Design

- Sampling design
  - observational
  - estimation
- Experimental design
  - manipulative
  - causal inference

To find out what happens when you change something,
it is necessary to change it

Box, Hunter, and Hunter (1978)

Design fundamentals

- Identify a population of inference: scope
- Identify sample or experimental unit
- Confounding — main issue
- Replication
- Randomization
- Control

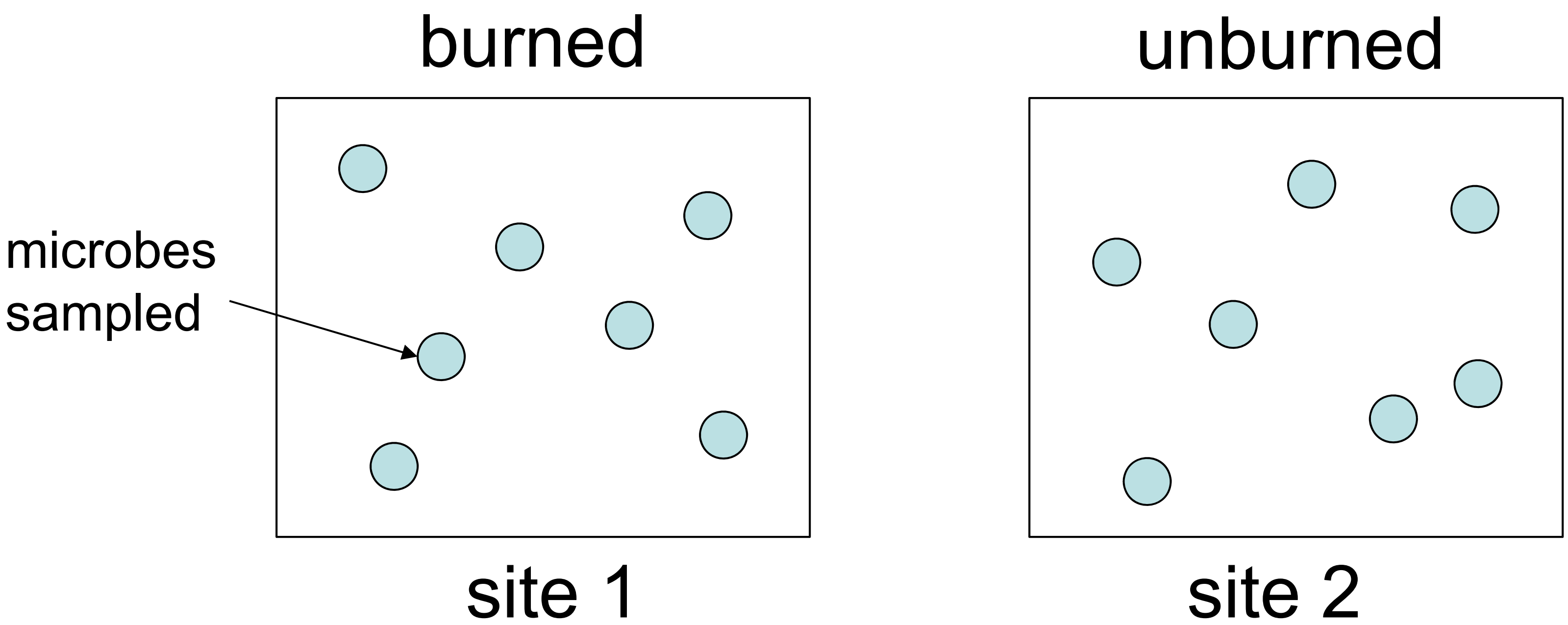
main remedies

Replication

- 1 replicate = confounded with unit
- How much replication?
  - depends on effect size and variance
  - rule of thumb: < 20 d.f. is treacherous
- Degrees of freedom (d.f.)

