

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

Stratified random sample

Effects parameterization

Diagram illustrating the effects parameterization for a stratified random sample. The model is defined by the following equations:

$$y_i \sim \text{Normal}(\mu_i, \sigma^2)$$
$$\mu_i = \beta_0 + \beta_1 x_{1,i}$$

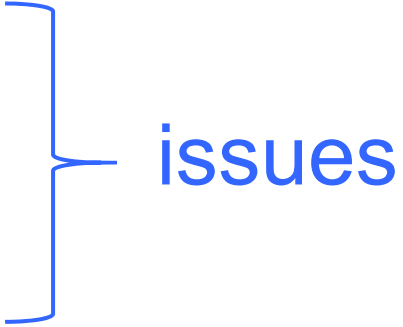
Annotations:

- μ_i is labeled "or whatever" (blue arrow pointing to μ_i).
- β_0 is labeled "species 1 (reference level)" (blue arrow pointing to β_0).
- $\beta_1 x_{1,i}$ is labeled "species 2" (blue arrow pointing to $\beta_1 x_{1,i}$).

A red curved arrow points from the μ_i in the second equation to the μ_i in the first equation.

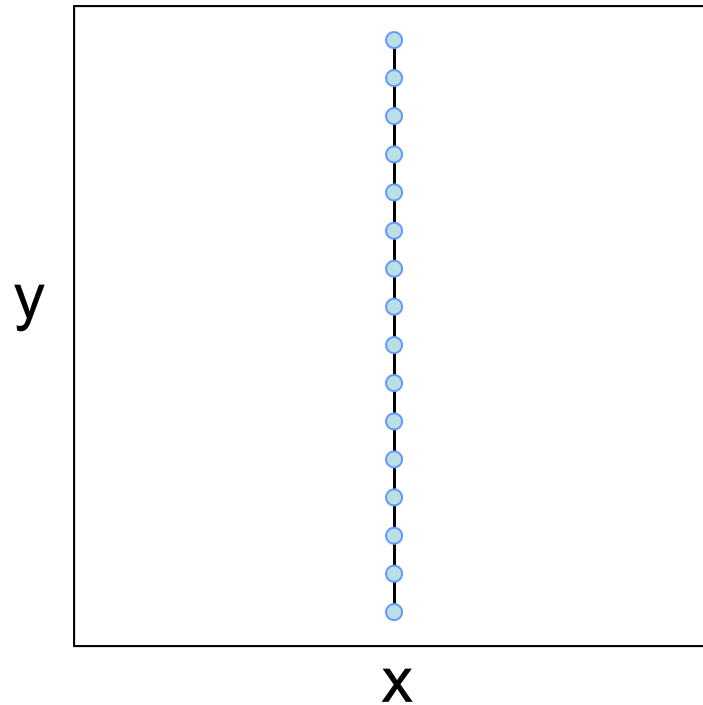
R code: `stan_lmer(y ~ species)`

Systematic sampling

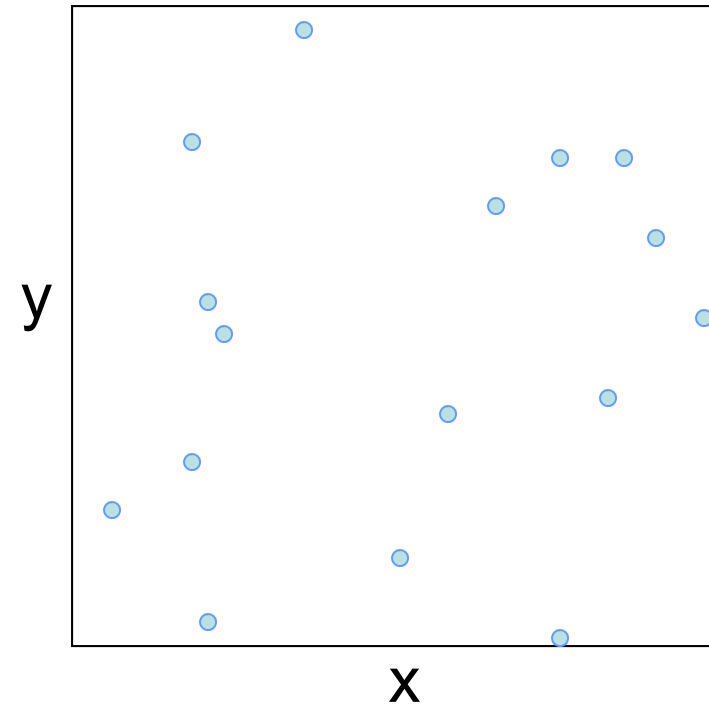
- Opposite of random
 - Examples
 - transects with equal spacing of samples
 - spatial grid
 - every Thursday
 - Bias
 - Autocorrelation
 - Scope
- 
- issues

