Design of Experiments

Peter Ralph

8 October – Advanced Biological Statistics

Experimental design

Goals of an experiment

What do we want to know?

Goals of an experiment

What do we want to know?

How do we measure it?

Goals of an experiment

What do we want to know?

How do we measure it?

This can be done by *observation* or *experiment*.

What's an experiment?

An *experiment* is a study in which the values of important variables (e.g., group membership, dosage) are determined by the *experimenters*.

Otherwise, it is an *observational* study.

What's an experiment?

An *experiment* is a study in which the values of important variables (e.g., group membership, dosage) are determined by the *experimenters*.

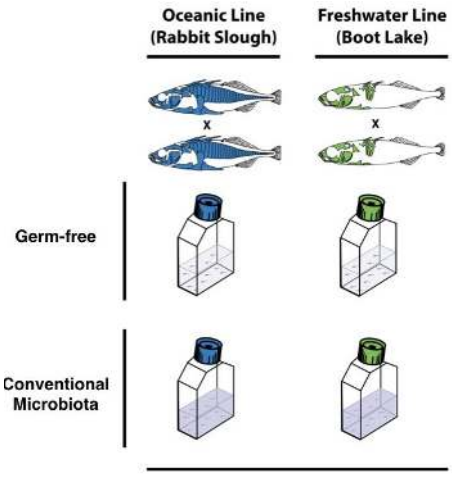
Otherwise, it is an *observational* study.

Note that controlling the set-up doesn't necessarily make it a *good* experiment.

A biological example to get us started

Say you perform an experiment on two different strains of stickleback fish, one from an ocean population (RS) and one from a freshwater lake (BP) by making them microbe free. Microbes in the gut are known to interact with the gut epithelium in ways that lead to a proper maturation of the immune system.

Experimental setup: You decide to carry out an experiment by treating multiple fish from each strain so that some of them have a conventional microbiota, and some of them are inoculated with only one bacterial species. You then measure the levels of gene expression in the stickleback gut using RNA-seq. Because you have a suspicion that the sex of the fish might be important, you track it too.



Term	Definition	Example
Measurement	A single piece of recorded information reflecting a characteristic of interest (e.g. length of a leaf, pH of a water aliquot, mass of an individual, number of individuals per quadrat etc)	Protein content of the milk of a single female koala
Observation	A single measured sampling or experimental unit (such as an individual, a quadrat, a site etc)	A small quantity of milk from a single koala
Population	All the possible observations that could be measured and the unit of which wish to draw conclusions about (note a statistical population need not be a viable biological population)	The milk of all female koalas
Sample	The (representative) subset of the population that are observed	A small quantity of milk collected from 15 captive female koalas[a]
Variable	A set of measurements of the same type that comprise the sample. The characteristic that differs (varies) from observation to observation	The protein content of koala milk.

from Logan, *Biostatistical Design and Analysis Using R*

What makes a good study?

- Will we have the *power* to detect the effect of interest?
  - What are the sources of *noise*?
  - How big do we expect the effect to be?
- How generalizable will the results be?
  - How *representative* is the sample? Of what group?
- What are possible *causal* explanations?
  - What are possible *confounding factors*?

Considerations

1. Where do the samples come from?
2. Sample size, replication, and balance across groups
3. Controls: setting up good comparisons
4. Randomization!

Considerations

1. Where do the samples come from?
2. Sample size, replication, and balance across groups
3. Controls: setting up good comparisons
4. Randomization!

For (2), remember that

$$(\text{margin of error}) \propto \frac{\sigma}{\sqrt{n}}.$$

Toxoplasma gondii infection rates and measures of aggregate neuroticism:

country	prevalence	N18	country	prevalence	N18
Argentina	52.7	51.3	Japan	12.3	50.7
Australia	28.0	48.6	Netherlands	24.5	48.6
Austria	36.0	48.3	Norway	8.6	47.4
Belgium	46.8	49.6	Peru	32.9	48.5
Brazil	66.9	53.7	Poland	46.5	50.7
China	24.3	53.1	South Korea	4.3	48.4
Croatia	37.4	49.3	Slovenia	30.9	50.6
Czech Rep	26.6	51.4	Spain	22.7	49.7
Denmark	22.0	50.3	Sweden	12.5	46.3
Ethiopia	16.4	48.8	Switzerland	36.7	47.5
France	45.0	52.7	Thailand	11.2	48.9
Germany	42.7	48.1	Turkey	46.8	51.4
Hungary	58.9	53.8	UK	6.6	50.1
Indonesia	46.2	50.0	USA	12.3	48.1
Ireland	25.0	50.1	Yugoslavia	66.8	51.1
Italy	32.6	52.6		NA	NA

Lafferty 2006, *Can the common brain parasite, Toxoplasma gondii, influence human culture?*

// reveal.js plugins