  = n – number of parameters

Confounding examples



burn and site are confounded

Confounding examples

Confounding in time

e.g. Process all of treatment 1

before lunch.

time 1

Environment 1

Process all of treatment 2

after lunch.

time 2

Environment 2

Pseudoreplication

- Replicates are grouped
- Grouping = confounding

Randomization

- Fixes confounding by shuffling potential confounders
- Random sampling: allows inference to population (scope)
- Random assignment: allows causal inference about a treatment

Simple random sample

- Number each individual in the population
- Use a random number generator to draw individuals at random
- Unbiased sample
- Ensures unbiased estimate

Stratified random sample

- Divide the statistical population into sub-populations
- Random sample within sub-populations
- Examples
  - male/female
  - different habitat types
  - species 1 / species 2

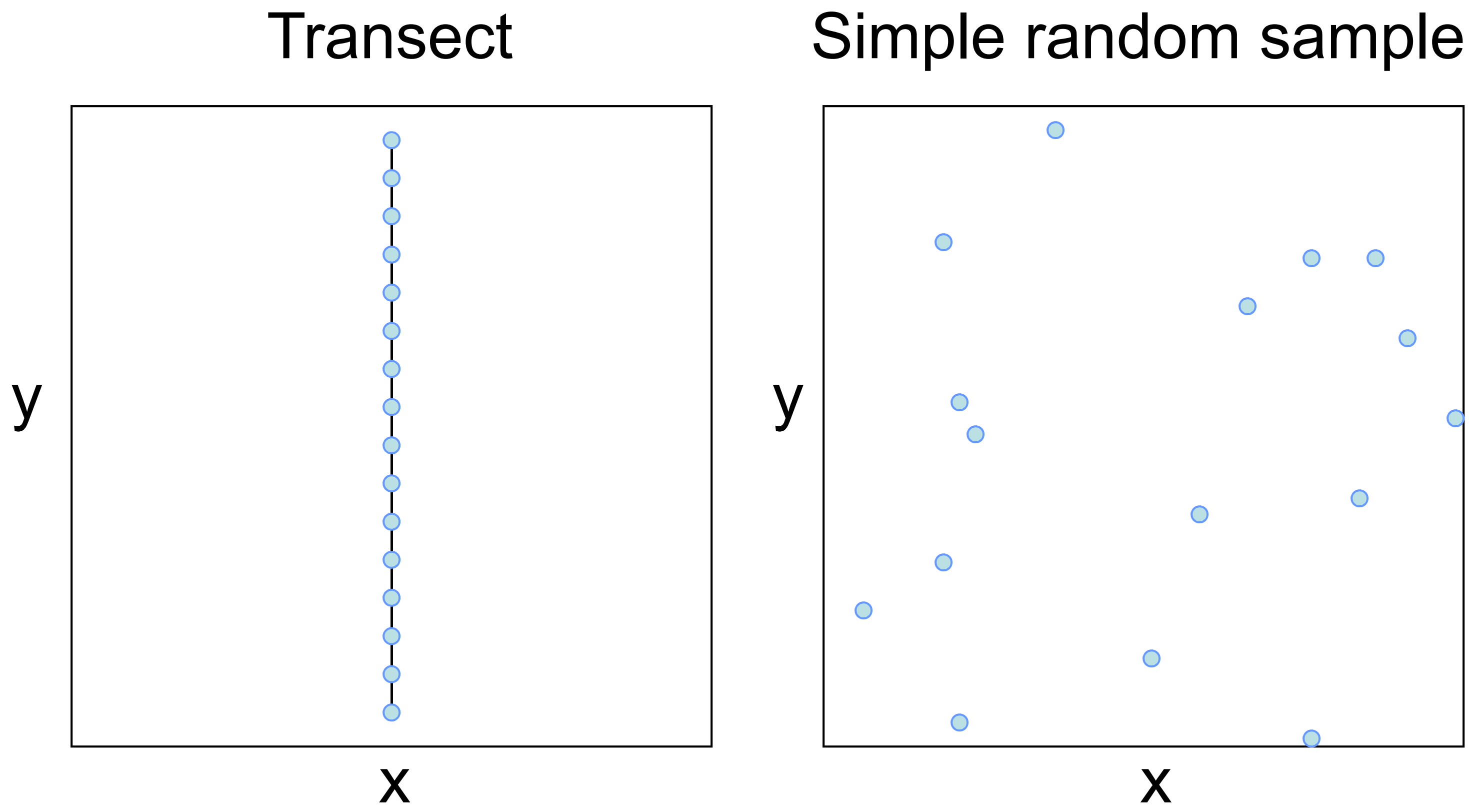
Nested random sample

- Example
  - trees / branches / leaves
- Randomly sample trees within forest
- Randomly sample branches within trees
- Randomly sample leaves within branches
- Scope: leaves within a forest

