

EXPERIMENTAL DESIGN IN LONG-TERM ECOLOGICAL RESEARCH

Christina Staudhammer, Associate Professor
Department of Biological Sciences
University of Alabama

Data collection and the scientific method

Slide 2

Data collection for a purpose

- The first step in rigorous exploration is formulating testable hypotheses or posing critical research questions
- To apply the scientific method, we must collect data that allow us to discriminate between different hypotheses
 - we collect data to:
 - estimate values of characteristics of the parent population
 - conduct hypothesis tests
- Before we collect data, we *plan and design data collection procedures in support of those hypotheses and/or questions*
- Data should be collected with a *purpose*
 - Independent variables (for explanation)
 - Dependent variables (for inference)
 - *Your research hypotheses/questions define what variables need to be measured*

The importance of planning your study design

Even if you do not do an "experiment", the concepts of experimental design are essential in supporting your research

- Data are collected with the purpose of supporting/refuting specific pre-formulated hypotheses

Data collection and the scientific method

Slide 4

Empirical (mensurative) approaches

- Empirical approaches are purely observational, e.g.:
 - Incremental growth of old growth trees in Boreal forest by species
 - Soil efflux in a longleaf pine forest over time
 - Change in NEE due to hurricane

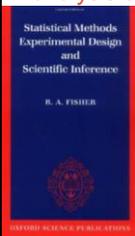
Experimental (manipulative) approaches

- An experiment is a set of actions and observations, performed to verify or falsify a hypothesis or research a causal relationship between phenomena.
 - a force applied to an experimental material: a *treatment*, e.g.:
 - Change in annual NEE in response to fertilizer
 - Change in NEE when exotic species are removed
 - Incremental growth of old growth trees under rainfall exclusion

Data collection and the scientific method

Slide 5

Requirements for statistically defensible analysis of data



- Randomization
 - Why?
- Replication
 - Why?
- Design Control
 - What does this mean?

Assures that our own biases do not enter the data.

Necessary to meet assumption of required by most statistical tests

Permits calculation of experimental error, "insurance" against chance events, Averages out "noise"

Use homogeneous experimental/sampling units, OR If material is heterogeneous, then use blocking

Randomization

- Random sampling ensures that population parameter estimates are unbiased, e.g.:
 - Plants randomly selected from population of interest
 - Fixed area plot locations randomly selected from within study area
- If we do not obtain a random sample, we reduce our inferential population
- Experimental units should be randomly allocated to treatment groups

Independence

- Lack of independence can arise over space, time, or can be due to genetics
- While independence is necessary for basic statistical techniques, there are often ways of appropriately accounting for dependencies in statistical analyses (more on this later)

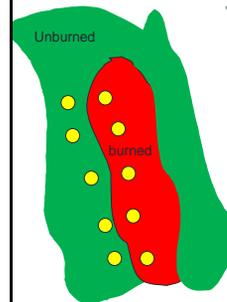
Replication

- In order to analyze data, we must have multiple observations of each factor combination we are interested in
 - If we have one factor we are interested in (e.g. two species), we must have at least two observations per species (4 obs) in order to assess the variability within species and between species
 - BUT NOTE: two is dangerous – what if one individual dies?
- Replication reduces the chances that we have inherent consistent differences in experimental units that receive the same treatment
 - i.e., we can be more confident in attributing differences to treatments rather than other factors

Replication, pseudoreplication, and confounding

- Biologists in particular often find it difficult to replicate the exact same conditions, e.g.:
 - Are two pots of soil the same?
 - Are two rivers the same?
- To properly replicate conditions, “pseudoreplicates” are often chosen, e.g.:
 - To assess the effect of burning on soil nutrient composition, we sample from adjacent burned and unburned areas

Example of sampling from burned and unburned areas?



- Are these really replicates?
 1. Since land is likely heterogeneous, these are not true replicates, but they are as good as it gets in ecology!
 2. Since the fire was applied to the entire area, we really have only one true replicate (in each of unburned and burned areas) with pseudoreplicates, or subsamples
 - We need *multiple fires* in order to appropriately evaluate impact of fire in general; otherwise, our inference is only to this fire

Pseudoreplication

- Pseudoreplication, according to Hurlbert (1984) is:

“the use of inferential statistics to test for treatment effects with data from experiments where either treatments are not replicated (though samples may be) or replicates are not statistically independent.”
- Heffner et al (1996) revisited this concept and distinguished a pseudoreplicate from a true replicate, which they defined as:

“the smallest experimental unit to which a treatment is independently applied.”
- True replication is required in order to make statistical inference, as it permits the estimation of variability within a treatment.

What is meant by “experimental design”?

