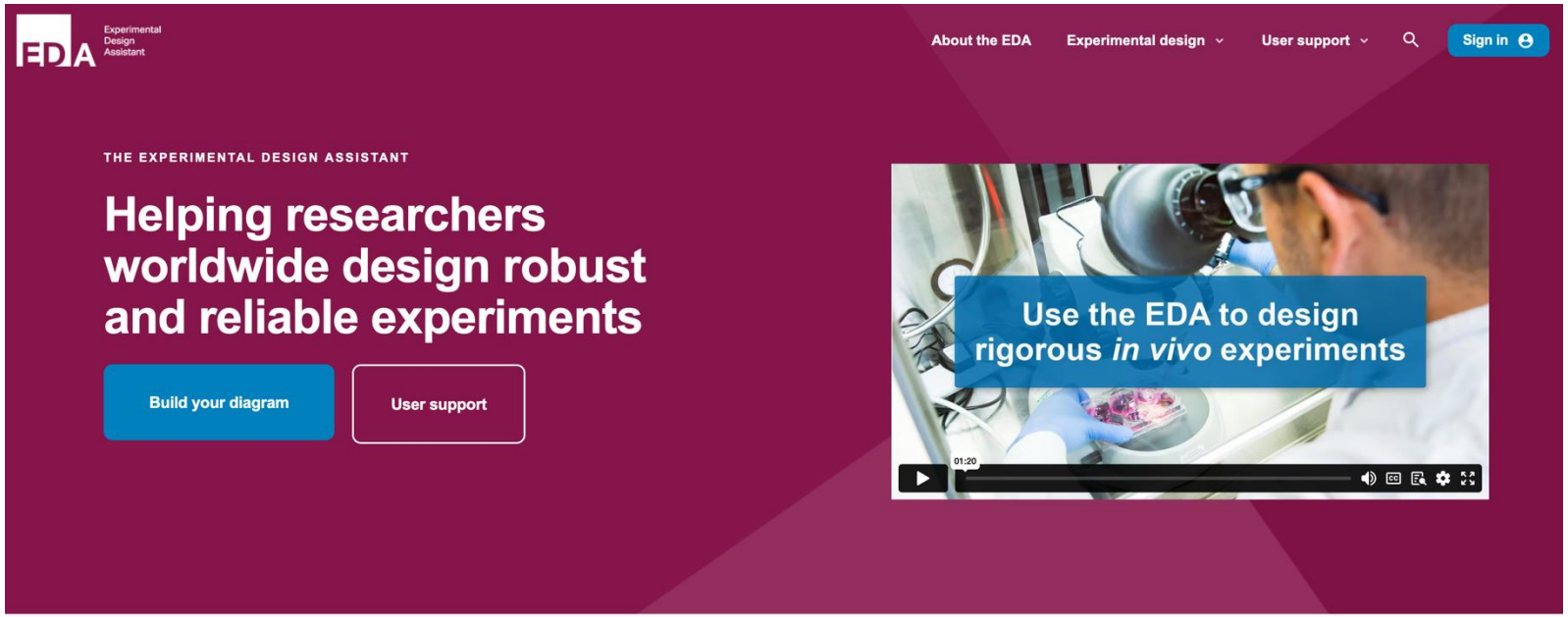
MP968 Workshop: Experimental Design

Dr Leighton Pritchard

leighton.Pritchard@strath.ac.uk

Register for NC3Rs EDA



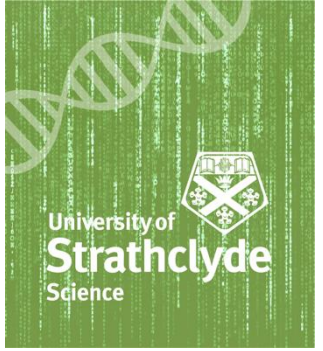
https://eda.nc3rs.org.uk/

https://sipbs-compbiol.github.io/MP968-Workshop_Experimental_Design

Learning Objectives

After this workshop I hope you will be able to…

1. Understand and explain the relationship between measurements, statistics, and experimental design.
2. Understand and explain that statistical models are devices that process data to produce estimates that support scientific insight.
3. Understand how assumptions and expectations about factors influencing an experiment are translated into effective experimental designs.
4. Use G*Power to estimate adequately-powered sample size for a statistical test.
5. Use the NC3Rs Experimental Design Assistant to lay out and analyse a simple experiment and share the design information.



Introduction

Experimental design and Statistics for people who would rather not be doing Statistics

What we'd all rather be doing

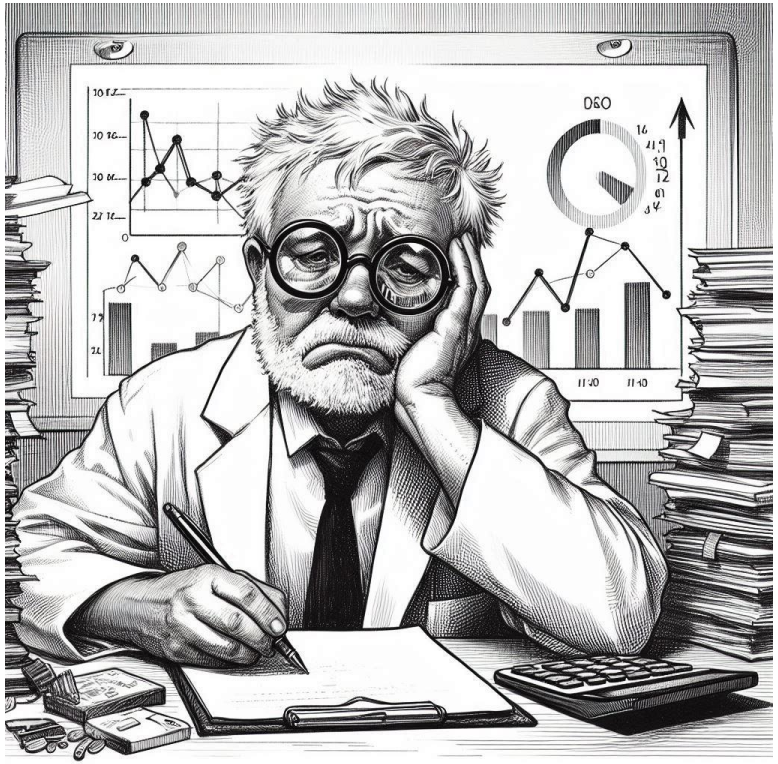
- **We are interested in the world around us**
  - Discovering how life and nature work
  - Finding ways to improve people's lives/reduce disease burden

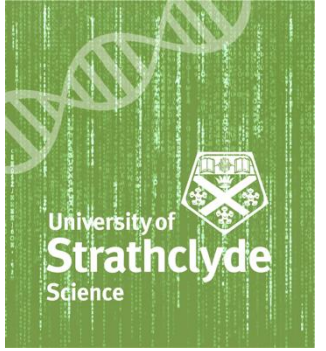
- **What we have to do**
  - Collect data
  - Design experiments to give us useful data
  - Design experiments that do not cause unnecessary suffering (3Rs)
  - Carry out appropriate statistical analysis to make the data scientifically useful

The case for Statistics

- We have models of the world around us
  - Our understanding of cause and effect is a model
  - We express our understanding with scientific models
  - Our experiments are models

- We connect data to models with Statistics
  - A branch of Applied Mathematics
  - How we understand our experiments guides statistical analysis
  - We can make Statistics serve scientific enquiry





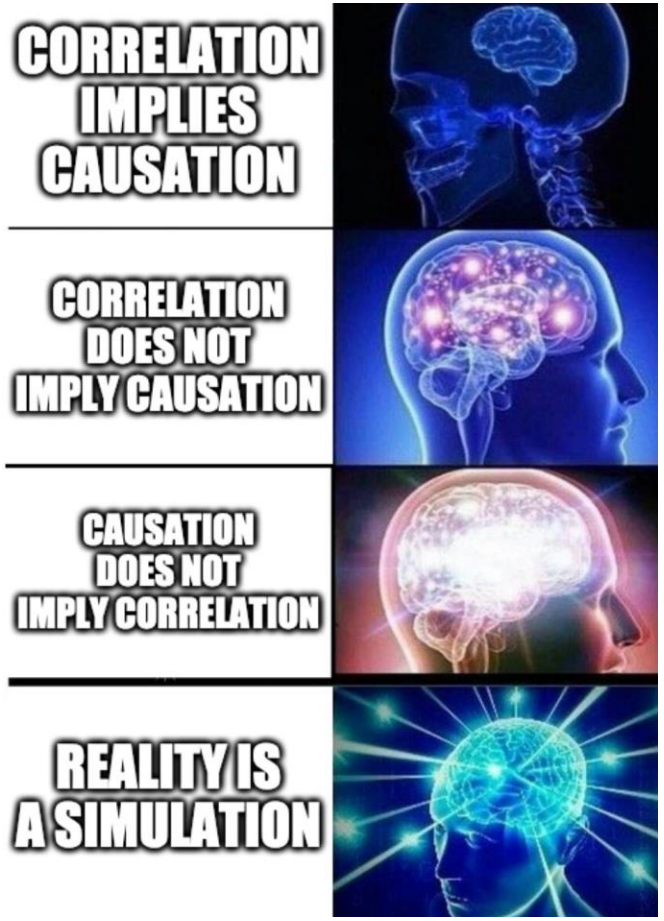
Cause and Effect

The experimental data you collect does not contain information about cause and effect

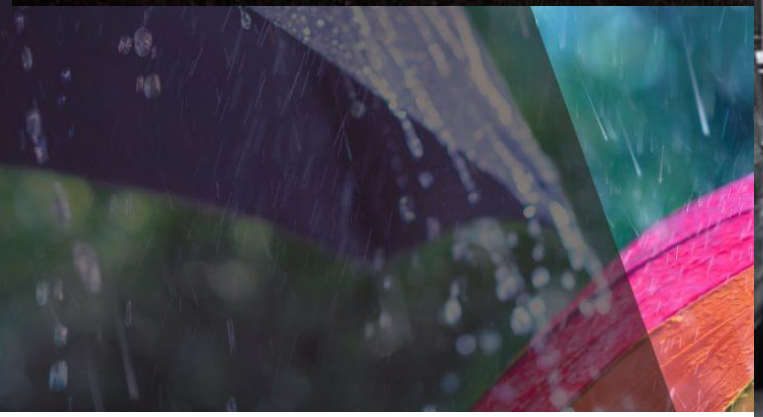
Science before Statistics

- For **statistical models** to provide scientific insight, they require additional **scientific (causal) models**.
- The **reasons** for a statistical analysis are not found in the data itself – they are found in the **causes** of the data.
- The **causes** of the data cannot be extracted from the data alone.
- We have to design the right experiment and statistical model to give us the estimate we need.

