Experimental design

Experimental design or why care about all this complexity?

- Experimental design is the development of plans for experimentation that recognize in advance statistical issues
- Good experimental design recognizes sources of uncontrolled variation and plans statistics around them
  - Increases power!

Non-demonic intrusion

- Avoid random confoundment

+N +N 0 0

Plants on sandy soil grow worse dominates nitrogen effect

An experiment is ...

- Controlled – comparison
  - Avoids year-to-year intrusion
- Manipulated – factor changed by experimenter
  - Avoids covariate intrusion in observational data
- Replicated – multiple instances measured
  - Avoids chance intrusions
  - Allows quantification of possibility of chance intrusion

Basic design

- One or more factors
  - E.g. add nitrogen, add phosphorous, add water
- One or more levels
  - E.g. +0, +5kg, +10kg

The key tradeoff

- Factors vs. levels vs. replicates
- Have a fixed # of measurements
  - One field season – need to measure over two week period for consistency
  - Can do 10/day x 6 days/week x 2 weeks
  - → 120
- Could allocate to:
  - 40 levels by 30 levels in 2 factors w/ no replicates
  - One factor with two levels and 60 replicates
  - And …
- # one missing thing from QE proposals

Correct answer

- Depends on the question
  - Usually 2-5 levels is adequate
  - Usually 1-3 factors is most that is manageable
  - Gotelli & Ellisons – rule of 10 – absent other information need more replicates
    - If seeking small effect sizes in noisy data need more replicates

Simplest case (one-way ANOVA)

- One factor (e.g. add nitrogen)
- Use control (untouched)
- Process control (touched as much as treatment)
- Treatment(s) (1-n)
- Randomize the site/treatment assignments with eye to interspersion
- Gotelli & Ellison – rule of 10 – 10 replicates per treatment
  - Get 10 replicates before adding treatment levels

Randomization

- Randomization:
  - Needs to be truly random beyond control of experimenter (coin flips, computer generation)
  - Prevents experimenter bias
  - Prevents non-demonic intrusion
  - Prevents confoundment if large enough replicates
- There are alternatives
  - Stratified/blocking – known variance to be managed
  - Interspersion better than random?

Example of randomization in field

+N	+H_2O	-
+H_2O	-	+N
+N	-	+H_2O

Sampling designs

- Random – choose objects to study randomly
- Stratified – choose objects to study by traits
  - Assumes know key traits
  - Good to ensure get rare objects studied
- Stratified-random
- Cluster-random or cluster-stratified
  - Related to blocking & repeated measures
- Example – elevation
  - Random – sample 0-6000 m from random # generator 6 times
  - Stratified – 3 each from 0-2000, 2-4, 4-6
  - Stratified-random
    - Many
    - Sample randomly, rejecting designs that don't have at least one in each of 0-2, 2-4,4-6,000

Hurlbert redux

- Controls (temporal change, procedural effects)
- Randomization & blind measurement if subjective (Experimenter bias)
- Noise (replication)
- Inherent variability among units & confoundment (replication, interspersion, concommitant measurements)

Hurlbert argues

- Interspersion more important than randomization usually
- Pseudoreplication is common – multiple examples of non-independent entities
- Only pseudoreplication if use NHST statistics (no statistics OK)

Oksanen

- 20 years of history
  - Pseudoreplication like “communist” in the McCarthy era
- Agree
  - Non replicated experiments not always realistic, necessary for science
- Disagree
  - Statistics should be used even if pseudoreplication
  - Pseudoreplication pejorative

One more thing to remember

- Pseudoreplication depends heavily on the question
- Measurements of leaf size
  - Are measurements of multiple leaves from one tree pseudoreplication?
    - Studying genetic variation in the population
    - Studying phenotypic plasticity
    - Studying shade vs. light leaf patterns

To cross or not to cross

- Example: add nitrogen, phosphorous, 3 levels each
- How to organize:
  - One-at-a-time
  - Factorial
- Appears 1 at a time more efficient (3 vs. 4)
- But factorial …
  - has replicates for individual factors built in (really 4 vs 6)
  - Allows interaction terms (effect of N increased or decreased by P?)