Controls how we apply *treatments* to observational units, or select data from different populations

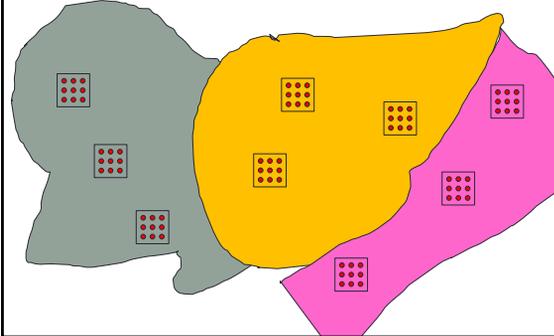
→ Controls how we analyze the data

- is often intimately related to the sampling design under which the data was collected

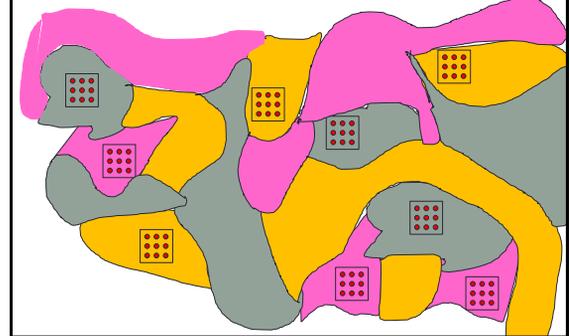
E.g., we want to describe longleaf pine regeneration in a 90 ha area with 3 understory types (20 ha in A, 30 ha in B, 40 ha in C)

- each understory type covers a contiguous and non-overlapping area, so we choose 3 1-ha areas, and within each install 9 grid plots
- OR, each understory type is patchy over our study area; we choose 3 random areas of each type, and within each install 9 grid plots

One experimental design option



Another experimental design option



What is meant by “experimental design”?

- 2

Controls how we apply *treatments* to observational units, or select data from different populations

→ Controls how we analyze the data

- is often intimately related to the sampling design under which the data was collected
- E.g., we want to describe disease presence in frogs under three diet regimes (9 each of low, medium, high protein), and have 3 blocks of space available (in three different locations)
 - In block #1, we observe 9 frogs with low protein, in block #2, we observe 9 frogs with medium protein, and in block #3, we observe 9 frogs with high protein
 - OR, 3 frogs with each of the diet regimes in each of block #1, #2, #3

One experimental design option

A	A	A	B	B	B	C	C	C
A	A	A	B	B	B	C	C	C
A	A	A	B	B	B	C	C	C

Another experimental design option

A	B	C	C	A	B	B	C	A
A	B	C	C	A	B	B	C	A
A	B	C	C	A	B	B	C	A

How are these designs different? Under what circumstances is each design more appropriate/more efficient

Assumptions of “traditional” statistical hypothesis testing

Note: most tests are robust to moderate violations

1. Samples are from a ~Normal population
 - If population is very skewed or multi-modal, tests not valid
 - Transformation can often fix this
2. Samples are from homoscedastic (equal variance) populations
 - Often, fixing #1 will fix this problem
3. Samples are randomly selected from the population
 - considered in the design stage of your experiment
4. Samples are independent
 - If samples are not independent, however, there are often ways to mitigate it in the analysis process

Sampling Design

- Sampling refers to the act of taking a subset of data to represent the whole
 - You don't have to eat the whole ox to know that it is tough. – (Anonymous, but frequently misattributed to Dr. Samuel Johnson)
- The goal is to improve efficiency of data collection, while avoiding mismatches between the sample and target population
- There are a wide variety of methods that can be employed to obtain individuals for measurement
- The choice of designs depends on
 - Nature and quality of the sampling frame
 - Availability of auxiliary information about units on the frame
 - Accuracy requirements, and the need to measure accuracy
 - Whether detailed analysis of the sample is expected
 - Cost/operational concerns

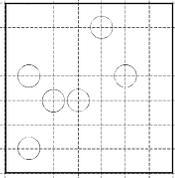
Sampling design Slide 22

Some random sampling methods

- Simple random sampling (SRS)
- Systematic sampling (SyRS)
- Stratified sampling (StRS)
- Others, for example...
 - Cluster sampling (CS)
 - Multi-stage sampling (MsS)
 - Multi-phase sampling (MpS)

Simple Random Sampling (SRS)

- Easiest way to sample
- The probability of obtaining any group of observations is the same
- The probability of obtaining any observation is the same for each sampling unit

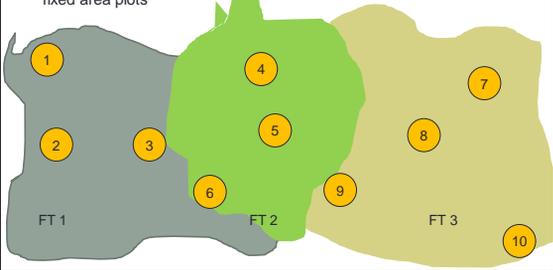


- *Effective only if the population to be sampled is homogenous/uniform variation*
- *If your population is heterogenous, stratified sampling may be more efficient!*

Sampling design Slide 24

SRS example

- For example: we are interested in characterizing biomass across a study area with three forest types, measuring woody plants within fixed area plots