Causal Inference

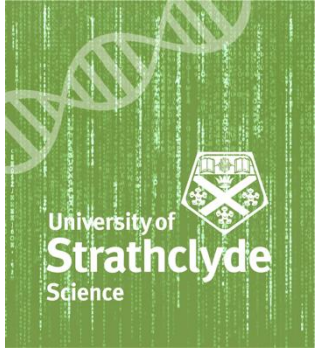


- **Correlation is not causation**
  - https://www.tylervigen.com/spurious-correlations
- Statistical analysis identifies **associations**, which can run in **either direction**
- Causal inference is **prediction** of **consequences of intervention**
  - "If I increase drug dose, blood glucose falls further"
- Causal inference is **imputation** of **unobserved interventions**
  - "I observed a change in marker level, so kidney function is altered"



University of Strathclyde Science

Causal Beliefs are Experimental Design



Measurement

Experimental measurements are estimates, and experimental design affects how good that estimate is

How wet is the Earth?

- **What proportion of the Earth's surface is water?**
- Can we measure this exactly, or only **estimate** it?
- If we're estimating, how **accurate** is our estimate, and **what affects the accuracy**?

How to measure the Earth's surface?

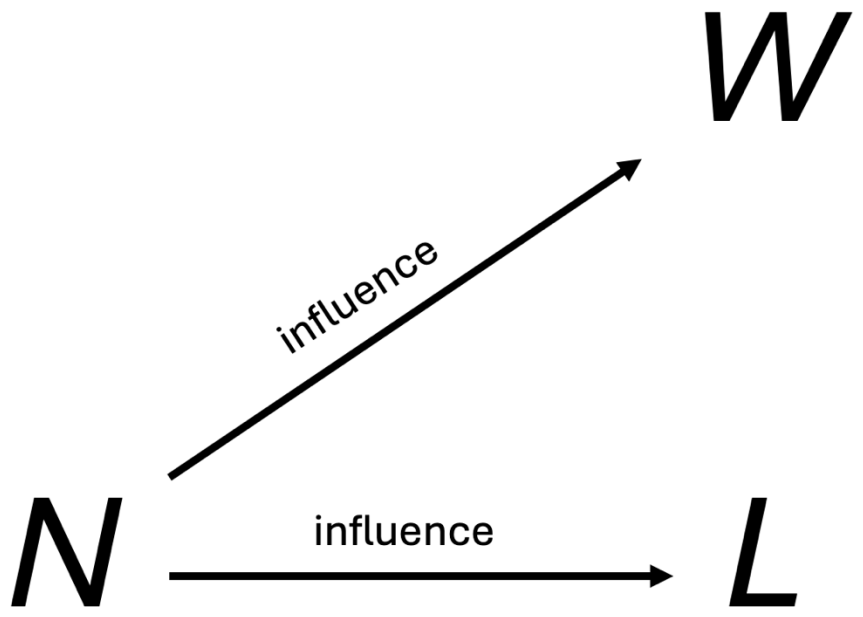
- Pour the water into a beaker?
- Measure all coastlines with a ruler?
  - In person?
  - In satellite photos?

- **We can't get an exact measure, so we need to estimate**

- **Let's spin the globe!** (An experiment!)

What affects our measurements?

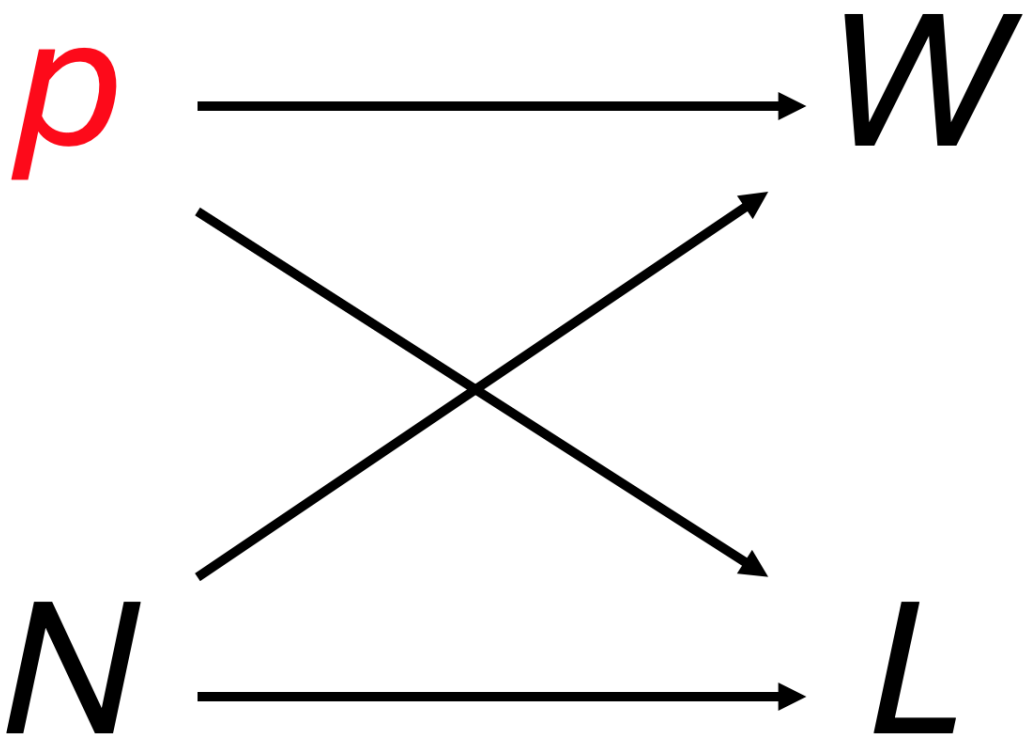
- **We measure the number of times we hit "water" (*W*), and how many times we hit "land" (*L*)**
  - These are our measured outcomes – *dependent variables*
- The number of times we spin the globe and make a measurement, *N*
  - *N* is an independent variable
- The more times we spin the globe, the larger *W* and *L* can be
- Represent this with a ***causal graph***



What affects our measurements?

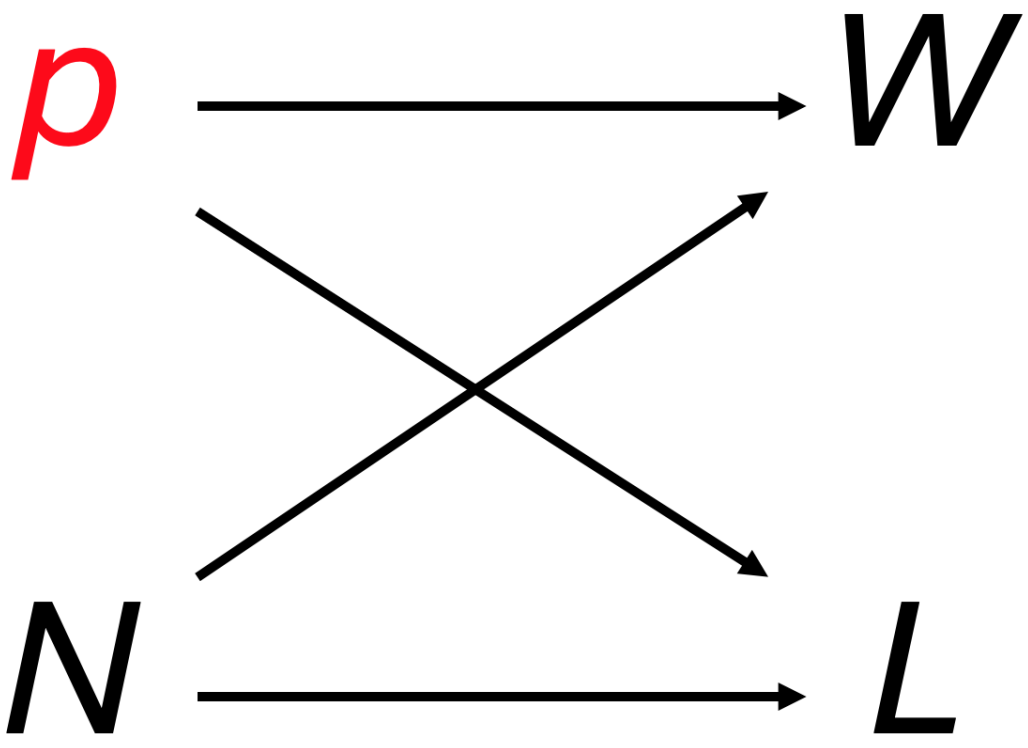
- *p*, our ***estimand***, is a variable representing the proportion of the globe that is covered with water
  - We **cannot directly observe** this value.
  - We can only measure/estimate it indirectly

- If *p* is large (more water on the Earth) we expect *W* to be large
- If *p* is small (less water on the Earth) we expect *L* to be large

