SEEC Stats Toolbox

Want to broaden your stats knowledge? Unsure of what you can do with your data? Still developing your proposal?

Join us for the monthly **SEEC Stats Toolbox** seminars where we introduce you to statistical methods that are useful for ecologists, environmental and conservation scientists.

Our next seminar:

Topic:	**Intro to Multivariate Analyses**
Who:	Dr Natasha Karenyi
When:	**Thursday 20 April 2017 (1-2pm)**
Where:	PD Hahn Lecture Theatre 3, PD Hahn Building Level 5, UCT

More details: www.seec.uct.ca.za

SEEC - Statistics in Ecology, Environment and Conservation

University of Cape Town · Universiteit van Kaapstad · iYunivesithi yaseKapa

Keep in touch

@SEEC_UCT

SEEC.UCT

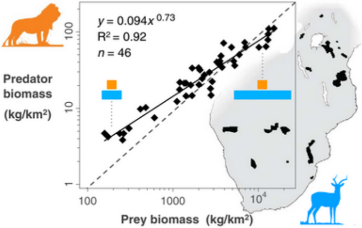
ask to be added to mailing list
(seecuct@gmail.com)

www seec.uct.ac.za

STA2007H -- POSTGRADUATE: STA2007H -- POSTGRADUATE

STUDY DESIGN & DATA ANALYSIS FOR SCIENTISTS

This course provides an introduction to applied statistical modelling. We cover linear regression, design and analysis of experiments, and generalized linear models. There will be a strong emphasis on the practical application of the above me thods, using open-source statistical software R.



Source: IA Hatton, KS McCann, JM Fryxell, TJ Davies, M Smerlak, ARE Sinclair, M Loreau 2015. The predator-prey power law: Biomass scaling across terrestrial and aquatic biomes. Science 349 (6252): aac6284

SEEC Stats Toolbox

Want to broaden your stats knowledge? Unsure of what you can do with your data? Still developing your proposal?

Join us for the monthly **SEEC Stats Toolbox** seminars where we introduce you to statistical methods that are useful for ecologists, environmental and conservation scientists.

Our next seminar:

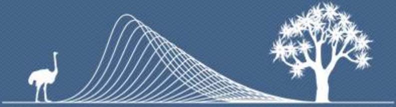
Topic: **Experimental and Survey Design**

Who: Prof Res Altwegg and Dr Birgit Erni

When: **Thursday 30 March 2017 (1-2pm)**

Where: PD Hahn Lecture Theatre 3,
PD Hahn Building Level 5, UCT

More details: www.seec.uct.ca.za



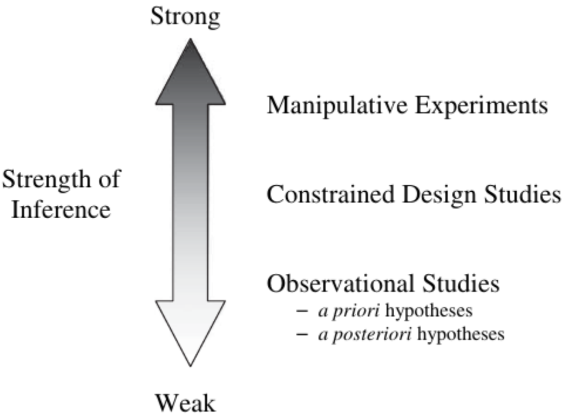
SEEC - Statistics in Ecology, Environment and Conservation

Experimental and Survey Design

- Why?
- What?
- How?

Yoccoz, N. G., J. D. Nichols, and T. Boulinier. 2001. Monitoring of biological diversity in space and time. Trends in Ecology & Evolution 16:446 - 453.

