Experimental Plan in Biology Research
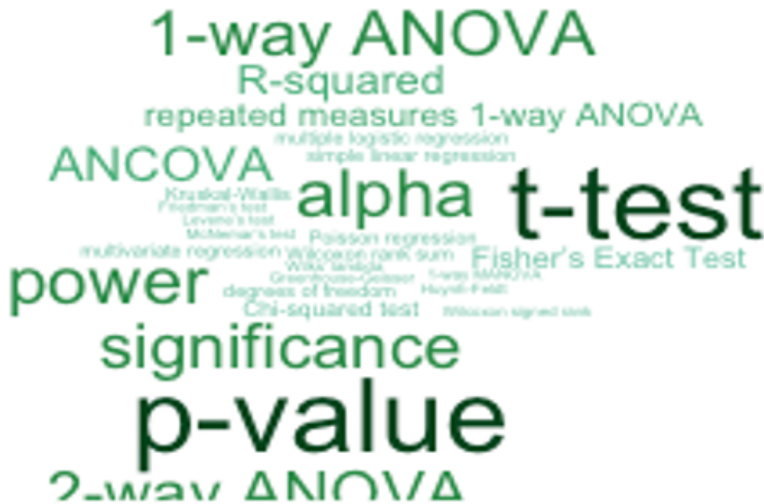## A Statistical Perspective

Terry Neeman

Australian National University

May 2022

Let's talk about statistics and its role in the sciences

What a first course in statistics emphasises:
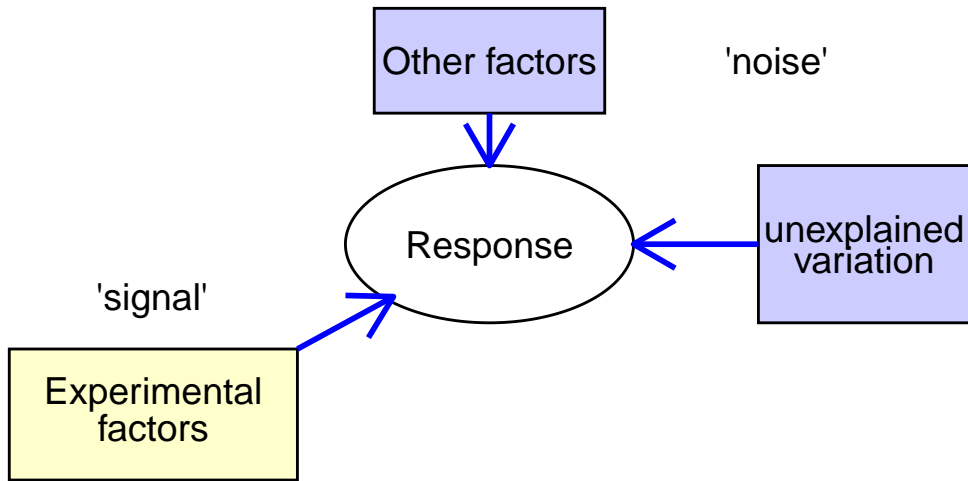
What statistics is Really About:

"Model" of an Experiment

- What is your research question?
- What is the main outcome of interest?
- What are the experimental factors?
- What other factors may influence outcome?

Statistical model: tool for turning data into information

Model of an Experiment

Part I: What is Statistical Framework for your Experiment?

Identify outcome measures and experimental factors

- What is role of Guanylate-binding protein (GBP) in autophagy and apoptosis in (mouse) intestinal cells?
- What is role of GBP on the production of interferon and other inflammatory markers in (mouse) intestinal cells?
- What is effect of inbreeding on associative learning in *Gambusia holbrooki*?
- What is effect of mutated UBF (UBFE210K) on RNA poly I transcription in mouse model?

► Do rainbow fish *M. Malanda* and *M. splendida* have different temperature preferences?

► Do rainbow fish *M. Malanda* and *M. splendida* show a preference for their own species?

► Do infectious agents cause growth anomalies (GA) in coral species?

The Experimental Plan

Points to Consider
- Type of Response
  - measuring instrument
  - how reliable, repeatable
- Experimental Factors
  - single factor/ multi-factorial
- Study conduct
  - ad hoc effect masking effect of experimental factor
  - multiple repeats of experiment
    - different litters
    - different reagents
    - gating strategy
    - environmental conditions

Study conduct introduces "noise" that can confound experiment
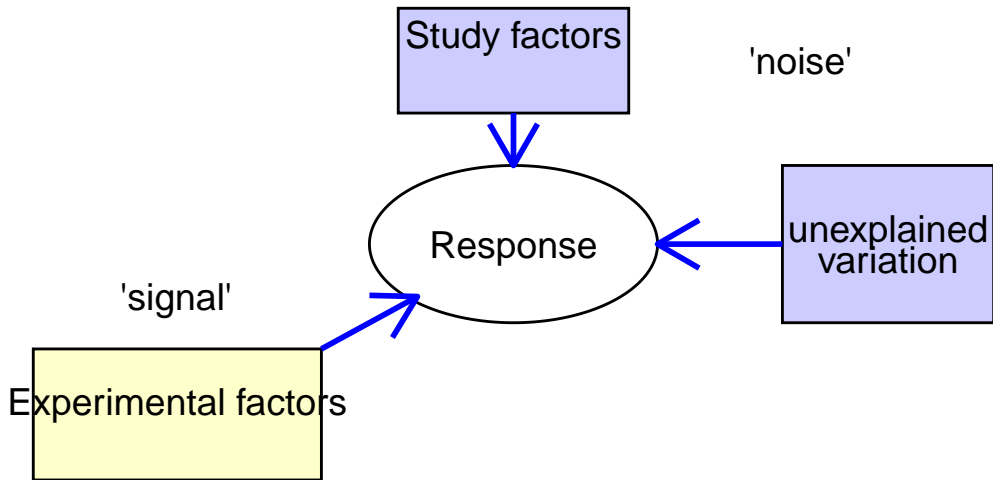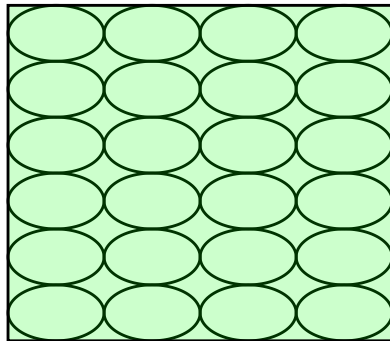
Example: What are problems with this experimental plan?