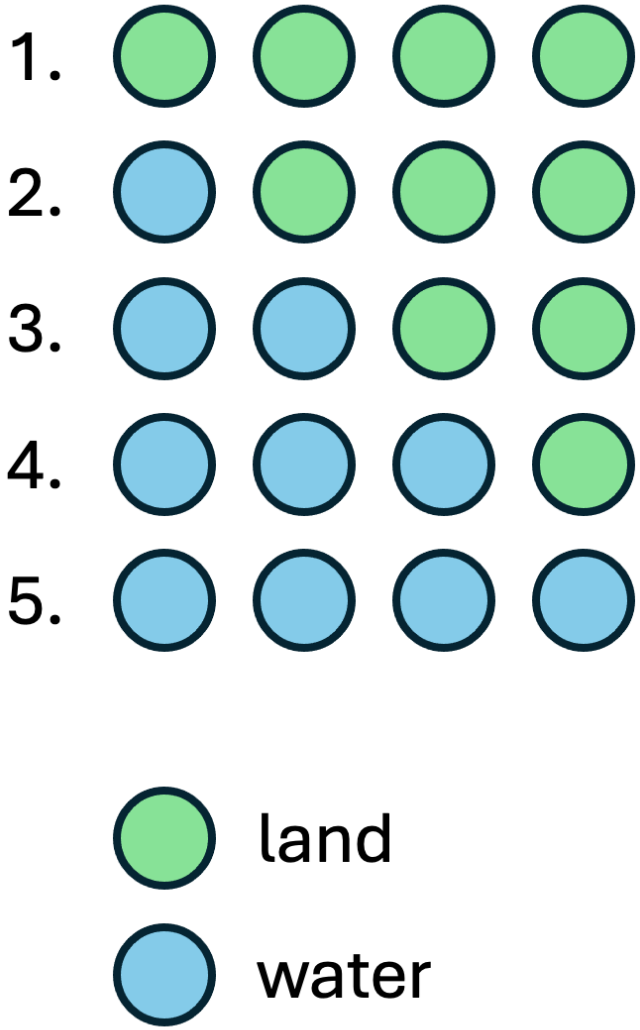


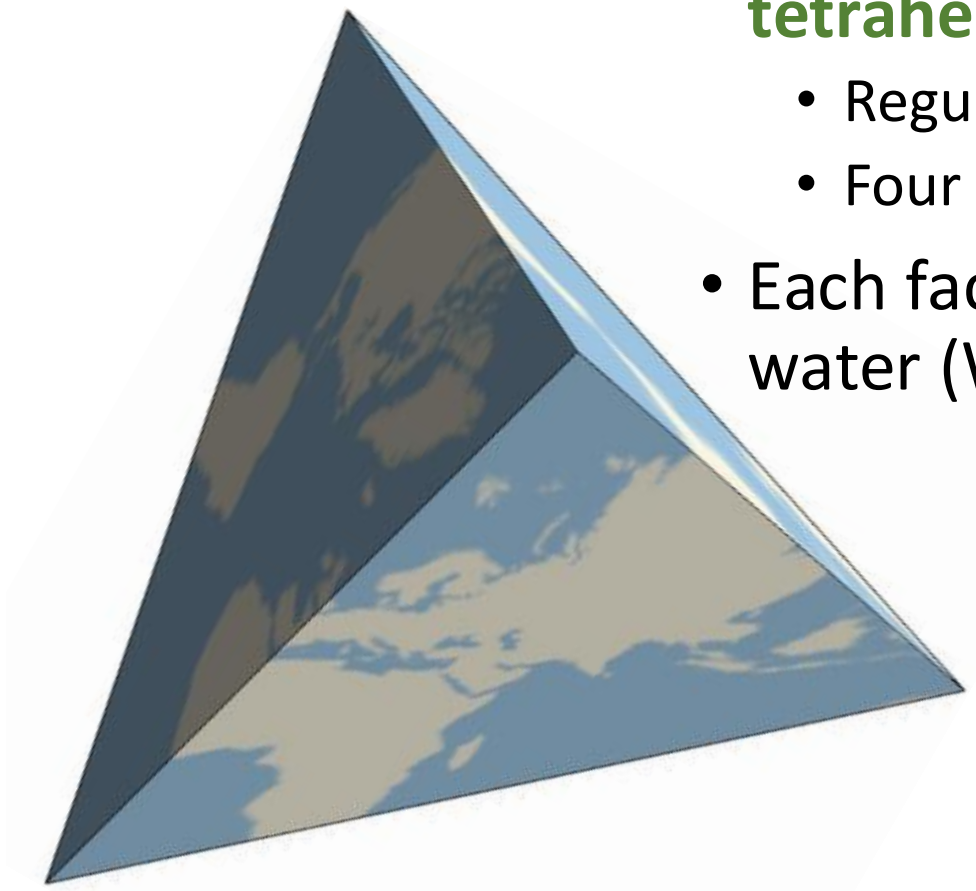
What the graph means

- There is no arrow linking *W* to *L* or *L* to *W*
  - They are *independent measurements* that do not influence each other
  - We have the same probability of observing a W or L, regardless of the last observation
- The graph does not tell us *how p* and *N* influence *W* and *L*, only that they *can*.
  - The researcher defines the relationships



Tetrahedron Earth

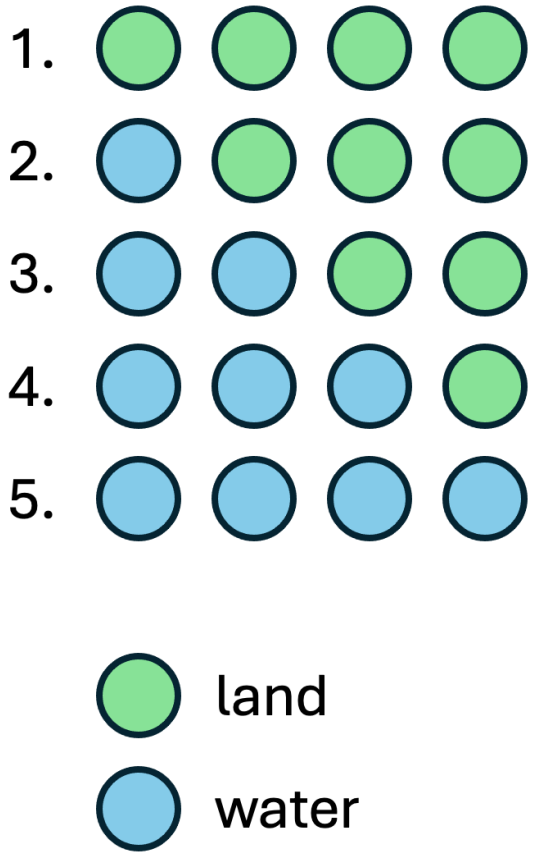




- Pretend the Earth is a **tetrahedron**.
  - Regular Platonic solid
  - Four faces
- Each face can either be water (W) or land (L)

Plausible explanations

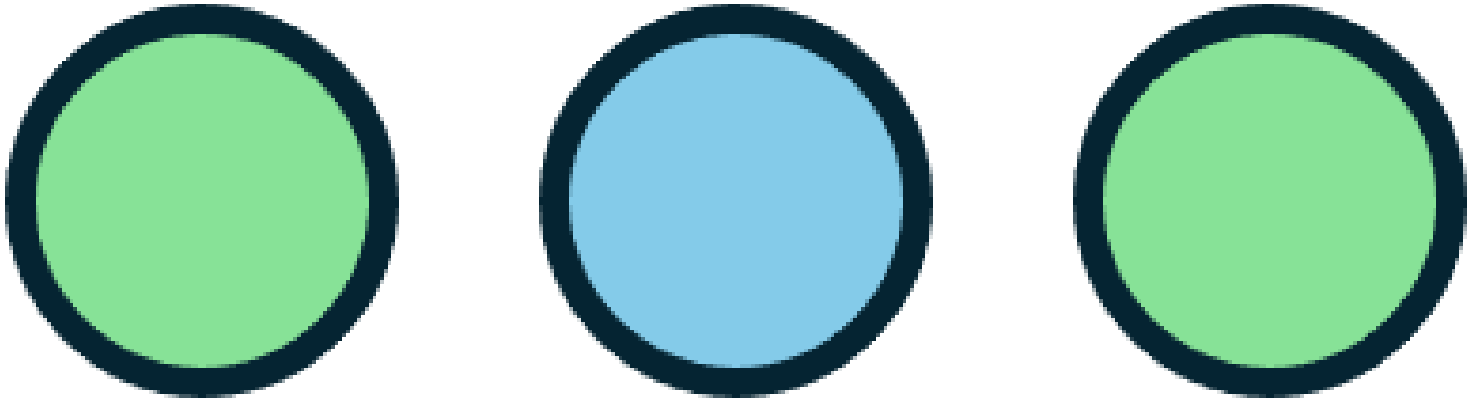
- For each possible explanation of the sample, count all the ways the sample could happen.
- Explanations with more ways to produce the sample are more plausible



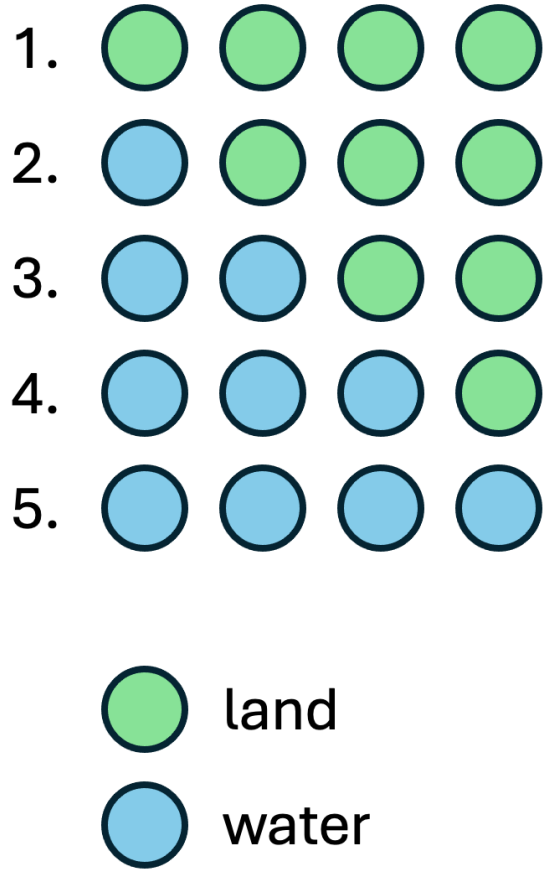
- For each **possible proportion of water on the globe**, count all the ways **the observed sample of tosses could happen.**
- **Proportions** with more ways to produce the observed sample are **more plausible**.

- (This is Bayesian statistics. Really)

A small dataset



Observations
L W L



How many ways?

- How many ways can the 25% water version of Tetrahedron Earth explain the observations?