Experimental and Survey Designs



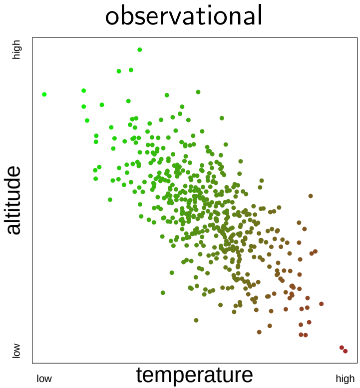
from Mackenzie et al 2006, Academic Press

Experiments: what they are and why we need them

- carefully controlled environment
- everything stays constant
- except the factor of interest (treatment factor)
- measure change in response
- infer causal relationships

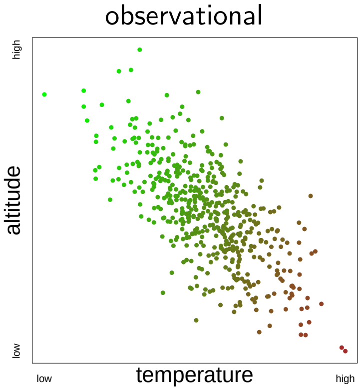
Observational vs experimental studies

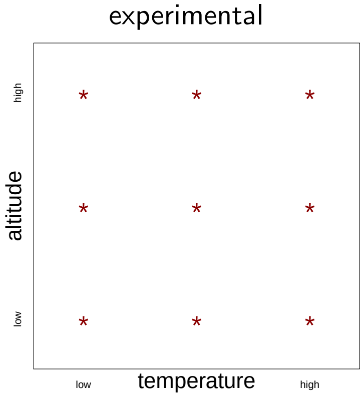
Hypothetical example: Vegetation in relation to temperature and altitude



Observational vs experimental studies

Hypothetical example: Vegetation in relation to temperature and altitude





Summary

Observational	Experimental
↓	↓
random sample	randomize
↓	↓
representative	equalize groups (average out all other factors)
↓	↓
make statements about the population	causal inference (cannot say anything about population in general)

Three principles in experimental design

1. Randomization
2. Replication
3. Reduce (unexplained variation)

Three principles in experimental design

1. Randomization
2. Replication
3. Reduce (unexplained variation)

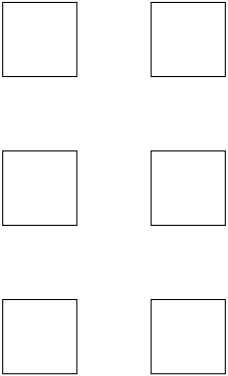
What is the experimental unit?

ED Principle 1: Randomization

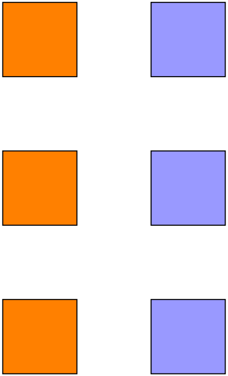
Randomization

is the assignment of treatments to experimental units at random.

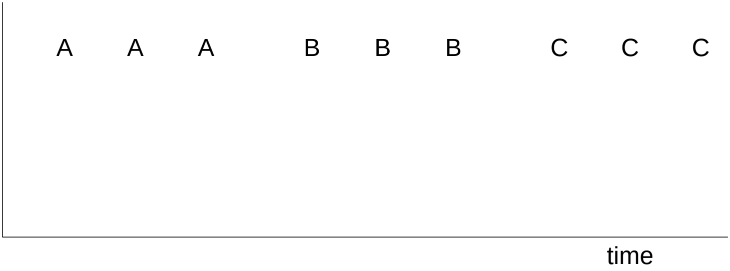
What happens if we don't randomize?



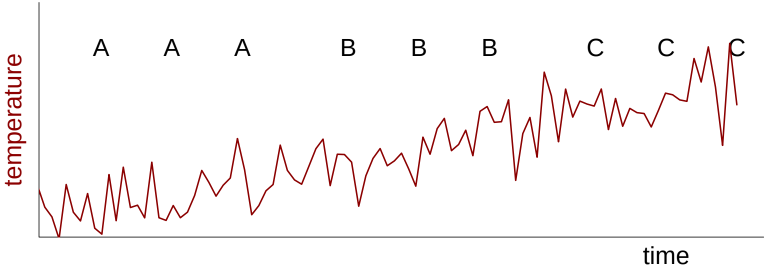
What happens if we don't randomize?



What happens if we don't randomize order?



What happens if we don't randomize order?



Summary: Why is randomization important?