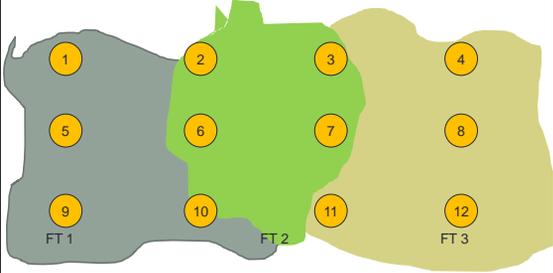
Systematic selection

- First unit is a probability-based selection, subsequent units are not
- Each unit in the population has the same probability of being selected
- Probabilities of different sets of units being included in the sample are not all equal
- Usually better for sampling across an environmental gradient
 - Can be worse if your sample selection coincides with an environmental gradient
 - Not advisable if you wish to characterize spatial variability

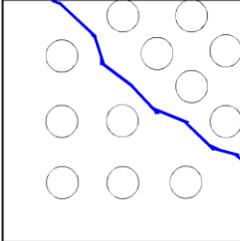
Sampling design Slide 26

SyRS example

- For example: we are interested in characterizing biomass across a study area with three forest types, measuring woody plants within fixed area plots



Stratified Random Sampling (StRS)



Population is subdivided

- sub-populations are sampled (at pre-determined rates) using simple random or systematic random sampling

Use when:

- There are naturally occurring mutually exclusive groups (strata) in the population
- Strata can be determined prior to sample selection

Slide 28

StRS – example 1

- For example: we are interested in characterizing biomass across a study area with three forest types, measuring woody plants within fixed area plots. We sample equally from each FT, though FT2 is smaller.

Slide 29

StRS – example 2

- For example: we are interested in characterizing biomass across a study area with three forest types, measuring woody plants within fixed area plots. We proportionally sample from each FT by area.

Comparison of stratified sampling to SRS

- Tradeoff between increased study cost/ complexity and the potential for increased efficiency
 - more accurate population parameter estimate may be obtained for the same number of sample units as SRS when the variability among strata is maximized and the variability within strata is minimized.
 - the same accuracy may be obtained, but with fewer sample units
- The decision to stratify must take into account how the cost of stratifying relates to the expected increase in estimate accuracy
- NOTE:
 - when modeling data obtained via stratified sampling, the strata must be included as a model factor or weight
 - The average and standard error of the stratified data are weighted

Slide 31

Simple random sampling with sub-sampling

- Secondary (or sub-) samples are selected randomly or systematically within primary samples
- NOTE: The average and standard error are calculated in a two-stage process, using primary sample averages

Some types of experimental designs

- Common Designs
 - Completely randomized design (CRD)
 - Randomized complete block (RCB)
 - Split-Plot Design (SPD)
 - Others (e.g., Latin Square Design)...
- Methods of treatment application
 - Repeated measures experiments
 - Factorial experiments

NOTE that experimental design concepts apply to both mensurative and manipulative experiments

Completely randomized designs (CRD)

- Treatments are randomly assigned to experimental units
- Units are randomly selected for the experiment from among the set of interest
- We assume that units are approximately homogeneous
 - E.g., we sample understory biomass in 0.01 ha plots under three irrigation regimes!

Irrigation A	Irrigation B	Irrigation C			

Completely randomized design – Analysis of Variance (ANOVA) table

- We would analyze this as a simple one-way ANOVA, or could (equivalently) use regression techniques
- Either is termed a *General Linear Model (GLM)*

Source	Degrees of freedom (DoF)	Mean Squares	F test
Irrigation	$k-1=2$	MS_{IRR}	$F_{(2,21)} = \frac{MS_{IRR}}{MS_E}$
Experimental Error	$k(n-1) = 21$	MS_E	
Total	$kn-1 = 23$		

As the number of experimental units ↑, experimental power ↑

What happens when the number of experimental units ↑?

Where: $k=3$ is the number of "treatments", $n=8$ is the number of experimental units per treatment

What about unbalanced designs? As long as n_i are not "too different", we can still use ANOVA techniques, but $EE\ DoF = \sum n_i - k$ and $Total\ DoF = \sum n_i - 1$

36

Fitting CRD models in R

```
> lm.irr <- lm(biomass ~ irrig, data=data.irr)
> anova(lm.irr)
> summary(lm.irr)
> plot(lm.irr)
> lsmeans(lm.irr, pairwise~irrig)
```

- The function `lm` estimates a linear model ($Y \sim X$) using data in the dataframe `data.irr`
- The function `anova` partitions the variation into its different sources (in this case, irrigation and error), and displays F-tests for each effect
- The function `summary` gives estimates of the model coefficients, standard errors, and t-tests, statistics on the model goodness of fit
- The function `plot` produces graphs to verify assumptions
- NOTE that character-valued X variable(s) are assumed to be categorical predictors, whereas numeric-valued X variables are assumed to be continuous predictors
- If your factors are numbered (e.g., 1=blue, 2=red, 3=green), then you will have to declare the variable as a factor

37

Fitting CRD models in R - output

R output

```
> lm.irr <- lm(biomass ~ irrig, data=data.irr)
> anova(lm.irr)
Analysis of Variance Table
```

Response: biomass

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Irrig	2	2021.0	1010.5	40.374	0.0003***
Residuals	21	525.26	25.0		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

What do these fit statistics tell us?