- Observation 1: L
  - 3 “land” faces, so 3 ways to get this
- Observation 2: W
  - 1 “water” face, so 1 way to get this
- Observation 3: L
  - 3 “land” faces, so 3 ways to get this

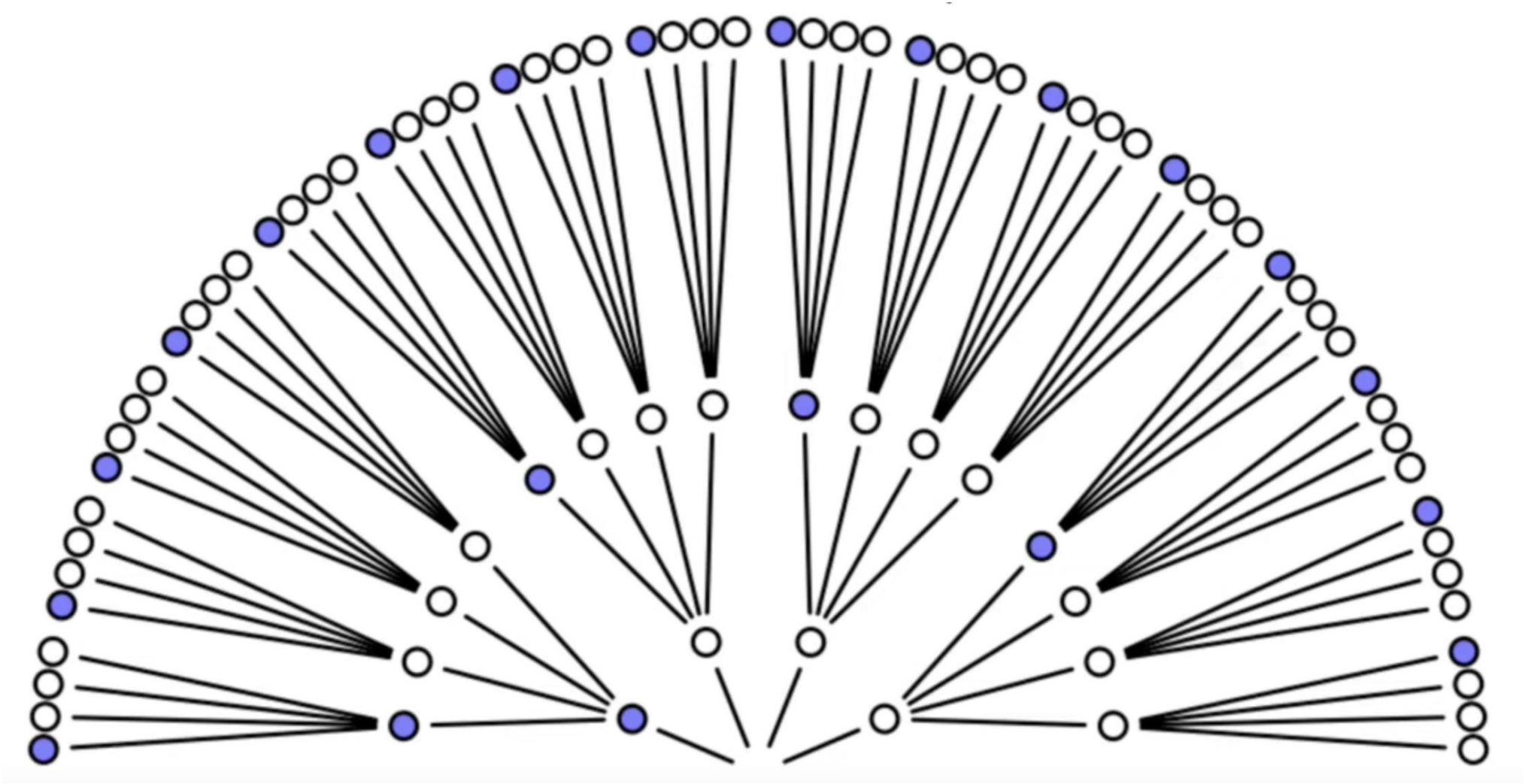
- **Ways to explain the observations:**
  - **3 x 1 x 3 = 9**

2.

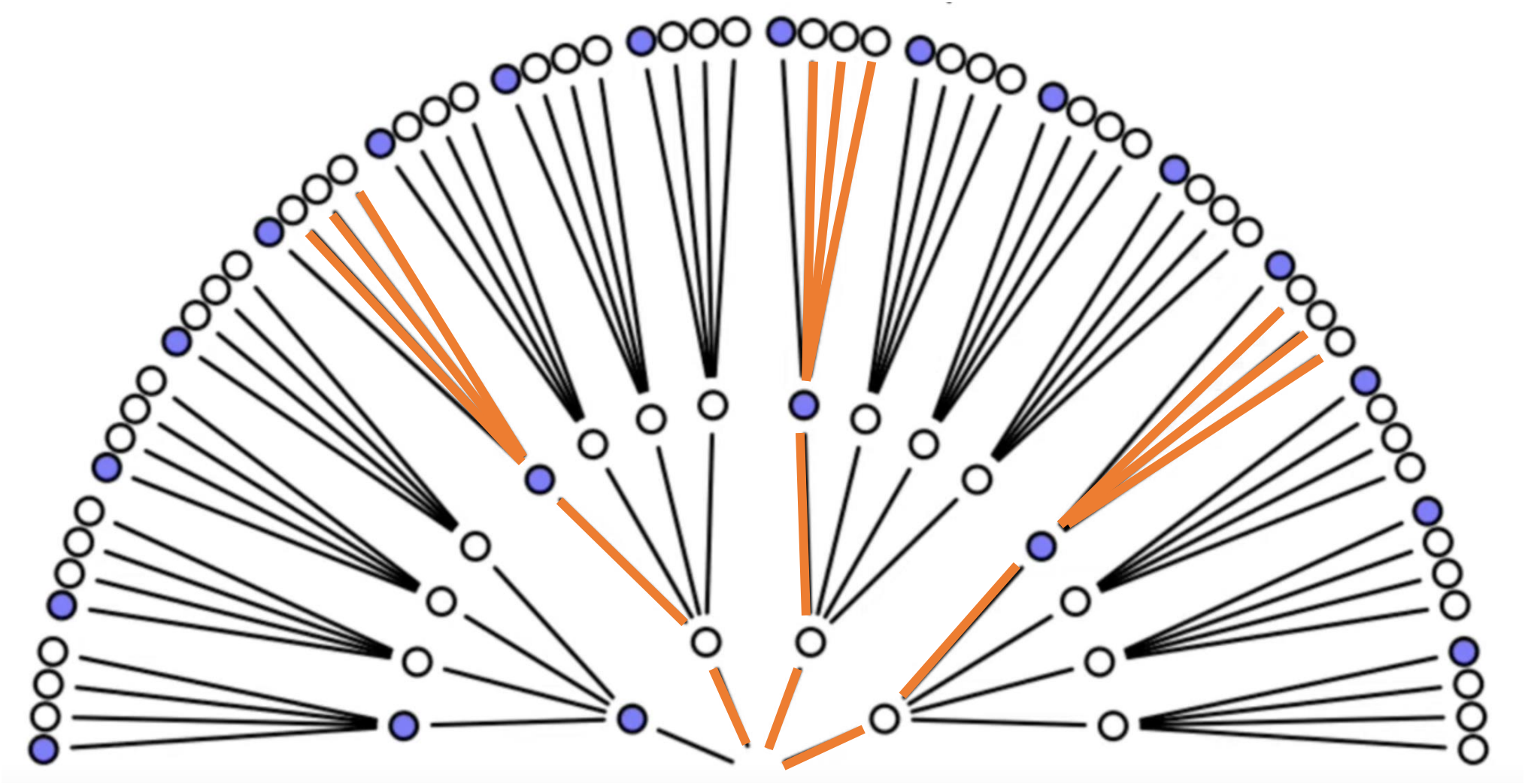
25% water Tetrahedron Earth

observations

Why nine ways to make LWL?



Why nine ways to make LWL?

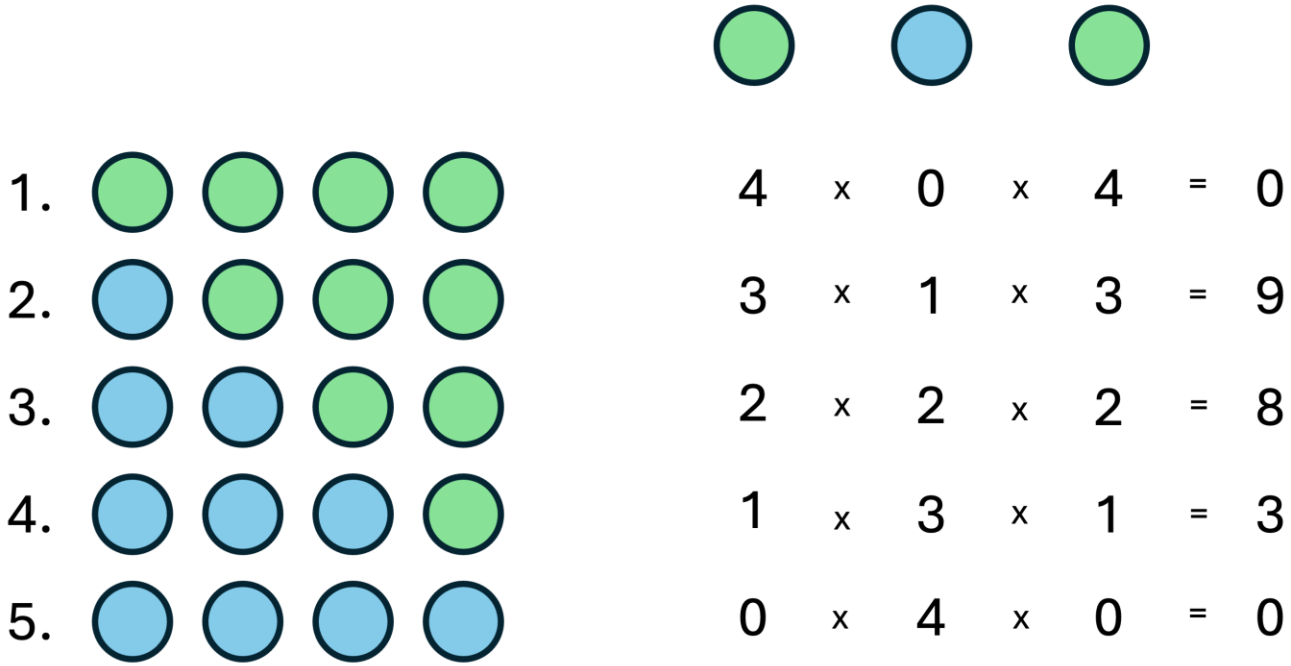


The most plausible Earth

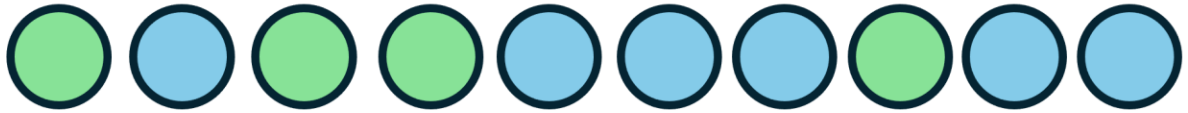
- We can **repeat this for all possible configurations** of Tetrahedron Earth, to calculate the most plausible configuration
  - This is the **most plausible value for *p***, the proportion of water on Tetrahedron Earth's surface.