Example:
spatial
sample

Transect



Simple random sample



Bias:

Autocorrelation:

Scope:

one x; gradient on y?

strong, systematic

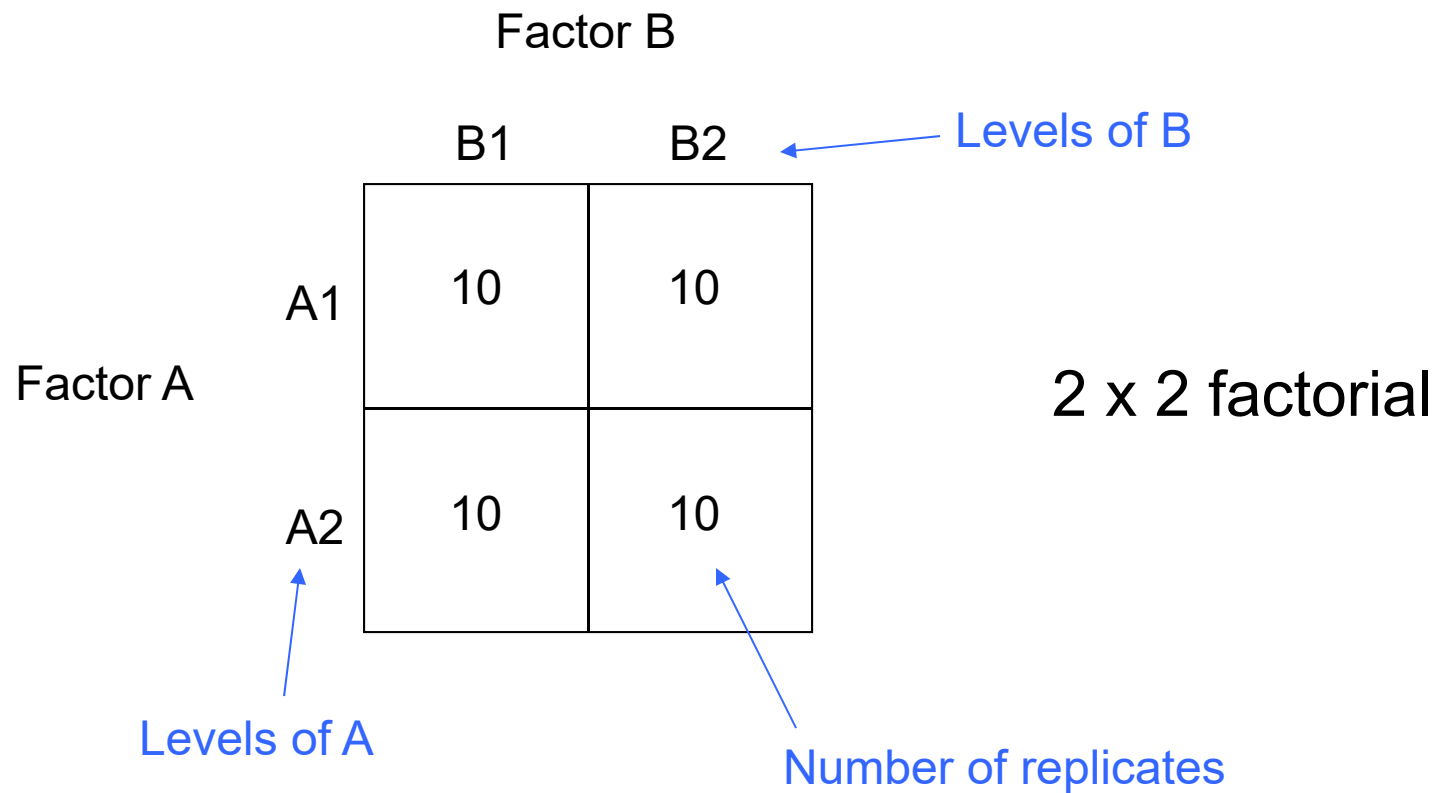
this transect

none

weak, diffuse

population

Factorial design



Advantage: estimate interaction effect

Factorial design

Effects parameterization

2 factors (A, B)

or whatever

$$y_i \sim \text{Normal}(\mu_i, \sigma^2)$$

Plot-scale stochastic model

$$\mu_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{1,i} x_{2,i}$$

factor A1, B1
(reference or control level)

factor B2

factor A2

interaction
A2 B2

R code: `stan_glm(y ~ factor_A * factor_B)`

Factorial design

- Many possibilities
 - $2 \times 2 \times 2 = \text{cube}$
 - $2 \times 2 \times 2 \times 2$
 - 3×2
 - 5×4
 - ...