Residual standard error: 1.09 on 21 degrees of freedom
Multiple R-squared: 0.9406, Adjusted R-squared: 0.9375
F-statistic: 40.4 on 2 and 21 DF, p-value: 0.0003

38

Fitting CRD models in R - output

R output

```
> summary(lm.irr)
```

Call:

```
lm(formula = lm(biomass ~ irrig, data = data.irr))
```

Residuals:

Min	1Q	Median	3Q	Max
-2.9233	-1.2752	-0.2657	1.3976	3.0226

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.0210	0.4111	4.926	0.0011 **
irrig.B	12.1991	0.6022	20.257	4.1e-08 ***
Irrig.C	17.9911	0.6022	29.874	1.7e-09 ***

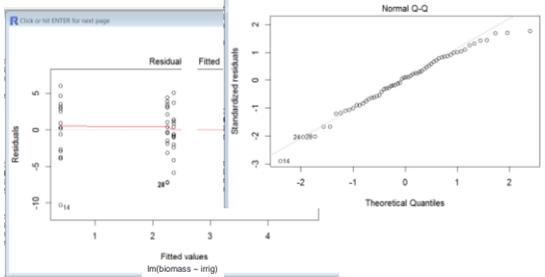
Residual standard error: 1.09 on 21 degrees of freedom
Multiple R-squared: 0.9406, Adjusted R-squared: 0.9375
F-statistic: 40.4 on 2 and 21 DF, p-value: 0.0003

39

Fitting CRD models in R - plots

R output

```
> plot(lm.irr)
```



40

Fitting CRD models in R – marginal means

R output

```
> lsmeans(lm.irr, pairwise~ irrig)
```

```
$lsmeans
  irrig  lsmean    SE df lower.CL upper.CL
  A      20.41037 1.762224 56 16.88021 23.94053
  B      24.77128 1.762224 56 21.24112 28.30143
  C      23.08582 1.762224 56 19.55566 26.61598
```

Confidence level used: 0.95

\$contrasts

contrast	estimate	SE	df	t.ratio	p.value
A - B	-4.360907	2.49216	56	-1.750	0.3082
A - C	-2.675448	2.49216	56	-1.074	0.7069
B - C	1.685459	2.49216	56	0.676	0.9056

F value adjustment: tukey method for comparing a family of 4 estimates

What happens if we measure repeatedly over time?

- For example, we:
 - collect multiple observations within a single experimental unit, e.g., biomass from m trees within n fixed area plots?
 - measure experimental units at multiple time periods, e.g., biomass on n trees in each of t years?
- ? Are observations within plots or measured repeatedly by year independent? probably not!
- ! And if not, we violate an assumption necessary for statistical hypothesis testing
- These are common occurrences in biology and other disciplines!
- Can lead to *pseudoreplication*
- * To appropriately analyze, we need to consider additional non-fixed effects

Models for data correlated over space/ time

- In many situations, researchers collect multiple elements of the same fixed area plot
 - E.g., models of biomass as a function of $k=3$ site qualities: we measure $n=15$ plots that each contain $m=4$ trees (45x4 trees total)
- E.g., models of biomass at $k=3$ sites on $n=15$ trees at $m=4$ times etc...

Models for data correlated over space/ time

- We then want to develop models for these elements
 - For tree-level data collected in fixed area plots
 - trees within the same plot are NOT independent; they are likely more alike than those in different plots
 - For data collected on the same exact trees over time
 - Measurements on the same tree over time are NOT independent; they are likely more alike than those taken on different trees
 - If we ignore these inter-relationships, estimates of the mean will still be unbiased, BUT we artificially inflate our DOF and deflate the standard errors → we are pretending to have more information than we actually have!

Mixed models for multiple measurements per experimental unit

- Knowledge of these correlations can be used to formulate the correct experimental error in our models
- Moreover, this knowledge can be useful in better understanding our data!

Experimental design Slide 45

Mixed models for multiple measurements per experimental unit (e.g., fixed area plots)

- E.g., models of biomass as a function of $k=3$ site qualities, where we measure $m=4$ trees in each of $n=15$ plots (60 trees total)

Source	Degrees of freedom	F test
Site	$k-1=2$	$F_{(2,42)} = MS_S / MS_E$
Experimental Error	$k(n-1) = 42$	
Within plot error	$nk(m-1)=135$	
Total	$knm-1 = 179$	

In the case of $k=3$ sites with $m=4$ trees per $n=15$ plots, each plot is a "subject"

Experimental design Slide 46

Mixed models for multiple measurements per experimental unit (e.g., repeated measures)

- E.g., models of biomass at $k=3$ sites on $n=15$ trees at $m=4$ times

Source	Degrees of freedom	F test
Site	$k-1=2$	MS_S / MS_E
Experimental Error	$k(n-1) = 42$	
time	$m-1=3$	MS_T / MS_W
Site x time	$(k-1)(m-1)=6$	$MS_{S \times T} / MS_W$
Within tree error	$k(n-1)(m-1)=126$	
Total	$knm-1 = 179$	

In the case of $k=4$ sites and $m=4$ measurements per $n=15$ trees, each tree is a "subject"

Experimental design Slide 47

Mixed models for multiple measurements per experimental unit (e.g., repeated measures)

- The most important aspect of the mixed model is the formulation of the F tests
- The site effect in the model are tested against the Experimental Error, whereas time is tested against the within-tree error
- This ensures that we appropriately account for within subject correlations

In the case of $k=4$ sites and $m=4$ measurements per $n=15$ trees, each tree is a "subject"

But this assumes our times are independent. But it is likely that we have correlations among times within tree...