- **25% water is the most plausible** (9 ways)
- But **50% water is almost as plausible** (8 ways)

- We have a *distribution* of plausibility



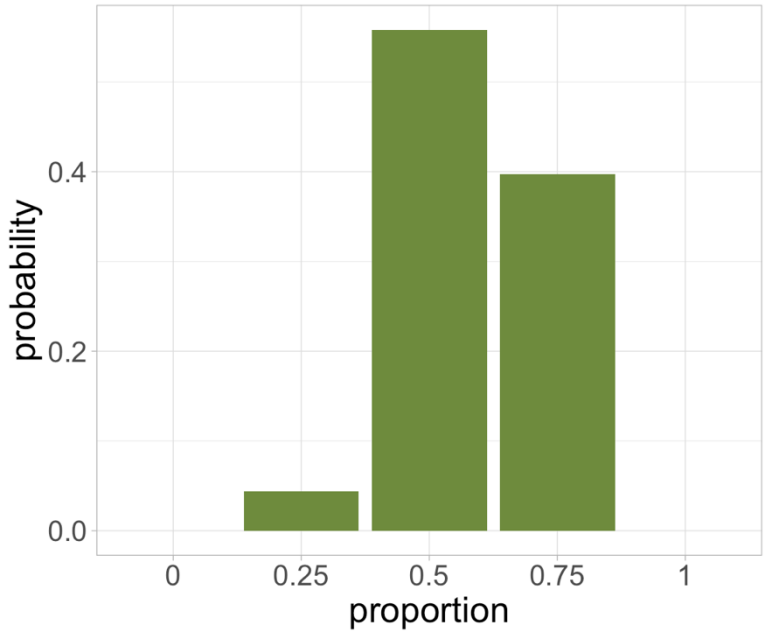
Make more observations



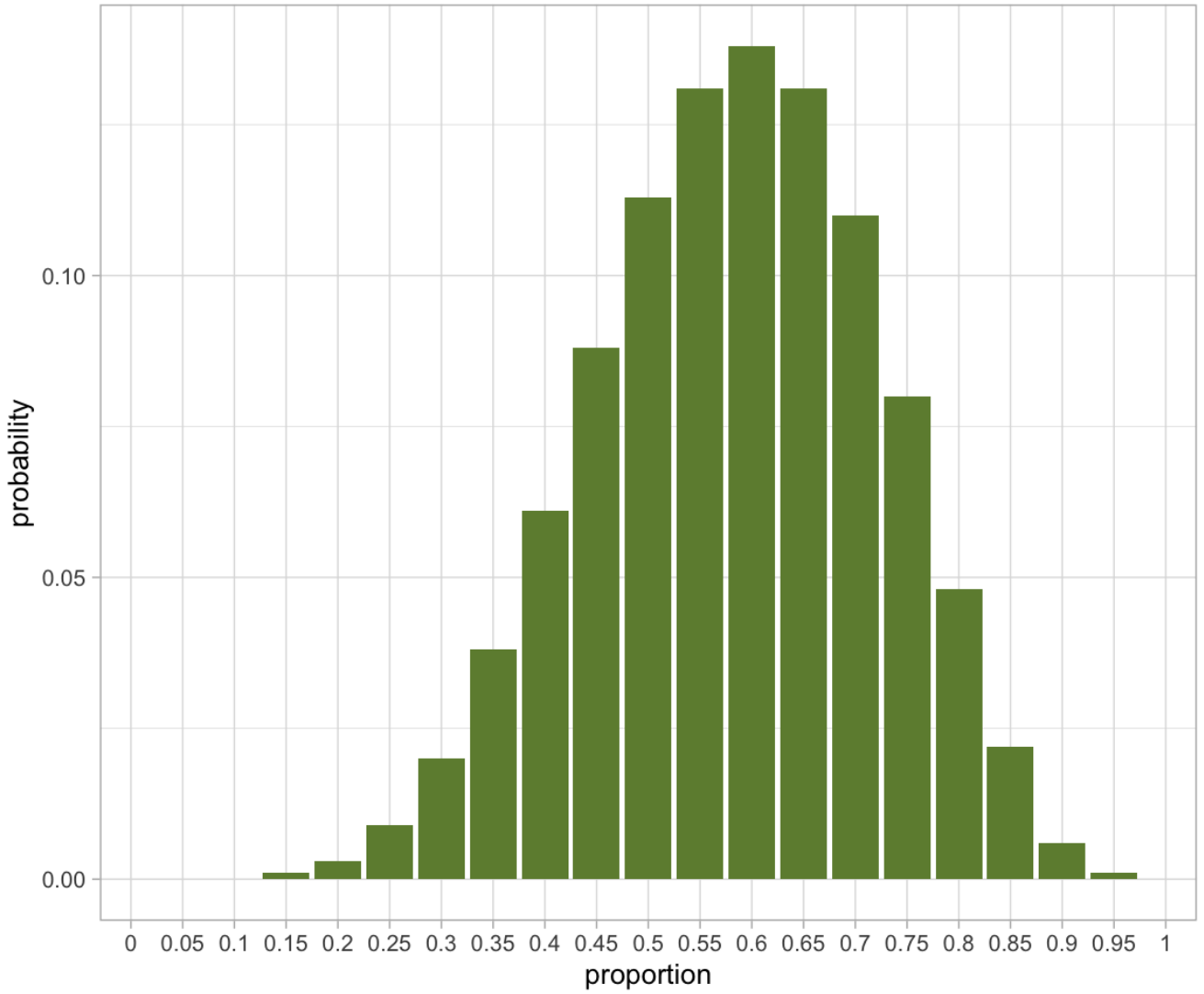
1.  $4 \times 0 \times 4 \times 4 \times 0 \times 0 \times 0 \times 4 \times 0 \times 0 = 4^4 \times 0^6 = 0$
2.  $3 \times 1 \times 3 \times 3 \times 1 \times 1 \times 1 \times 3 \times 1 \times 1 = 3^4 \times 1^6 = 81$
3.  $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 2^4 \times 2^6 = 1024$
4.  $1 \times 3 \times 1 \times 1 \times 3 \times 3 \times 3 \times 1 \times 3 \times 3 = 1^4 \times 3^6 = 729$
5.  $0 \times 4 \times 0 \times 0 \times 4 \times 4 \times 4 \times 0 \times 4 \times 4 = 0^4 \times 4^6 = 0$

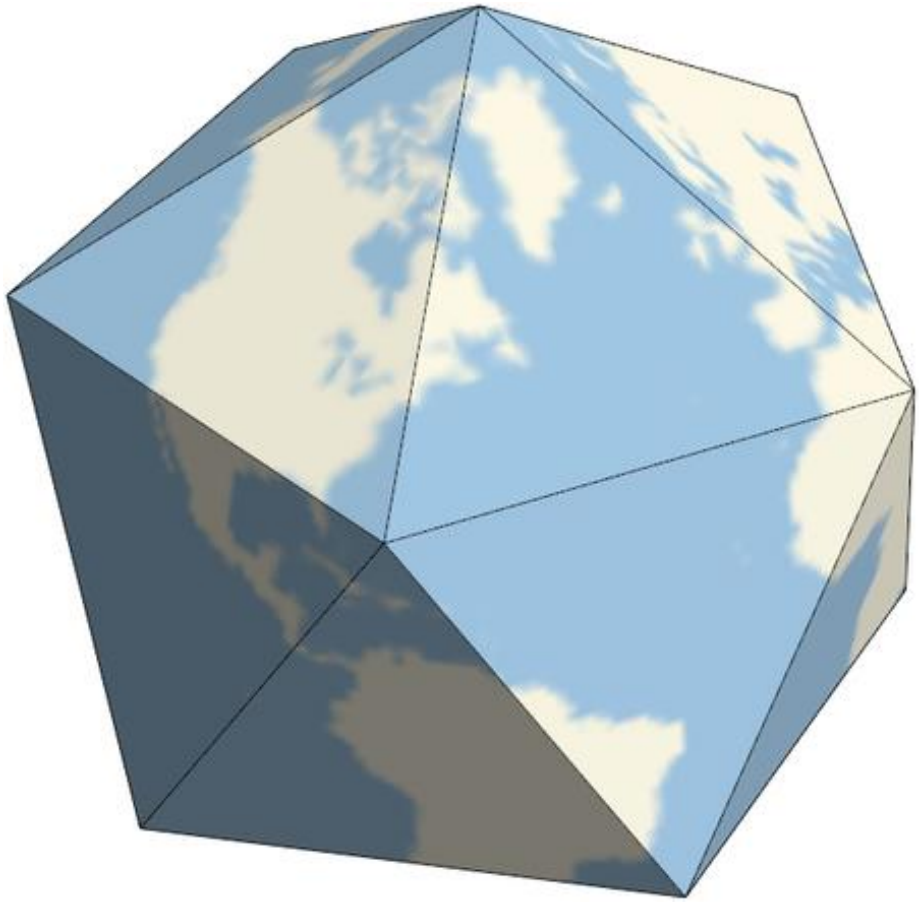
The most probable planet?

		Probability
1. 	$4^4 \times 0^6 = 0$	0
2. 	$3^4 \times 1^6 = 81$	0.044
3. 	$2^4 \times 2^6 = 1024$	0.558
4. 	$1^4 \times 3^6 = 729$	0.397
5. 	$0^4 \times 4^6 = 0$	0
	Sum: 1834	