0	+N	+P

	0	+N
0	0,0	0,+N
+P	+P,0	+P,+N

Designing factorial experiments

- Replicates - 10 replicates/treatment harder but include replicates!
  - No interaction term without replicates
- Balanced – same # of replicates per cell
- Full vs. partial
  - Usually an issue for 3+ factors – can't do all combinations so do a carefully chosen subset

Fractional factorial

- 2^{3-1}
  - Two levels per factor, 3 factors=8
  - But do only 4 (2^2)
  - Effect of X1=(4+6)/2-(1+7)/2
  - Must remain
    - Balanced (every variable at high 4 times)
    - Orthogonal
- Assumes weak or no interaction terms
  - Saves cost/effort
- Also response surface methodology
  - Measures y as a function of continuous x variables

	X1	X2	X3
1	-1	-1	-1
2	+1	-1	-1
3	-1	+1	-1
4	+1	+1	-1
5	-1	-1	+1
6	+1	-1	+1
7	-1	+1	+1
8	+1	+1	+1

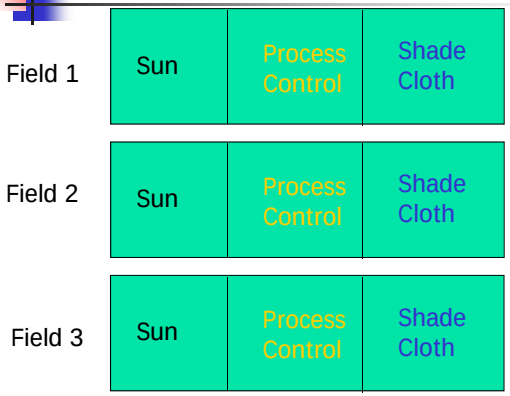
Blocking and repeated measures

- Recall in ANOVA
  - F=MS-between-groups/MS-within-groups
    - F=treatment effect/noise
    - Big F means reject null
- Deal with known-unknown sources of variance
  - Know variance between plots exists but don't know (and don't care) about its nature
  - Inflates ε which is denominator in F test
  - Reduces significance
- If you can pull "known unknown" variation into a systematic term and out of noise, you increase power

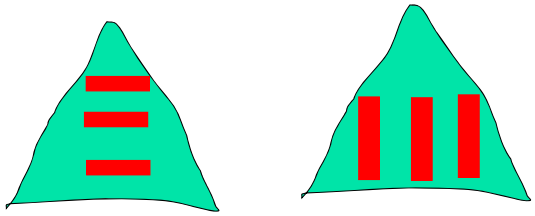
Blocking

- Lump similar (usually spatially close) plots together
- Analagous to the pairwise t-test
  - E.g. t-test where you have pairs of siblings & one gets treatment and one doesn't
- Called "within-subject" in psychology
  - Also equivalent to giving same plant two different treatments and measuring response each time
- Pulls variation into the block and out of the explained error term (ε)
- Treat like a two-way ANOVA
  - $y=t_i+b_j+\varepsilon$ or in R: y~b+t (block goes first)
  - Is B random or fixed?
  - Does treatment by block interaction matter?
  - Highly controversial – find arguments for all four combinations
    - But debate is sub rosa

Blocking design



Good vs Bad Blocking



Repeated measures (ANOVAR)

- Also called within-subject, longitudinal, panel
- Similar to blocking but the "plot" is an individual organism and the blocking factor is time
- Many scenarios:
  - Using same individuals to reduce variance (Photosynthesis~Time+CO_2Treatment)
  - Growth, recuritment or other variable changing over time (Mass~ Time+Diet+Diet:Time
  - Before/After (Number of individuals~Stream+BeforeAfterNitrogen

ANOVAR & Sphericity

- Sphericity (also circularity) is an extension of the assumption of homoscedasticity
- Data is spherical if the variance of the difference between different treatment levels (summed across blocks/individuals) is constant
- If only have two levels can't be a problem
- In blocking this is usually a good assumption
- In ANOVAR this is a good assumption if time is unimportant or if time intervals are equal
- Dealing with
  - Mauchley's test of sphericity (but sensitive to normality)
  - If not spherical, correct df, SS with Greenhouse-Geiser or Huynh-Feldt epsilon corrections gives correct p

	T=0	T=10	T=100
Plant A	5	6	15
Plant B	4	6	16

$[(6-5)^2+(6-4)^2]/2$ vs.
$[(15-6)^2+(16-16)^2]/2$

Fractional blocking or BIB

- An extension of fractional crossed designs
- Two-way (sun/shade and 0/+N) or one factor + blocking
- Ideally want a “complete” or “full” or “orthogonal” design (same # of replicates for each pairwise combination)
- Not always possible
  - Have 3 growth chambers or using prexisting plots that are divided into 3 but want four nitrogen levels
  - Next best is a fractional design (BIB design=balanced incomplete block)
  - Same # of replicates for each treatment
  - Model same as complete block (y~b+t)