Repeated times model Source	Degrees of freedom	F test
Site	$k-1=2$	MS_S/MS_E
Experimental Error	$k(n-1) = 42$	
time	$m-1=3$	MS_t/MS_w
Site x time	$(k-1)(m-1)=6$	MS_{St}/MS_w
Within tree error	$k(n-1)(m-1)=126$	
Total	$knm-1 = 179$	

Slide 48

How to formulate the appropriate model?

- The observations are "clustered" within a "subject" (e.g., plot for fixed area example, tree for repeated measures example)
 - the observations, and their residuals, are not independent, but correlated.
- There are two ways to deal with this correlation
 - A Marginal or Population Averaged approach.
 - A Mixed Model

Slide 49

The Marginal (Population Averaged) approach

- Instead of modeling correlation among residuals, the covariance structure of the residuals is modeled
 - While in linear models, observations are assumed independent, in marginal models, residuals from a single subject are assumed related.
 - Covariances among subjects are assumed non-zero
 - covariances among residuals from each subject are estimated
- not truly a mixed model, although you can use mixed methods to estimate them.
- (In SAS or SPSS, you use a repeated statement instead of a random statement)

Slide 50

The Mixed Model approach

- The model is altered by controlling for subject as a factor in the model
- Residuals are re-defined as the distance between the observed value and the *mean value for that subject*
- Subjects are not fixed effects in the model but instead are treated as a *random effect*
 - This uses less degrees of freedom

Fixed versus random effects

- FIXED effects**
 - An effect is fixed if all possible levels about which inferences will be made are represented
 - A level of a fixed effect is an unknown constant, which does not vary
 - If we were to repeat the study, we would choose the same factor levels
 - Examples
 - Regression models are fixed effects models, as X is assumed fixed
 - Most effects that we purposely study are considered fixed
- RANDOM effects**
 - Effects are random if the levels represent only a random sample of possible levels
 - Sub-sampling, clustering, and random selection of treatments result in random effects in models
 - If we were to repeat the study, a different set of effect levels would be obtained

Slide 52

How to fit a mixed model with subsamples?

Recall: biomass as a function of $k=3$ site qualities, where we measure $m=4$ trees in each of $n=15$ plots (60 trees total)

```
> library(nlme)
> data.sq$plot <- as.factor(data.sq$plot)
> lme.sq <- lme(biomass ~ quality, random = ~1|plot, data=data.sq)
> anova(lme.sq)
> summary(lme.sq)
> plot(lme.sq)
```

- The function lme estimates a linear mixed effects model (Y~X) using data in the dataframe data.sq
- A random effect is added to account for grouping of trees within plots
 - ~1|plot fits a model with a random intercept for each plot
- The functions summary, anova, plot are used in the same manner as with the simpler model

NOTE: in order for this to work properly in R, you *must* have unique plot numbers, e.g., you cannot have a plot 1 in each site quality!!

53

R output: mixed model with subsamples

```
> anova(lm.sq)
              numDF  denDF  F-value p-value
(Intercept)      1    135 29.516138 <.0001
quality          2     42  4.722407  0.0152
```

Note the difference in denDF

```
> summary(lme.sq)
Linear mixed-effects model fit by REML
Data: data.sq
      AIC      BIC    logLik
344.7039 342.842 -166.3519
```

Random effects:

```
Formula: ~1 | plot
(Intercept) Residual
StdDev:    1.582772 4.060305
```

Estimates of the variance among plots versus within plots

Fixed effects: biomass ~ quality

```
Value Std.Error DF  t-value p-value
(Intercept)  1.212249  1.264953 135  0.9583354  0.3437
qualityB     3.316992  1.788913  42  1.8541942  0.0822
qualityC    -0.029265  1.788913  42 -0.0163590  0.9872
```

These tests are for the effect versus the base (A)

54

R output: mixed model with

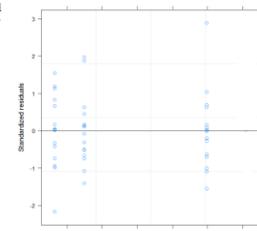
```
Correlation:
              (Intr)  quality.B
qualityB     -0.707
qualityC     -0.707    0.500
```

This is not the correlation between the variables. It is the expected correlation of the model coefficients. This might indicate multicollinearity; it indicates that if you did the experiment again and the coefficient for A got smaller, it is likely that those of B and C would get larger

```
Standardized Within-Group Residuals:
      Min       Q1       Med
-1.8651688 -0.6058632 -0.0108787
```

Number of Observations: 60
Number of Groups: 12

```
> plot(lme.sq)
```



This graph should have no pattern

55

How to fit a mixed model with repeated times?