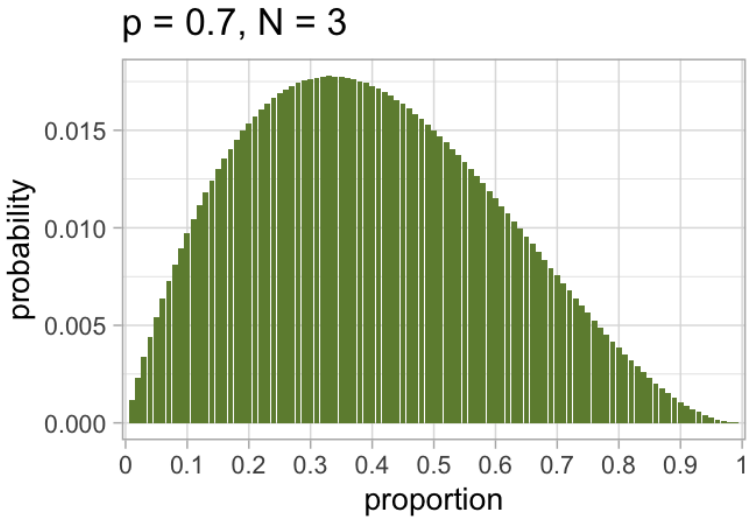


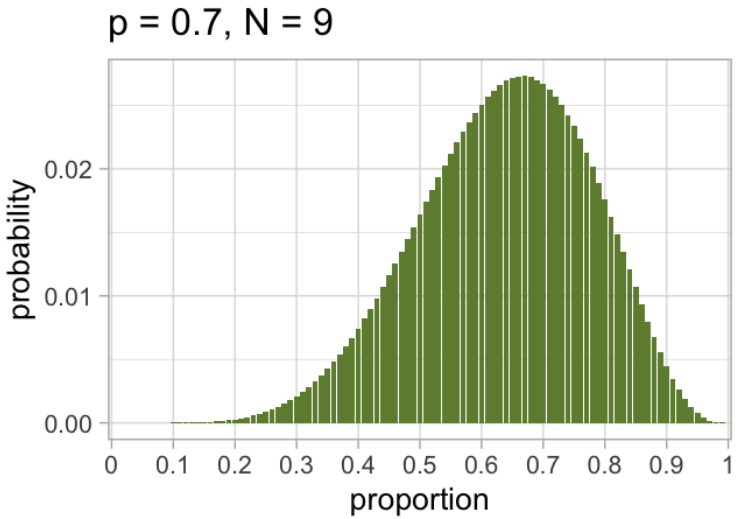
Icosahedron Earth

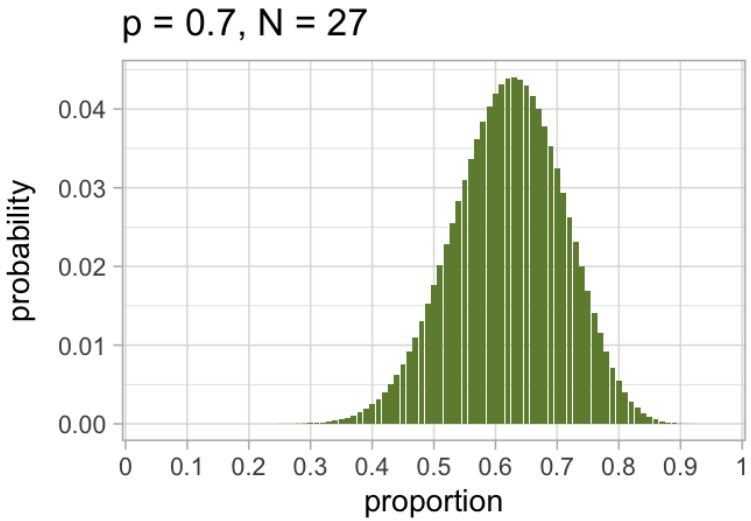


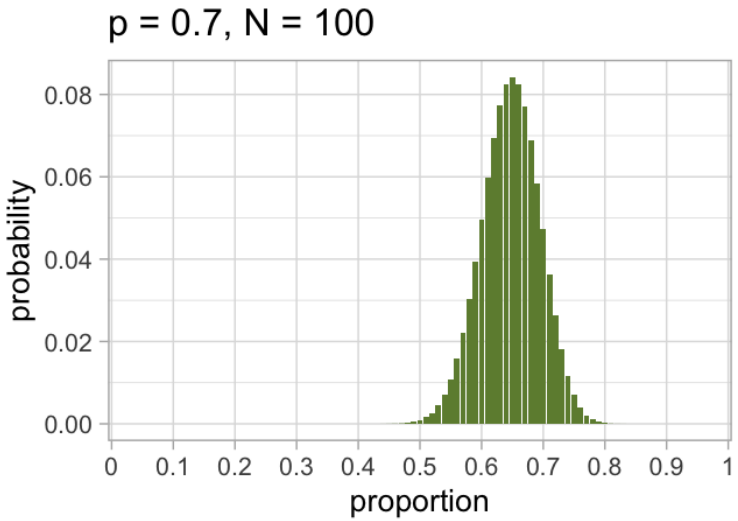


Make more observations (100 faces)







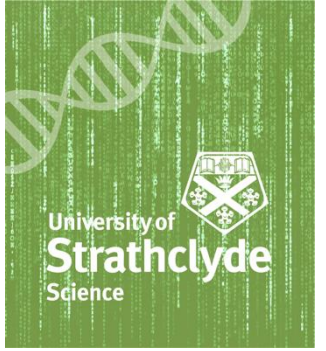


Sample size

- **Sample size** is a central concern of experimental design

- Too few observations = **uncertainty** in results, or **incorrect** results
- Too many observations = **wasted time and resources**
- In animal experiments, **we also want to minimize animal suffering** – as few experimental subjects as necessary **(3Rs)**

- We want to collect an optimal number of observations to give **robust results at an acceptable (financial and ethical) cost**



Statistical Power

Statistical power tells us what conclusions can reasonably be drawn from statistical analysis of our experiment

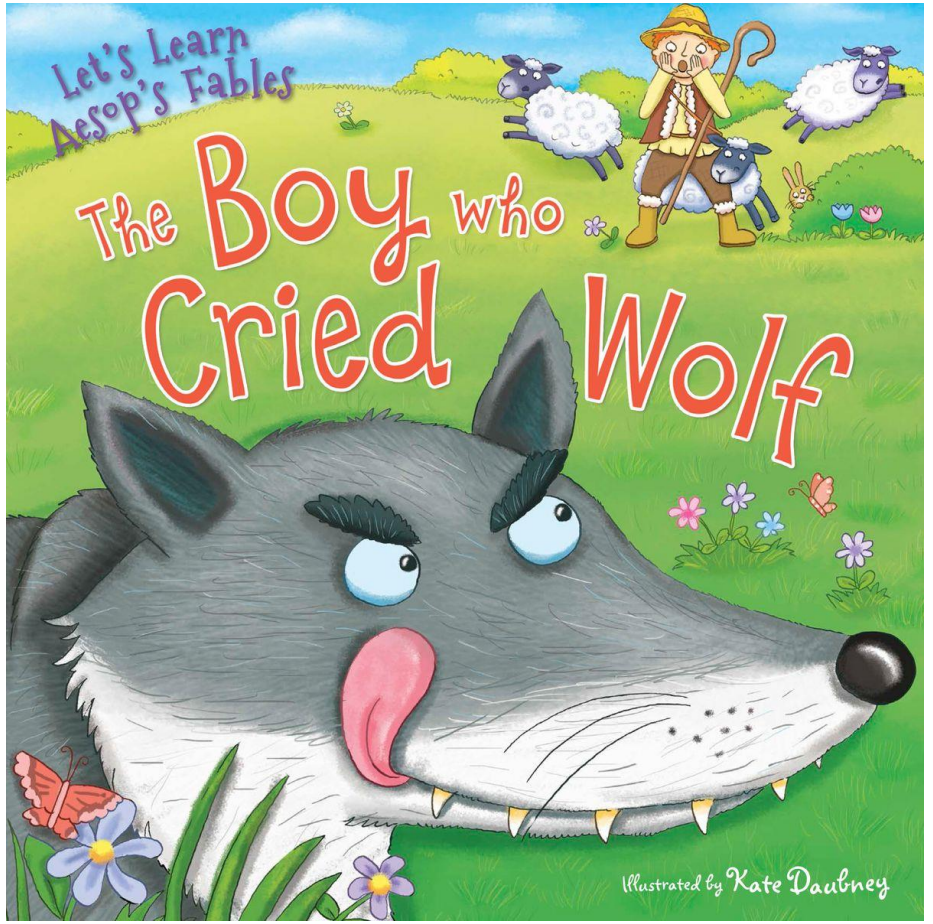
Statistical power

- **Statistical power** is the probability, *before a study is performed*, that a particular comparison will achieve "statistical significance" at some predetermined level.
- More precisely, **statistical power is the probability that the study will not return a false negative result** (known in statistics jargon as a "Type II error")

- As researchers we have two choices to make:
  - **What threshold of statistical significance** ("type I errors", *p*-value, or $\boldsymbol{\alpha}$) **is appropriate**?
  - **What rate of false negative results** ("type II errors" or 1-$\boldsymbol{\beta}$) **is acceptable** to us at that significance level? (i.e. what *statistical power* do we want?)

Type I and Type II errors

	WOLF 	NO WOLF
CRIES 	Boy cries wolf There's a wolf Village is happy	Boy cries wolf There's no wolf Village is unhappy FALSE POSITIVE TYPE I ERROR
DOES NOT CRY	Boy doesn't cry wolf There's a wolf Village is unhappy FALSE NEGATIVE TYPE II ERROR	Boy doesn't cry wolf There's no wolf Village is happy



Statistical power and sample size

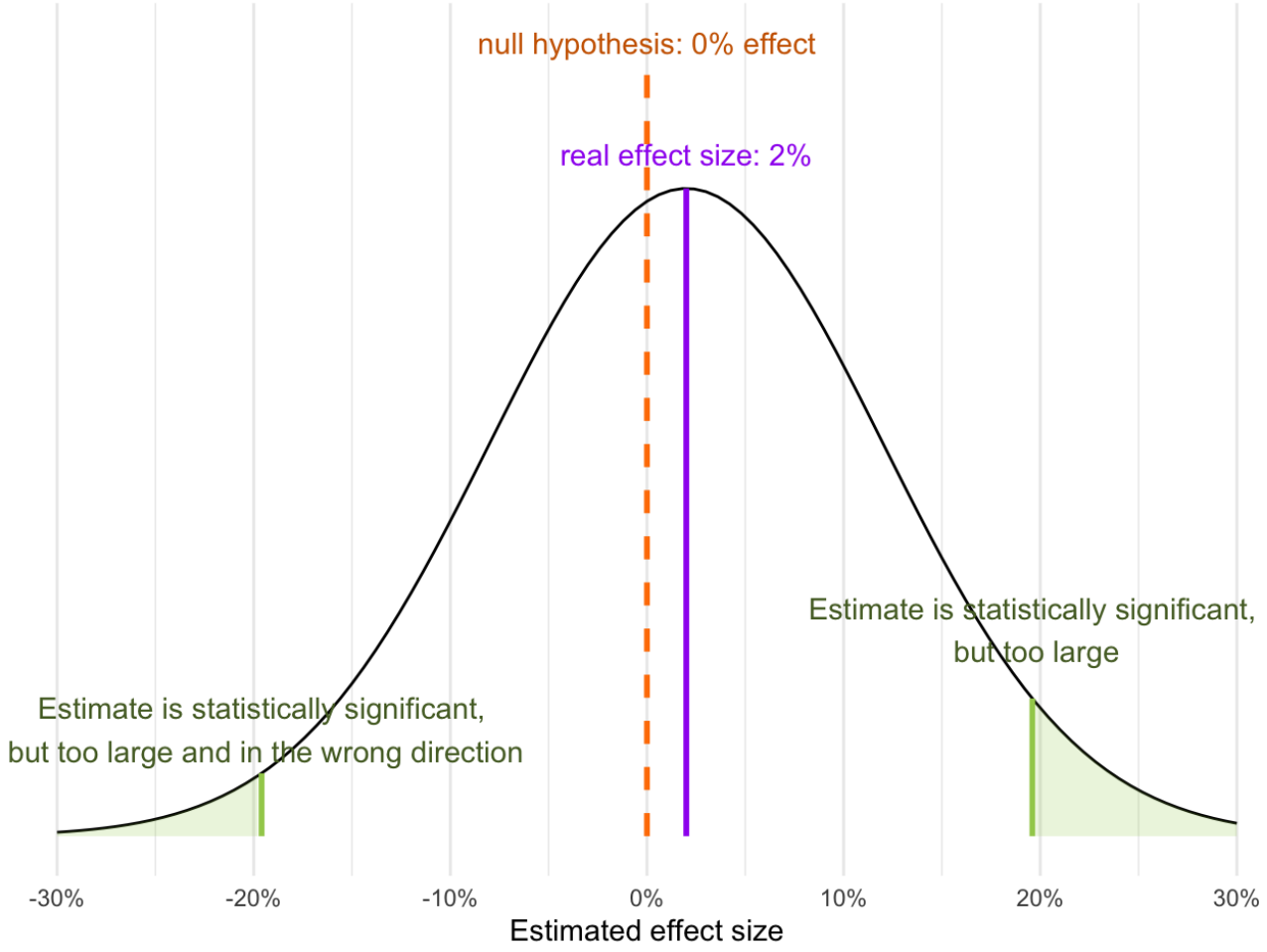
- **Sample size** is the number of experimental units (often, but not always, individuals in this context) used per group in the experiment.
- The sample size for an experiment involving animals should be determined by a *statistical power calculation* and chosen such that the experiment is neither **underpowered** nor **overpowered**.
  - **Underpowered**: too few subjects to interpret the experiment properly
  - **Overpowered**: so many subjects that we are wasteful of time and/or resources, or cause excess suffering