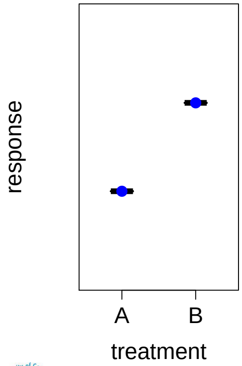
- average out systematic effects (anticipated or not), extraneous factors not directly controlled (prevent **confounding** with unknown factors)
- ensures that treatment is independent of experimental unit's properties

ED Principle 2: Replication

each treatment is assigned to > 1 *experimental unit, randomly and independently*

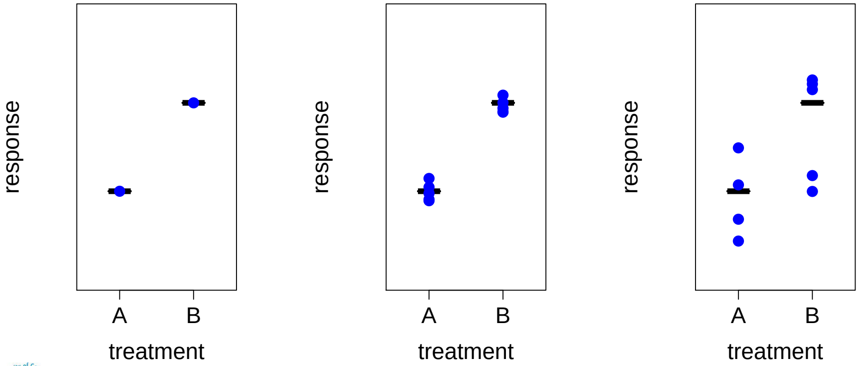
Uncertainty – Variability

If all experimental units were exactly identical, it would be easy to establish cause and effect relationships.

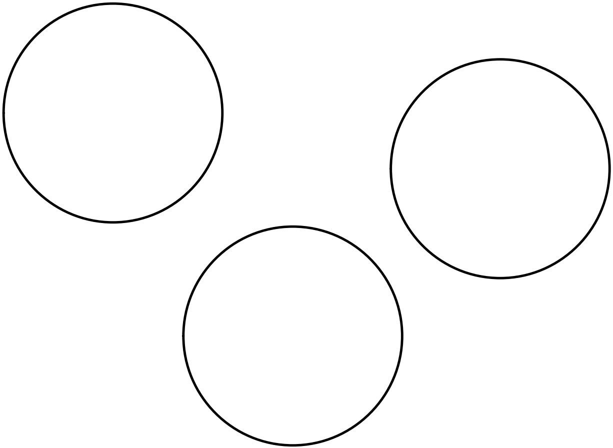


Uncertainty – Variability

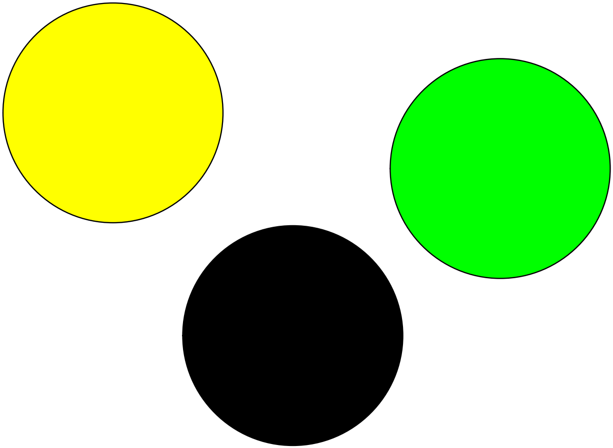
If all experimental units were exactly identical, it would be easy to establish cause and effect relationships.



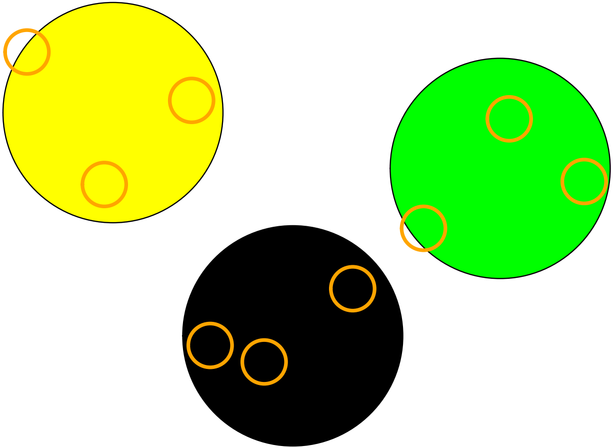
Is this proper replication?