Systematic sampling

- Opposite of random
- Examples
  - transects with equal spacing of samples
  - spatial grid
  - every Thursday
- Bias
- Autocorrelation
- Scope

Example: spatial sample

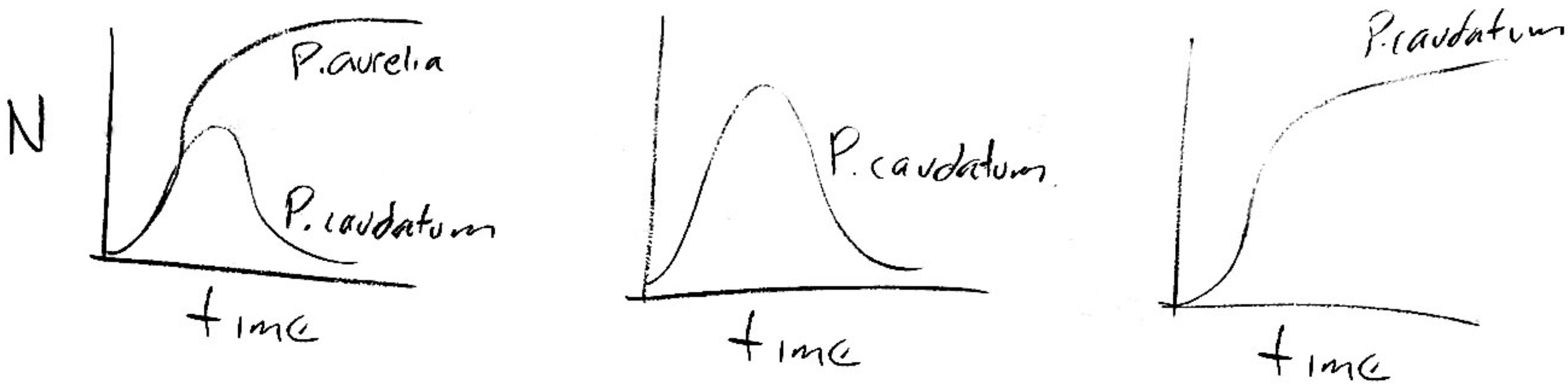


	Transect	Simple random sample
Bias:	one x; gradient on y?	none
Autocorrelation:	strong, systematic	weak, diffuse
Scope:	this transect	population

Controls

Systematically controlling for potential confounders

Classic example: Does *Paramecium aurelia* exclude *P. caudatum*?



P. aurelia present
actual outcome

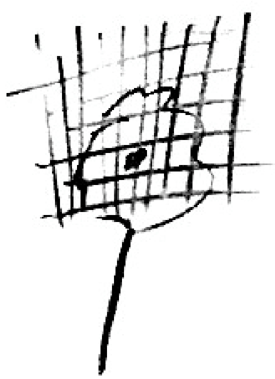
P. aurelia absent
possible outcome 1

P. aurelia absent
possible outcome 2

Presence of *P. aurelia* is confounded with time.
We need a control (absence of *P. aurelia*) to distinguish the two possible outcomes through time.

Controls

- Cage effects
  - examples: pollinator exclusion, herbivore exclusion
  - exclusion is confounded with changes to the environment caused by the cage



mesh bag

control 1: no bag

control 2: bag with holes

e.g. exclude pollinators from flowers. Control 2 attempts to measure confounding effect of environment while allowing pollinators. Very difficult issue to control.

Controls

- Handling effects
  - confounder: handling changes behavior
- Example: hormone treatment
  - catching and injecting an animal changes it's behavior
  - control: catch and sham inject