- **There is no universally applicable sample size.**
  - An appropriate sample size should be calculated specifically for each experiment.

Underpowered studies

- Studies are underpowered if the **expected effect size is small in relation to the variation** (e.g. standard error) of the measurement.
  - Too few subjects
  - Noisy nuisance variables
  - Too small an effect size

- It is **NOT** true that if the effects are large enough to be seen in a small study they must be real large effects



Ethical costs

- As stated in the NC3Rs EDA guide,
  - **“Under-powered *in vivo* experiments waste time and resources, lead to unnecessary animal suffering, and result in erroneous biological conclusions.”**
- Similarly, **over-powered** experiments in which more animals than necessary are used to establish a result also lead to **unnecessary animal suffering** and are **unethical**.
  - “Ethically, when working with animals we need to conduct a harm–benefit analysis to ensure the animal use is justified for the scientific gain. Experiments should be robust, not use more or fewer animals than necessary, and truly add to the knowledge base of science.” (Karp and Fry (2021)).

Increasing statistical power

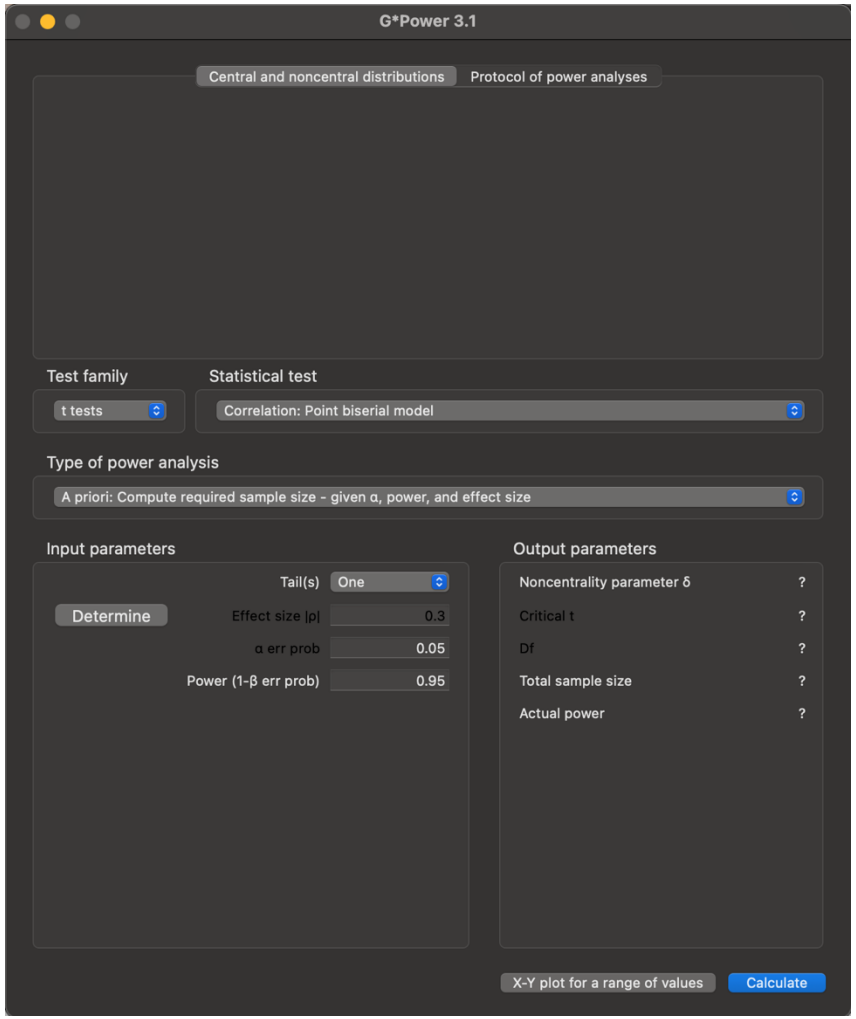
- **Reduce the variability in the experiment**
  - Control for nuisance effects
  - Make precise measurements/estimates
- **Increase the number of experimental units**
  - Using more experimental units reduces the estimated error of the mean
  - 2x sample size reduces standard error 1.4x
- **Increase the effect size**
  - More effective than increasing sample size (2x effect size better than 1.4x reduction in standard error)

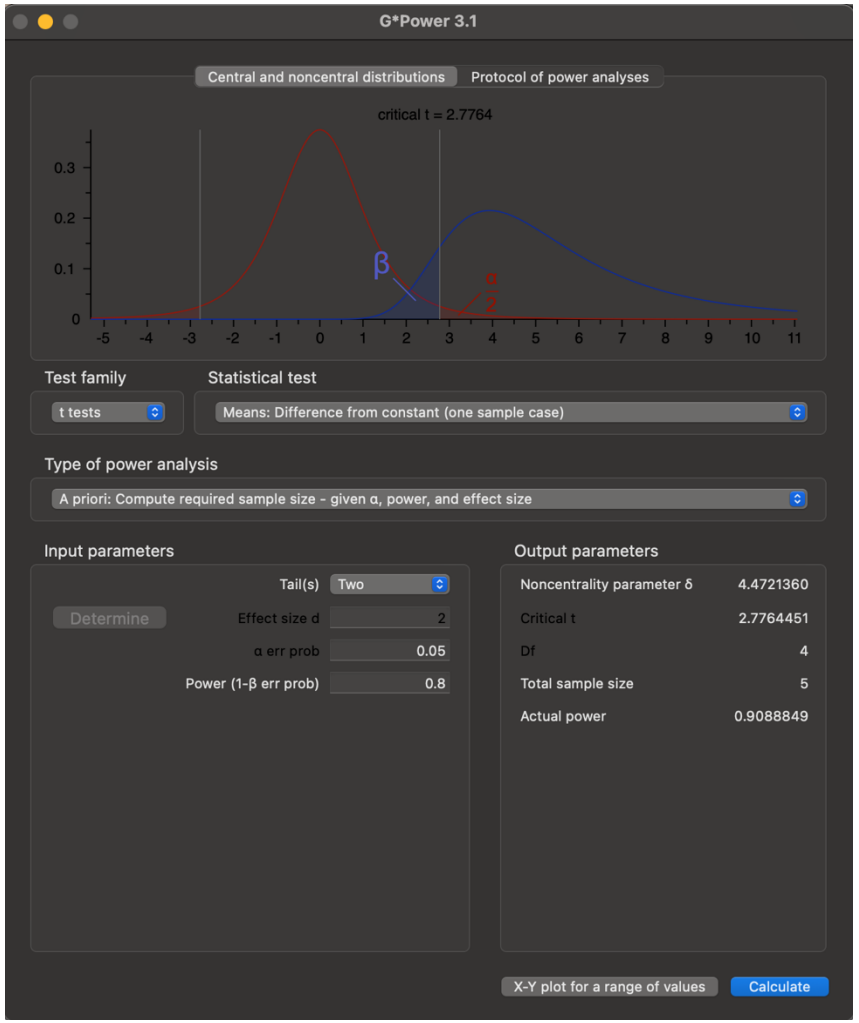
Live Demo: G*Power

- **Is the average weight of a group of mice statistically different from 25g?**

- It's a simple experiment, but **we still need to answer some questions**:
  - What statistical test do we want to perform?
  - What do we want to know from the analysis?
  - What kind of power analysis do we want to carry out?
  - What statistical power do we want?
  - What is the expected effect size, and expected variation in the sample?

Live Demo: G*Power





Let’s have a wee break

We’ve got a long walkthrough coming up…

NC3Rs EDA

Online experimental design, critique, and analysis tool walkthrough

Our Experiment

- **Test for the effect of a novel drug on glucose levels in the plasma of diabetic mice.**

- "We will use NONcNZO10/LtJ mice (JAX) to represent a polygenic form of diabetes. Mice in the **treatment** group will receive a single subcutaneous injection of **drug A** (up to 30mg/kg). Mice in the **control** group will receive a single subcutaneous injection of **vehicle**. Mice will be randomly assigned to receive either drug A or vehicle only without regard to the sex of the animal. 48 hours after administration of drug or vehicle, the blood glucose level will be measured."

Our beliefs about the experiment

- We are testing the **effect of a new drug** - *drug A* - **on plasma glucose levels**
  - There is a single experimental *variable* of interest: whether drug A is present (treatment) or absent (control)
  - The plasma glucose level is our *outcome measure*
- Our experimental subjects are diabetic (NONcNZO10/LtJ) mice
- We will divide individuals into **two groups**, by different *pharmacological treatment*
  - Group 1 will receive the vehicle with no active drug (the *control*)
  - Group 2 will receive the vehicle, containing active drug (the *treatment*)
- Individuals will be **allocated** to each group **randomly**, by *complete randomisation*
- We are **not testing for the effect of sex on drug performance** as an experimental *variable of interest*

Causal Relationships

The outcome variable, plasma glucose concentration, is dependent on

1. the pharmacological effect of drug A
2. the pharmacological effect of the vehicle
3. individual differences between experimental subjects (mice)

Effect of drug A

- “treatment” vs “no treatment”

- In statistical terms, the presence or absence of drug A is our **independent variable** or **explanatory variable**.
- Whether drug A is administered or not is entirely under our control, as experimenters.