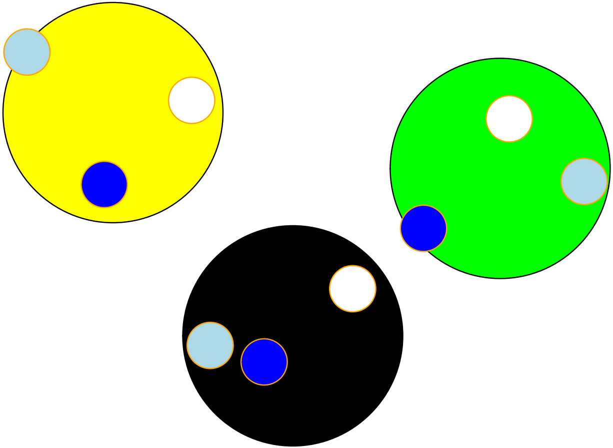
Is this proper replication?



Is this proper replication?



Is this proper replication?

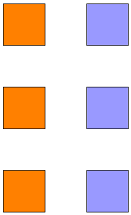


ED Principle 3: Reduce unexplained variation

1. keep all other factors constant
2. blocking

Blocking:

is the grouping of similar (homogeneous) experimental units into blocks.



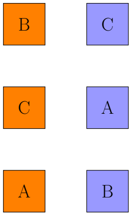
- The experimental units in one block are similar.
- Each treatment occurs once in each block.
- Common blocking factors: age, spatial location, time (day or year)
- each block is a replicate

ED Principle 3: Reduce unexplained variation

1. keep all other factors constant
2. blocking

Blocking:

is the grouping of similar (homogeneous) experimental units into blocks.



- The experimental units in one block are similar.
- Each treatment occurs once in each block.
- Common blocking factors: age, spatial location, time (day or year)
- each block is a replicate

Controls

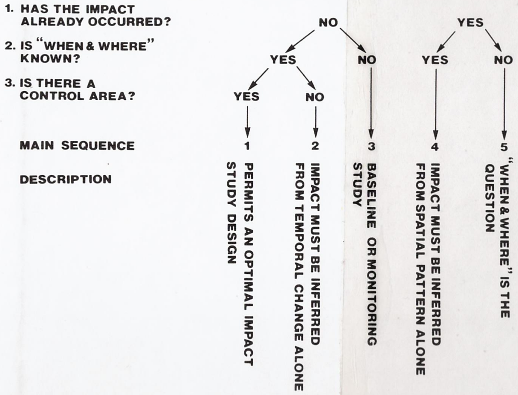
- to measure the effect of a treatment, we need to compare response with to without treatment
- eliminate alternative explanations for results

Part 2: Constrained Design Studies

One or several principles of experimental design cannot be adhered to:

- impact studies
- observational studies

Sampling Designs for Impact Assessment



The decision key to the "main sequence" categories of environmental studies.

from Green 1979, John Wiley & Sons

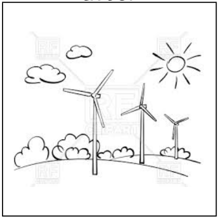
Example: Impact of a new wind energy farm

Problem formulation:

- not an experiment, but impact assessment
- cannot randomly choose location
- only one farm (no replication of 'treatment')
- impact: effect of wind farm on bird behaviour around turbines
- hypothesis: change in the local density of raptors with presence of wind farm compared to absence of wind farm
- sampling method: vantage point surveys

Control and Replicate Time and Location

after



Control and Replicate Time and Location

before

temporal control

after

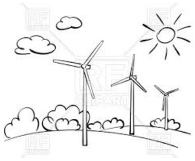
Control and Replicate Time and Location

before

temporal control

o o o: temporal replication

after



o o o

Control and Replicate Time and Location

before

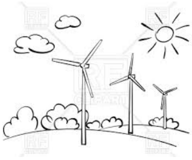
temporal control

o o o: temporal replication

spatial control

o o o: temporal replication

after



o o o

spatial control

o o o

Random Sampling

Purpose: sample must be representative (of the population)

- where: where to place the vantage points
- when: times of day, season, weather conditions

Observational studies

No manipulation

- Cause and effect unclear
- Influence of unmeasured variables can never be ruled out (confounding)