Recall: biomass at $k=3$ sites on $n=15$ trees at $m=4$ times

```
> library(nlme)
> data.rm$time <- as.factor(data.rm$time)
> lme.rm <- lme(biomass ~ site*time, random = ~1|tree,
data=data.rm)
> anova(lme.rm)
> summary(lme.rm)
> plot(lme.rm)
```

- The function `lme` estimates a linear mixed effects model (Y~X) using data in the dataframe `data.rm`
- `site*time = site + time + site:time`
- A random effect is added to account for grouping of measurements on the same tree
- The functions `summary`, `anova`, `plot` are used in the same manner as with the simpler model

56

R output: mixed model with repeated times

```
> anova(lm.rm)
              numDF  denDF  F-value p-value
(Intercept)      1    126 92.46865 <.0001
site            2     42  3.59848  0.0189
time           3    126 35.55504 <.0001
site:time       6    126  0.50806  0.8673
```

```
> summary(lm.rm)
Linear mixed-effects model fit by REML
Data: data.rm
      AIC      BIC    logLik
1172.093 1233.503 -428.0465
```

Random effects:

```
Formula: ~1 | tree
(Intercept) Residual
StdDev:    3.448301 2.019734
```

Fixed effects: biomass ~ site * time

```
Value Std.Error DF  t-value p-value
(Intercept)  0.733993  1.0318303 126  0.711351  0.4779
siteB       1.017526  1.4592283  42  0.697304  0.4885
siteC       3.325862  1.4592283  42  2.290368  0.0094
time2       2.275000  0.7375026 126  3.084843  0.0024
time3       2.629211  0.7375026 126  3.425019  0.0005
time4       2.667666  0.7375026 126  3.617162  0.0004
siteB:time2  0.375345  1.0429862 126  0.359876  0.7194 ...
```

57

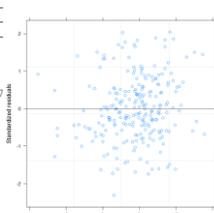
R output: mixed model with repeated times - 2

```
Correlation:
              (Intr)  siteB  siteC  time2  time3  time4  sitB:2  sitC:2  ...
siteB         -0.623
siteC         -0.623  0.200
time2         -0.357  0.253  0.253
time3         -0.357  0.253  0.253  0.200
time4         -0.357  0.253  0.253  0.200  0.200
siteB:time2   0.253 -0.357 -0.179 -0.623 -0.354 -0.354
siteC:time2   0.253 -0.179 -0.357 -0.623 -
siteB:time3   0.253 -0.357 -0.179 -0.354 -
siteC:time3   0.253 -0.179 -0.357 -0.354 -
...
```

```
Standardized Within-Group Residuals:
      Min       Q1       Med
-2.31861551 -0.58095354 -0.05834473  0.53737
```

Number of Observations: 180
Number of Groups: 12

```
> plot(lm.rm)
```



58

Are random intercepts enough?

Random intercepts model

- Intercepts are allowed to vary
- biomass is predicted by an intercept that varies across trees
- assumes that slopes are fixed (the same pattern across time)
- information about intra-subject correlations are helpful in determining whether there is correlation among measurements on the same subject

Random slopes model

- Slopes are allowed to vary
- slopes are different across trees
- assumes that intercepts are fixed

Random intercepts and slopes model

- includes both random intercepts and random slopes
- most complex
- both intercepts and slopes are allowed to vary across trees, meaning that they are different across times

59

How to fit a mixed model with random slope and intercept

Recall: biomass at k=3 sites on n=15 trees at m=4 times

```
> lme.rm <- lme(biomass ~ site*time, random = time|tree, data=data.rm)
> anova(lme.rm)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	126	42.96620	<.0001
site	2	42	3.44695	0.0226
time	3	126	44.31872	<.0001
site:time	6	126	0.63711	0.7642

With only random intercept:

```
(Intercept) 1 126 92.46
site         2  42  3.59
time        3 126 35.55
site:time   6  126  0.50
```

Does the AIC indicate a better model?