Factorial design

Factor B

Levels of B

		B1	B2
Factor A	A1	10	10
	A2	10	10

Levels of A

Number of replicates

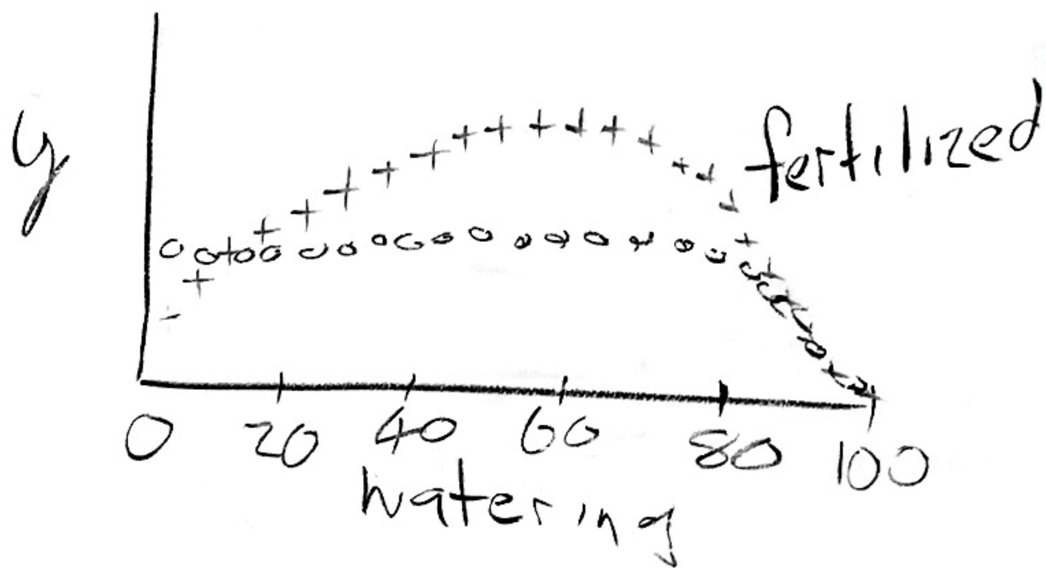
2 x 2 factorial

Advantage: allows us to estimate interactions

Factorial versus response surface design

		Water 20	40	60	80	100
Fertilize	+	5	5	5	5	5
	-	5	5	5	5	5

50 experimental units
no interaction
# parameters = 7
df = 50 - 7 = 43
with interaction
# parameters = 11
df = 50 - 11 = 39

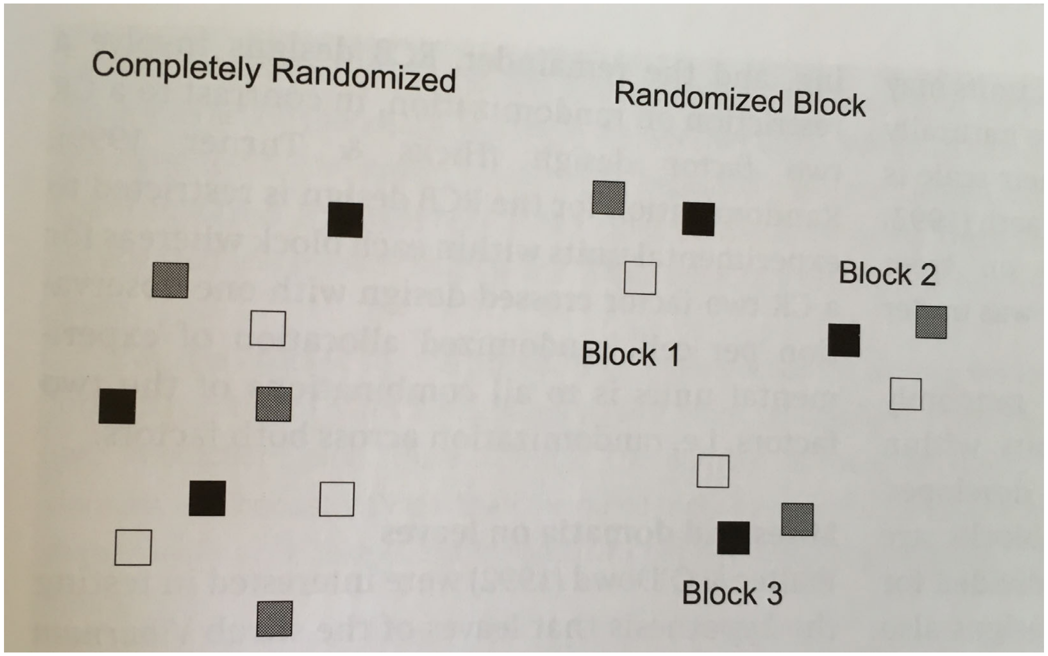


50 experimental units
3 parameters per curve
df = 50 - 7 = 43
5 parameters per curve
df = 50 - 11 = 39