“control” vs “treatment”

- Drug A, like most drugs, is carried in a *vehicle* - a substance expected to be inert in the context of the treatment.

- In our experiment we use two *pharmacological interventions*:

1. injection of a vehicle (*control*)
2. injection of a vehicle containing drug A (*treatment*)

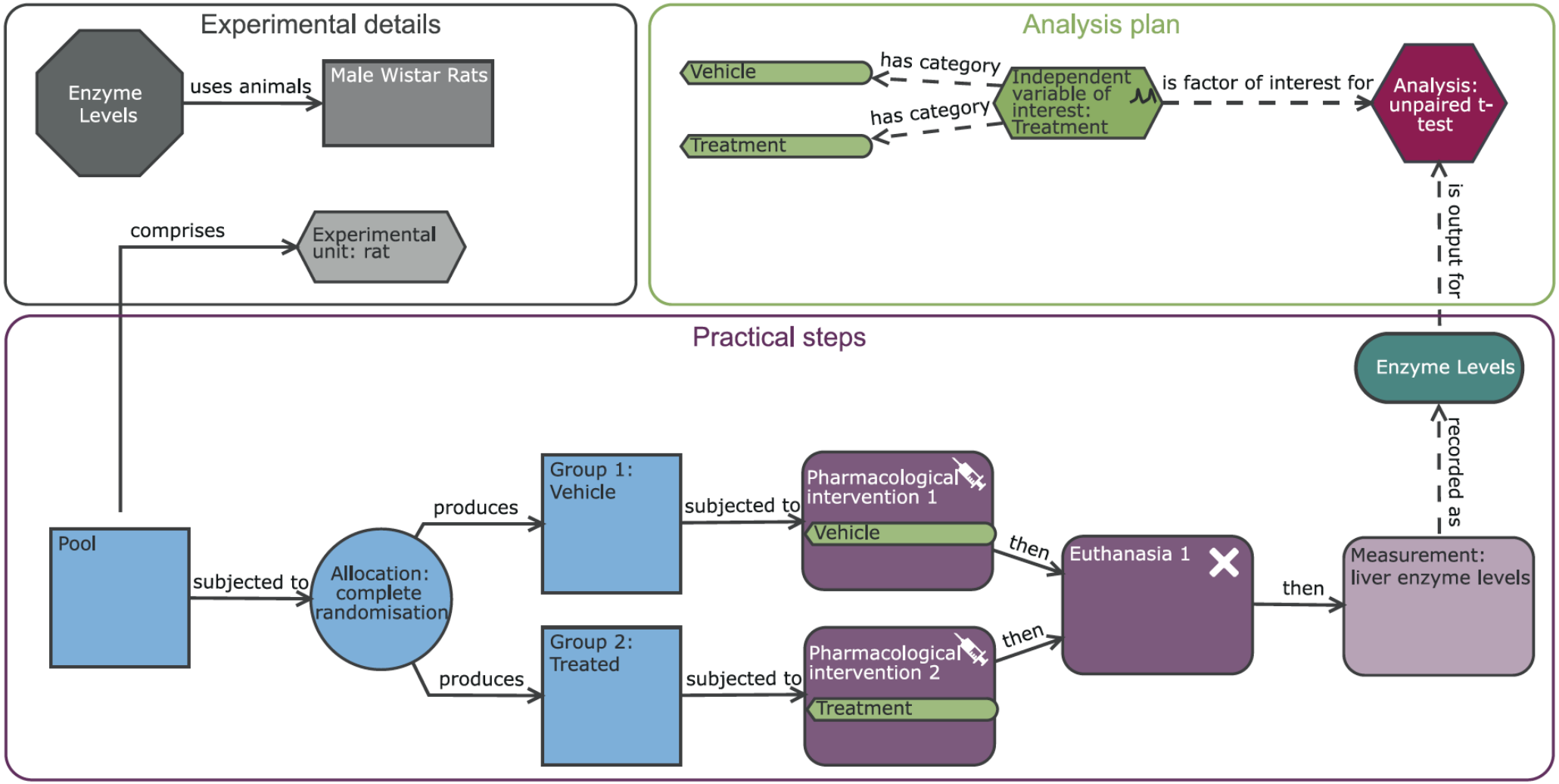
Assignment of individuals to groups

- underlying differences between individuals - is accounted for by randomising subjects into experimental groups:

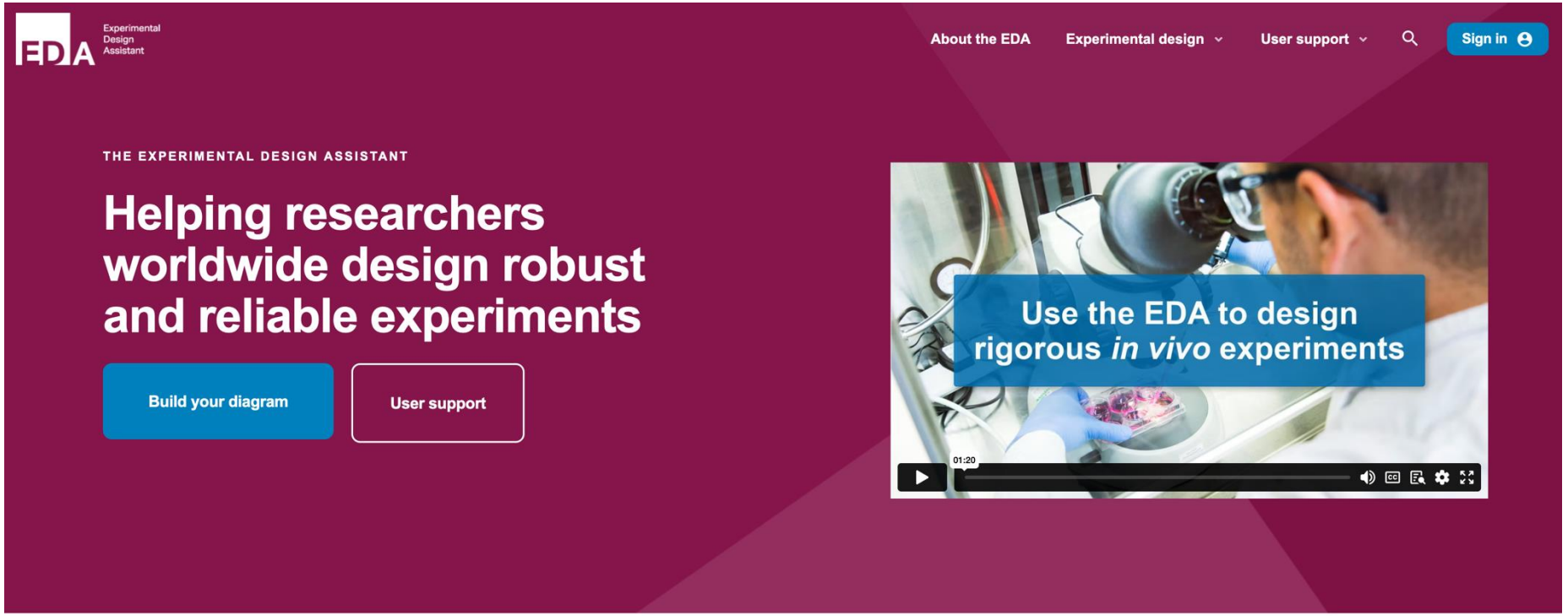
- Treatment (vehicle + drug)
- Control (vehicle)

- we *assume* that all individuals are "drawn from the same population,"
  - each mouse is a random choice from a single pool whose underlying plasma glucose level is distributed according to some kind of regular (potentially parameterisable) statistical distribution

EDA Diagrams



Live Demo: NC3Rs EDA



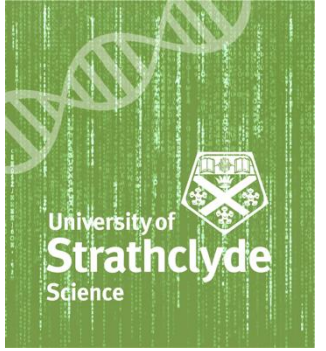
A free resource from the **NC3Rs** used by over 19,000 researchers worldwide to help design robust experiments more likely to yield reliable and reproducible results.

The EDA helps you build a diagram representing your experimental plan, which can be critiqued by the system to provide **bespoke feedback**.

The EDA also:

- Recommends statistical analysis methods
- Provides support for randomisation and blinding
- Performs sample size calculations

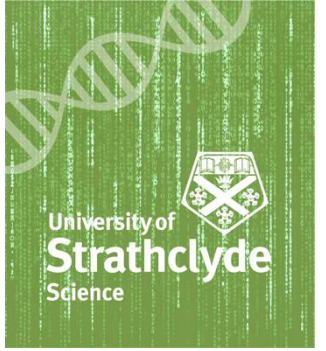
The EDA website provides information about **experimental design concepts**, and how to apply these in your experiments.



Summary

What I hope I've been able to explain

Cause and Effect



- **the data you collect does not contain information about causes and effects**

- **your beliefs and assumptions about an experiment are the causal model for that experiment**

Sample size and measurement

- **the values we collect in an experiment lead to *estimates* of "true" values**
- **causal graphs explain our beliefs about what variables affect our experimental measurements**
- **the number of experimental units affects the precision and confidence we can have in our estimated value**

Power calculations

- **statistical power tells us what conclusions can reasonably be drawn from statistical analysis**
- ***underpowered experiments* lead to uninformative results and misleading conclusions**
- **it is usually better to design your experiment to increase effect size/minimise noisy variation than to increase sample size**
- **how to estimate sample size for a desired statistical power**

NC3Rs EDA

- **How to design, critique, and analyse an experiment using the NC3Rs EDA tool, and share the resulting design with others.**

Final notes

- Please see the **support materials** online at
  - https://sipbs-compbiol.github.io/MP968-Workshop_Experimental_Design

- Links to these slides, and the support material, can be found on the **MP968 MyPlace page**
  - https://classes.myplace.strath.ac.uk/course/view.php?id=15691

- **Please fill out the short feedback questions on MyPlace**
  - https://classes.myplace.strath.ac.uk/mod/quiz/view.php?id=2042783