```
> summary(lme.rm)
Linear mixed-effects model fit by REML
Data: data.a.rm
      AIC      BIC    logLik
1174.107 1266.222 -560.0537
...
      Value Std.Error DF t-value p-value
(Intercept) 0.733993 0.9333719 126 0.786389 0.4327
siteB       1.017526 1.3199872  42 0.770860 0.4440
siteC       3.925862 1.3199872  42 2.974166 0.0043
time2       2.275080 0.8674875 126 2.622608 0.0095
time3       2.629211 0.6337064 126 4.148941 0.0001
time4       2.667666 0.6950092 126 3.838318 0.0002
```

The estimates are the same, but the standard errors are very different!

60

What correlation pattern do we expect among observations on the same subject?

- The models we fit assumed a compound symmetric correlation structure (CS) among measurements taken on the same subject (trees in the same plots or times on the same tree)
- What if we think measurements taken closer together in time/space might be more correlated than those taken farther apart?

General form of a variance-covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{21}^2 & \dots & \sigma_{t1}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \dots & \sigma_{t2}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{t1}^2 & \sigma_{t2}^2 & \dots & \sigma_{tt}^2 \end{bmatrix}$$

Diagonal elements are the variances among observations from different subjects taken at the same time
 Off-diagonal elements are the co-variances between observations taken at different times

Variance components – type matrix (VC)

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

In a fixed effect model, we assume:

- variances among observations from different subjects taken at the same time (diagonal elements) are equal (**homoscedastic!**)
- co-variances between observations taken at different times (off-diagonal elements) are zero (**independent!**)

Compound Symmetric (CS) Variance-covariance matrix

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix}$$

- Variances among observations from different subjects taken at the same time (diagonal elements) are the equal (**homoscedastic!**)
- Co-variances between observations taken at different times (off-diagonal elements) are equal

→ Regardless of time between measurements, observations from same subject are equally correlated

Autoregressive order 1 (AR(1)) variance-covariance structure

- Variances among obs from different subjects taken at the same time (diag. elements) are the equal (homoscedastic!)
- Covariances between obs taken at different times (off-diag. elements) are correlated, with constant decay ρ

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}$$

Correlations decrease as time between obs. increases

Add to lme call: `corr=covAR1()`

What if our experimental/sampling area is not homogenous?

- **Blocking**
 - A block is a group of homogeneous experimental units
 - Blocks are chosen so as to maximize variation among blocks with the aim of minimizing the variation within blocks
- **Reasons for blocking**
 - To remove block-to-block variation from the experimental error (which should increase precision)
 - To allow more uniform treatment comparisons
 - To allow the researcher to sample a wider range of conditions

Blocking set-up example #1: 4 treatments and 3 blocks

- At least one replication is grouped in a homogeneous area

A	B	D	A
C	D	B	C
B	A	D	C

**Just replication
(i.e., CRD)**

A	B	D	C
C	D	B	A
B	A	D	C

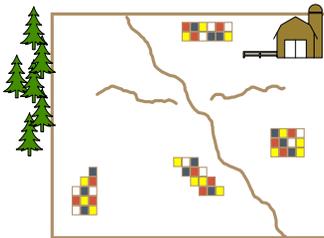
**Blocked
replication**

Block #1

Block #2

Block #3

Blocking set-up example #2: 4 treatments, 3 blocks, and 3 replicates



Criteria for blocking

- Proximity or known patterns of variation in the field
 - gradients due to fertility, soil type
 - animals (experimental units) in a pen (block), or plants in a greenhouse
- Time
 - season, time of planting / harvesting
- Management of experimental tasks / Control of observer error
 - individuals collecting data
 - runs in the laboratory
- Physical characteristics
 - Height class, maturity level
- Natural groupings
 - branches (experimental units) on a tree (block)

Advantages of the Randomized Complete Block (RCB) Design

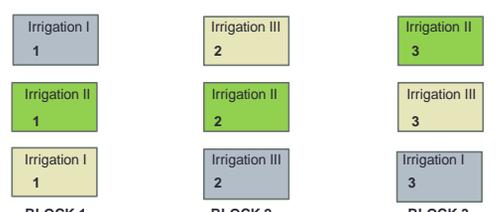
- Can remove site variation from experimental error and thus increase precision
- When an operation cannot be completed on all plots at one time, can be used to remove variation between "runs"
- By placing blocks under different conditions, it can broaden the scope of the trial
- Can accommodate any number of treatments and any number of blocks, but each treatment should be replicated the same number of times in each block
- Statistical analysis is fairly simple

Disadvantages of the RCB

- Missing data can cause some difficulty in the analysis
- If there is more than one source of unwanted variation, the design can be less efficient
- If the plots are uniform, then RCB is less efficient than CRD
- As treatment or entry numbers increase, more heterogeneous area is introduced and effective blocking becomes more difficult (a Split plot or lattice designs may be more appropriate)
- Blocks are not a fixed effect, and therefore inference about particular blocks is inappropriate (better to use a factorial analysis)

Randomized complete block designs (RCB)

- Blocks are chosen so that the experimental material within block is homogeneous – and generally we do NOT care to make inferences about blocks (it is a 'nuisance' variable)
- Treatments are randomly assigned *within block* (restricted randomization)



Randomized complete block designs – ANOVA table

- We would analyze as a two-way ANOVA – also a GLM

Source	Degrees of freedom	Mean Squares	F test
Block	n-1=2	MS _B	$F_{(2,4)} = \frac{MS_{IRR}}{MS_E}$
Irrigation	k-1=2	MS _{IRR}	
Experimental Error	(k-1)(n-1) = 4	MS _E	
Total	kn-1 = 8		

Experimental error is partitioned so that we separate out block-to-block variation → lose DOF but (hopefully) decrease Exp.Error

73

How to fit a mixed model with blocking?