Top Shelf
WT
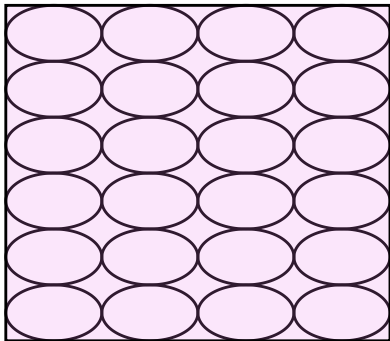
Bottom Shelf
mutant

Group Exercise:

For an experiment you plan to do next semester:
- ▶ Aspect of study conduct that can contribute to "noise".
- ▶ Identify a technology in your experiment and its reliability.
- ▶ How many times will you repeat experiment? What will change between repeats?

From Experimental Plan to Statistical Model

Example 1

What is role of GBP8 in apoptosis in (mouse) intestinal cells?

- ▶ Outcome measure: apoptosis activity
- ▶ Experimental factors: GBP8 expressed (Yes/No)
- ▶ Other things to consider: litter, multiple measures/mouse, experiment
- ▶ Question: does mean activity differ between WT and GBP8-/- intestinal cells?

$$model1 = lm(activity \sim GBP8 + (1|litter) + (1|mouseID))$$

Example 2

What is effect of inbreeding on associative learning in *Gambusia holbrooki*?

- ▶ Outcome measure: result on cognitive task (?)
- ▶ Experimental factors: inbred (Yes/No)
- ▶ Other things to consider: ancestry
- ▶ Possible question: does time to complete task differ between inbred and outbred mosquitofish?

$$model2 = coxph(Surv(time, complete) \sim inbred + (1|family) + (1|tank))$$

Example 3

What is effect of mutated UBF on RNA poly I transcription in UBFE210K mouse model?

- ▶ Outcome measure: normalised Ct-value for pre-RNA expression?
- ▶ Experimental factors: tissue type (bone marrow, spleen, b-cells)
- ▶ Other things to consider: PCR run, position on plate
- ▶ Possible question: does mean Ct-value differ between tissue types?

$model3 = lm(Ct \sim tissue_type + (1|plate) + (1|row) + (1|column) + (1|sample))$

Example 4

Do rainbow fish *M. Malanda* and *M. splendida* show a preference for their own species?

- ▶ Outcome measure: prefer *M. Malanda* (Yes/No)
- ▶ Experimental factors: species (2)
- ▶ Other things to consider: repeated measures of subject, potential mates in tank.
- ▶ Possible question: Does preference for *M. Malanda* depend on species?

$model4 = glmmTMB(M.Malanda \sim species + (1|subjectID) + (1|targetfish))$

Example 5

Do infectious agents cause growth anomalies (GA) in coral species?

▶ Outcome measure: develop GA (Yes/No)

▶ Experimental factors: position in tank, coral species

▶ Other things to consider: tank effects, clear definition of GA

▶ Possible question: Are coral touching lesion area more likely to develop GA compared with coral further away?

$$model5 = glmmTMB(GA \sim position * species + (1|tank))$$

Group Exercise:

For an experiment you plan to do next semester:
- ▶ What is the response variable?
- ▶ What are the experimental factors?
- ▶ Write down a model of your experiment.

Model-centred thinking Part 3: R/RStudio, modelling and reproducible research

R: your friendly data analysis tool

- Advantages:
  - Great graphics tools for data visualisation and pattern recognition
  - Easy to model, interpret and summarise data
  - Data organisation
  - Essential for creating transparent, reproducible workflow
- Disadvantages:

R: your friendly data analysis tool

- MORE Advantages:
  - Other researchers can follow logic of analysis
  - Code/Data available to future researchers
  - Easy to get statistical help
  - International community of R users
    - latest analytic techniques
    - online expert advice
- Disadvantages:

R: your friendly data analysis tool

- ▶ MORE Advantages:
  - ▶ Code compiles to well-formatted document
  - ▶ R coding skills valued inside/outside academia
- ▶ Disadvantages:
  - ▶ learning curve
  - ▶ use or lose
  - ▶ too few local R trainers

Statistics and R: How can we help?

- Training workshops in R
- Workshops on Statistical Modelling
- Data analysis workshops
- Weekly biostatistics/bioinformatics drop-in

Contact us any time through e-mail or BDSI admin

Example of analysis using R