Observational studies

No manipulation

- Cause and effect unclear
- Influence of unmeasured variables can never be ruled out (confounding)

Two things to worry about:

1. Observation process (Solution: distance sampling, repeated visits, multi-observer designs)
2. Representativeness

Observational studies

1. Identify target population

Observational studies

1. Identify target population
2. Random sampling: no assumptions about the population needed

Observational studies

1. Identify target population
2. Random sampling: no assumptions about the population needed
    - every unit has a known probability of being in the sample
    - sample is drawn according to these probabilities
    - selection probabilities are used when making estimates

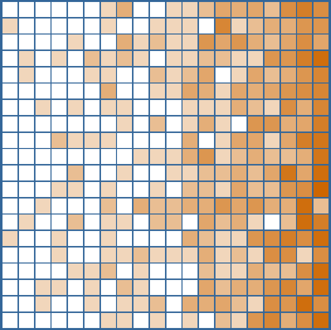
Observational studies

1. Identify target population
2. Random sampling: no assumptions about the population needed
    - every unit has a known probability of being in the sample
    - sample is drawn according to these probabilities
    - selection probabilities are used when making estimates
3. Model-based sampling: unbiased if model describes population well

Random sampling

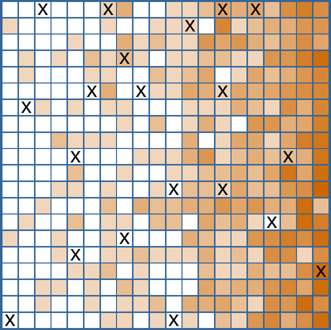
- Simple random sampling
- Stratified random sampling
- Cluster sampling
- Multistage sampling
- Adaptive sampling

Density of hadeda ibises



What is the average density of hadeda ibises in this landscape?

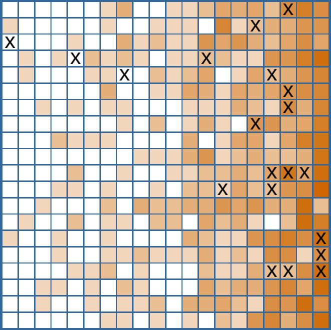
Simple random sampling



$\mu = 1.98$
$\bar{x} = 1.35$
$95\% CI : 0.21 - 2.49$

Unbiased regardless of how densities are distributed.

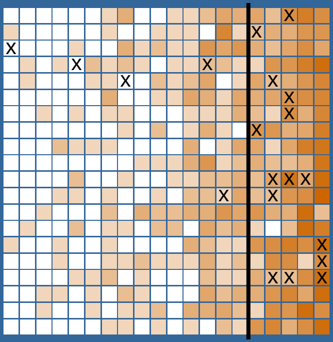
Convenience sampling



$\mu = 1.98$
$\bar{x} = 3.05$
$95\% CI : 1.91 - 4.18$

Preferentially sampling high-density areas leads to estimates that are biased high.

Stratified random sampling



$\mu = 1.98$

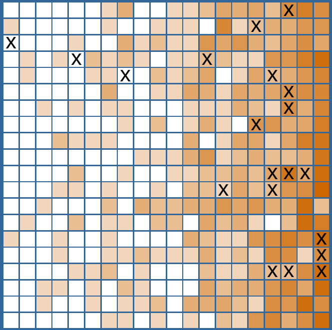
$\bar{x} = \frac{1}{N} \sum_{h=1}^{L} N_h \bar{x}_h = 1.75$

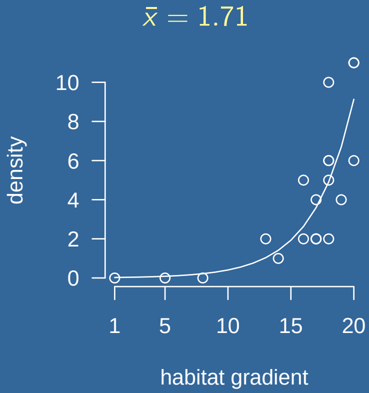
$s^2 = \sum_{h=1}^{L} (\frac{N_h}{N})^2 (\frac{N_h - n_h}{N_h}) \frac{s_h^2}{n_h}$

$95\% CI : 1.05 - 2.44$

Strata should be as homogeneous as possible; blocking is a related idea in experimental design.