```

> data.rcb$block <- as.factor(data.rcb$block)
> lme.rcb <- lme(biomass ~ irr, random = ~1|block/irr, data=data.rcb)
> anova(lme.rcb)
> summary(lme.rcb)
> plot(lme.rcb)

```

- The function lme estimates a linear mixed effects model (Y~X) using data in the dataframe data.rcb
- Block is not a fixed effect
- Irrigation types are nested inside each block in the random effect
- The functions summary, anova, plot are used in the same manner as with the other analyses

Split plot (SP) design

- Developed by agronomists to save money on experimental layout where there are two or more factors involved, and where it is much easier to apply one treatment to a large area, e.g.:
 - we want to test effect of burning/not burning in combination with two harvesting methods (it is difficult to burn just a small area), or
 - we want to test two fertilizers and three irrigation regimes (its much easier to control irrigation over a large area)
- Restriction on randomization is in two parts: first the main plot is randomized, then sub-plot is randomized within it
- Can be used with any basic designs (CRD, RCB, etc.)

Split plot design layout over RCB – irrigation and fertilization example

- E.g., we sample understory biomass in 0.01 ha plots under three irrigation and two fertilization regimes
- We have 3 blocks we can use
- We wish to study three levels of irrigation and two levels of fertilizer
 - We could do a factorial (fertilizer x irrigation) over an RCB....
 - BUT It is very difficult to irrigate a small area
- We decide to first allocate areas of each block to the three irrigation regimes, then overlay fertilizer treatments over each irrigation plot

Split plot design layout over RCB – irrigation and fertilization example

- First, set out irrigation treatments in main plots

Block 1	i ₁	i ₂	i ₃
Block 2	i ₁	i ₃	i ₂
Block 3	i ₃	i ₁	i ₂

- Then overlay fertilizer treatments over each irrigation plot

Block 1	f ₁	f ₂	f ₂	f ₁	f ₁	f ₂
Block 2	f ₂	f ₁	f ₁	f ₂	f ₂	f ₁
Block 3	f ₂	f ₁	f ₁	f ₂	f ₁	f ₂

- The final layout is then:

Block 1	i ₁ f ₁	i ₁ f ₂	i ₂ f ₂	i ₂ f ₁	i ₃ f ₁	i ₃ f ₂
Block 2	i ₁ f ₂	i ₁ f ₁	i ₃ f ₁	i ₃ f ₂	i ₂ f ₂	i ₂ f ₁
Block 3	i ₃ f ₂	i ₃ f ₁	i ₁ f ₁	i ₁ f ₂	i ₂ f ₁	i ₂ f ₂

Issues with Split plot Designs

- There are two different types or sizes of experimental units
 - Main treatment effects are estimated from the whole plots
 - Sub-plot and interaction of whole and sub-plot effects are estimated from sub-plots
- *the main and sub-plots have different precision*
- Observations from different subplots in the same whole plot may be correlated
 - The correlation between any two subplots on the same whole plot is equal across plots
 - The correlation between observations in different whole plots is zero
- *the covariance matrix of observations within a whole plot is compound symmetric*

Issues with Split plot Designs – cont'd

- *the main and sub-plots have different precision*
- *the covariance matrix of observations within a whole plot is compound symmetric*

Implications:

- The partitioning of the sums of squares is altered
 - The experimental error is split in to two parts: those for the underlying design's main treatment effects, and those for the sub-plots
 - The main treatment SS and tests remain unchanged from the underlying design

Analysis of Variance Table (for Split-Plot CRD)

Source	DoF	SS	MS
Treatment (Main)	$k-1$	SS_{TR}	MS_{TR}
Exp. Err. #1	$k(n-1)$	SS_{E1}	MS_{E1}
Subunit	$m-1$	SS_S	MS_S
Treat. x Subunit	$(k-1)(m-1)$	$SS_{T \times S}$	$MS_{T \times S}$
Exp. Err. #2	$k(n-1)(m-1)$	SS_{E2}	MS_{E2}
Total	$nkm-1$	SS_T	

How do we construct F tests?

If we have a 'fixed effects' model, then main treatments are tested against EE#1 and subunits are tested against EE#2

This looks just like our repeated measures example!

Conclusions: Why does design matter?

- Experimental designs have HUGE impacts on how we collect and analyze the data
 - How we set up the experiment controls:
 - What effects are testable
 - What error terms are appropriate
 - The number of 'true replicates'
- Controls are meant to allow us to eliminate as many artifacts as possible introduced by our experimental procedure, e.g.:
 - In drug studies, placebo group gets a sugar pill/saline shot in order to simulate the same stress as those would undergo while taking the real drug
 - Animals that are handled often undergo stress, so those that are 'controls' should get handled as well

Take home messages

- In the design stage, be sure to be *very clear* about how you intend to collect the data!
 - Draw a picture
 - Make a table
 - Consider 'confounding factors', such as aquaria or greenhouse space or other things that might introduce bias
- Using well-studies designs enables us to easily analyze data and construct uncertainty estimates