ANOVA – unbalanced or not fully crossed

- Unbalanced – all cells have data, but some cells more than others
  - Plague of locusts
  - Observational (non-experimental data)
- Three choices
  - Randomly throw away a site everywhere else (best if have the power)
  - Add an average site (OK and more realistic)
  - Run calculations on unbalanced data

Calculations on unbalanced data

- Everything works nicely in balanced calculations but starts to fall apart in unbalanced
- There are six ways of calculating sum-squares for ANOVA – Type I-VI
  - All the same when balanced
  - Type I takes terms in order specified
    - Good in that sum squares add up
    - Jarring to have dependent on order
    - Statisticians say this is best – have to think order out
  - Type II
    - Social scientists like this
  - Type III
    - Independent of order
    - Sum squares don’t add up
    - Ecologists say this is best
  - Type IV-VI ignored
- Calculations:
  - Most packages only Type I
  - SAS can add a /SS1 or /SS3 switch
  - R: anova command does type I
  - R: Anova command (package car) does type II, III

Still worse: missing data

- If missing cells
  - E.g. no aquatic passerines
- Also if mixed (random + fixed)
- Then even type III doesn’t work
  - Need to use maximum likelihood
  - Systat doesn’t do this
  - SAS has proc Mixed (vs. GLM)
  - R has lme (package nlme) vs lm

Nested models

- A factor that clearly is contained within another:
  - Batches/brands of pesticide within pesticide type
  - Genus within family
  - Subplots within plots
- Really a spectrum
  - Fully crossed (balanced, orthogonal)
  - Balanced incomplete block (also orthogonal)
  - Unbalanced
  - Missing data
  - Nested

Spectrum

Orthogonal/Balanced

	0N	+N	++N
Wild	4	4	4
Bred	4	4	4

Balanced Incomplete
(each field holds one breed
Only room for 2 nitrogen treatments,
3 fields per breed (0,+; 0,++; +,++)

	0N	+N	++N
Wild	2	2	2
Bred	2	2	2

Unbalanced

	0N	+N	++N
Wild	3	4	4
Bred	4	4	4

Missing

	!Pass	Pass
Aquat	20	0
Terrest	23	102

Nested is end of spectrum

	DDT	Roundup
DDT Batch1	5	-
DDT Batch2	5	-
Roundup Batch 1	-	5
Roundup Batch 2	-	5

DDT	Batch 1	5
	Batch 2	5
Roundup	Batch 1	5
	Batch 2	5

Terms for design

- Scale 1
  - Balanced
  - Unbalanced
  - Missing
  - Nested (special structure missing)
- Scale 2
  - Complete (all treatment combos in each block)
  - Incomplete/fractional/partial

Split plot design

- Two treatments + blocking + nesting
- Lump all levels of factor 1 within a block (site)
- Randomly assign sites to levels of factor 2
- Example (Gotelli)
  - Substrate: Cement/Slate/Granite
  - Predation: Control/PredInCage/PredExcludeCage
- Each site can have only one of Predation but can have all three substrates
- Y~P+B/S

ED for Regression

- Make sure get full range of independent variable (as large a range as possible)
- Make sure sample uniformly over range

BACI ED

- Before-after, control-impact
- Used when cost of manipulation prohibitively large to replicate in space
- EG effect of acidification of lake on rotifer populations
  - May only be able to acidify one lake
  - Can pick another as a "control"
  - No replication, no variance, no error, no GLM?
- "Replicate" in time instead
  - Before perturbation is control
  - After is treatment
  - Replication comes from many observations over time
  - Form of ANOVAR
  - Best analyzed as time series data
- Pulse vs. press distinction
  - Pulse acidify and allow to return to normal
  - Press acidify and keep at near constant acidfication

Summary Experimental Design

- Fixed vs. random
  - Need to know definition
  - Impacts on calculation (lme, proc mixed)
- Fully crossed&balanced/BIB/ unbalanced/nested
  - Type III, calculations
- Experimental design
  - Blocking & ANOVAR assign "unknown" variance to a factor and increase power
  - Simultaneously deals with nonindependence
  - Blocking usually treated as a random factor
- ANCOVA & regression ED
  - Explore the full range of independent variable evenly
  - ANOVA-regression design tradeoff
- BACI ED, analysis methods to be covered later