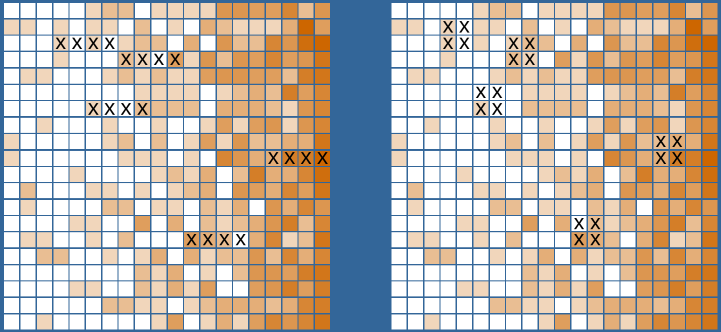
Model-based sampling





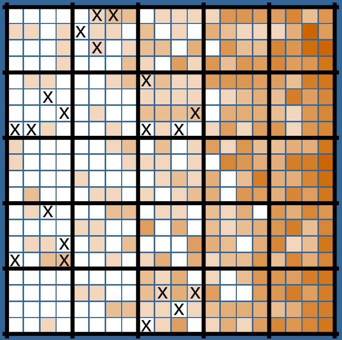
Can be unbiased if the model is good; no random sampling needed.

Cluster sampling



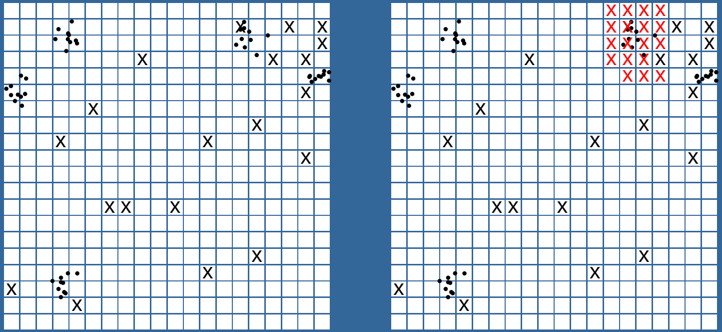
- Clusters selected at random
- Most efficient if variability within cluster is high and variability among clusters is low

Multistage sampling



1. Select large units at random
2. Select small units at random within selected large units

Adaptive sampling



1. Select units at random
2. If species detected, sample surrounding units

Random sampling

- Probability of being in the sample is known for each unit
- This probability is used when making inference
- Lots of designs to choose from
- Some designs are more effective than others for your situation
- Things to think about:
    - Scale of variability and spatial correlation (do a pilot)
    - Do you have informative covariates? (stratification; model-based sampling)
    - Costs of travelling to sites
- Things get complicated quickly

Summary: Nine principles of good study design

1. formulate problem (design, predictors, hypothesis ...)
2. replicate time, location
3. random sample
4. control (space and time)
5. conduct a pilot study
6. worry about observation process
7. subsampling, stratified sampling where necessary
8. appropriate sample unit size
9. use an appropriate statistical model

Adapted from: Green 1979, John Wiley & Sons

Key references

Experimental design:

- Underwood, A. J. 1997. Experiments in ecology: their logical design and interpretation using analysis of variance. Cambridge University Press.
- Montgomery, D. C. 2012. Design and Analysis of Experiments, 8th edition. Wiley.

Constrained designs / impact studies:

- Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. John Wiley & Sons.

Survey design:

- Thompson, S. K. 2012. Sampling. 3rd edition. John Wiley & Sons.

SEEC Stats Toolbox

Want to broaden your stats knowledge? Unsure of what you can do with your data? Still developing your proposal?

Join us for the monthly **SEEC Stats Toolbox** seminars where we introduce you to statistical methods that are useful for ecologists, environmental and conservation scientists.

Our next seminar:

Topic: **Intro to Multivariate Analyses**

Who: Dr Natasha Karenyi

When: **Thursday 20 April 2017 (1-2pm)**

Where: PD Hahn Lecture Theatre 3,
PD Hahn Building Level 5, UCT

More details: www.seec.uct.ca.za

SEEC - Statistics in Ecology, Environment and Conservation