Advantage: can get much better nonlinear resolution for same replication

Multilevel designs

- Randomized block



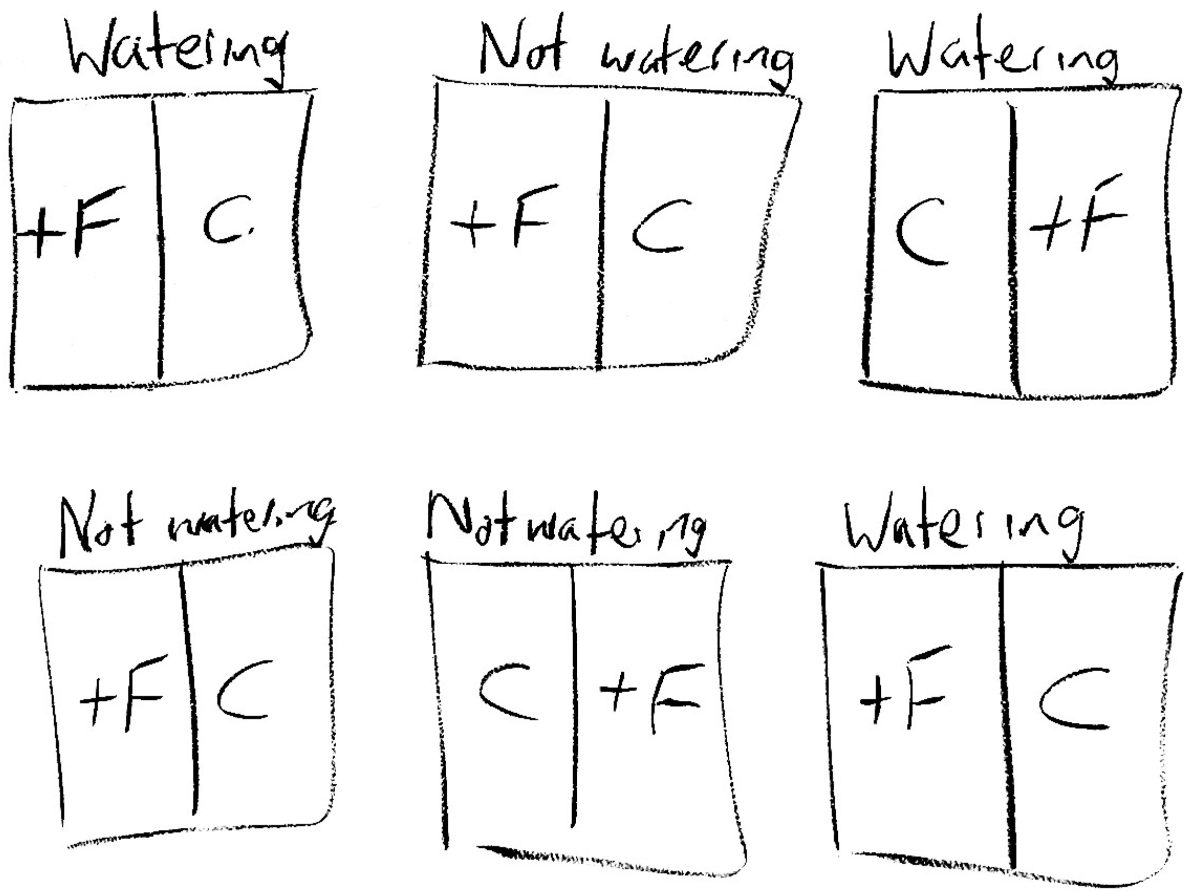
Example spatial design with three treatments (box colors)

Contrasted with completely randomized design

Pros: account for large scale variation
Cons: penalty for more complex model
Whether it helps depends on this tradeoff

Multilevel designs

- Split plot



Plots are split into sub-plots.

Watering treatment is at large scale (plot), fertilizer treatment is at small scale (sub-plot).

Pro: watering simpler
Con: replication of large scale factor is reduced (3)
Con: penalty for model complexity (need a grouping variable)

Boulder county trails

Example from ecology undergrad field class
Effect of distance from hiking trails

How would you design this?

Think about this in terms of
- scope of inference
- amount of replication
- logistics
- grouping vs no grouping
- autocorrelation
- pseudoreplication