Factorial versus response surface design

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

50 experimental units
no interaction

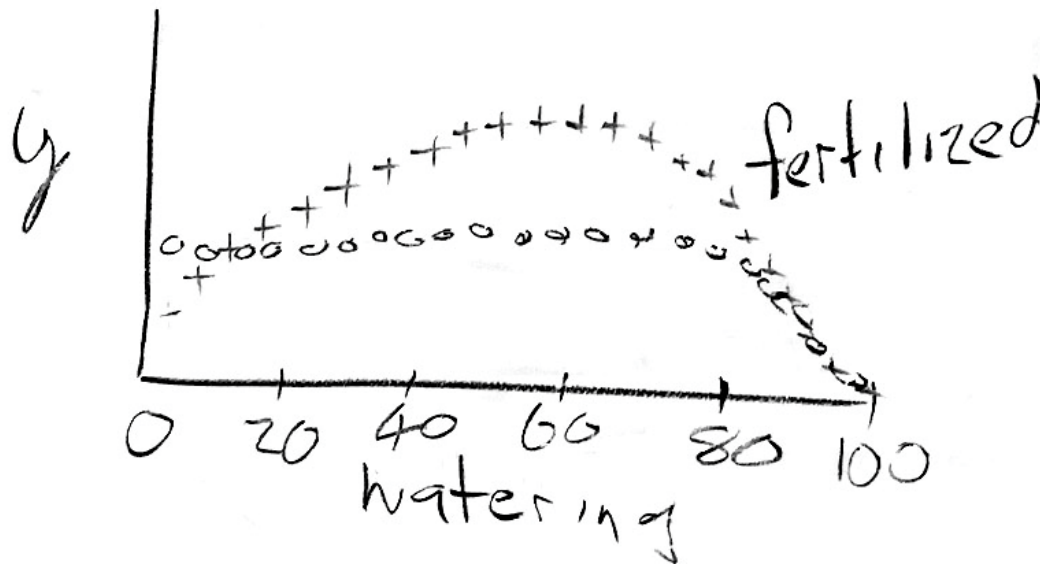
parameters = 7

df = 50 - 7 = 43

with interaction

parameters = 11

df = 50 - 11 = 39



50 experimental units

3 parameters per curve

df = 50 - 7 = 43

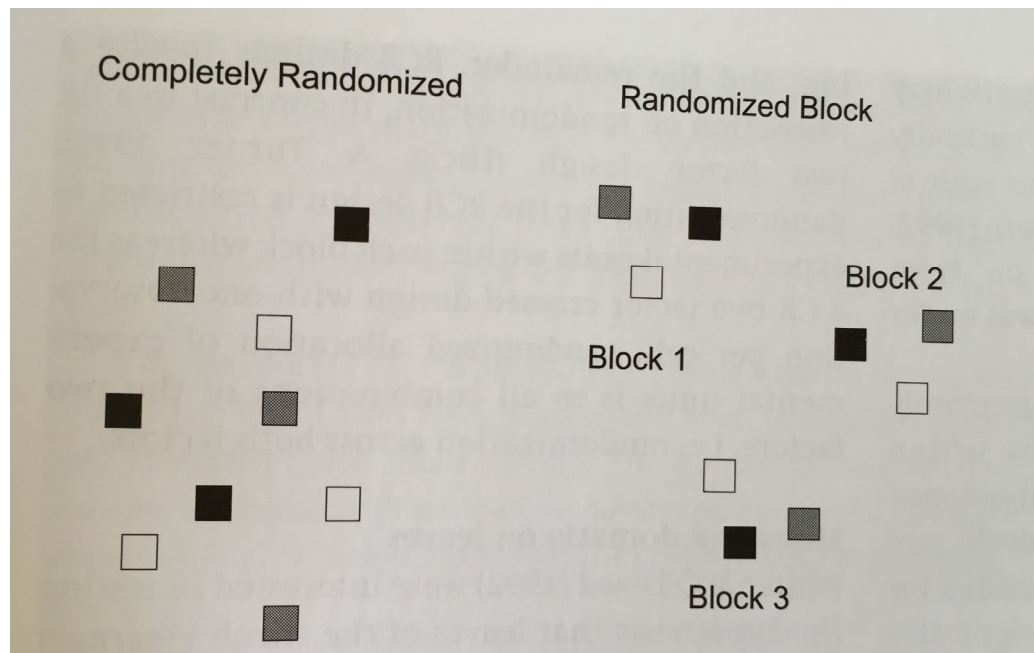
5 parameters per curve

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

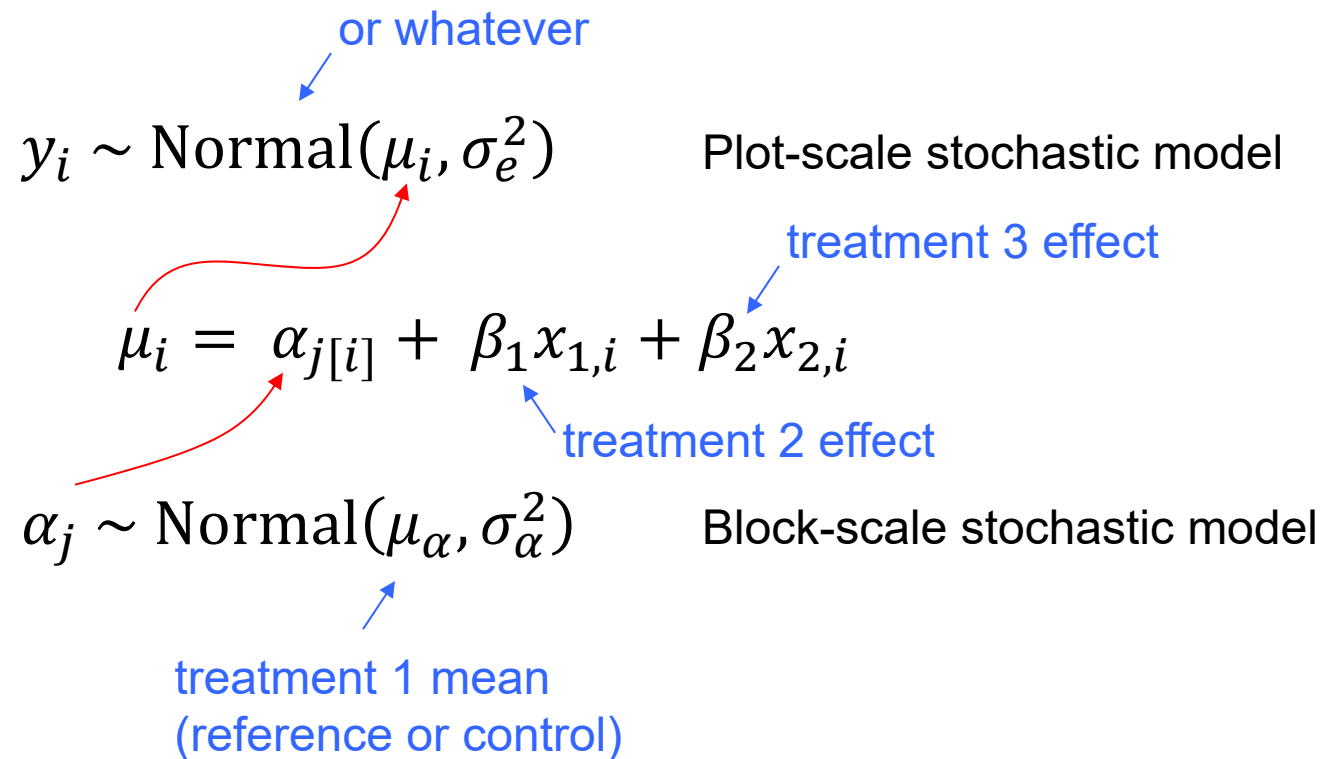
Cons: penalty for more complex model (grouping variable)

Whether it helps depends on this **tradeoff**

Randomized block

Effects parameterization

3 treatments



R code: `stan_lmer(y ~ treatment + (1|block))`

Additive decomposition form

Effects parameterization

observed value for plot i

treatment 2 effect

deviation of plot i from mean (aka residual deviation)

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + b_{j[i]} + e_i$$

treatment 1 mean (reference or control)

treatment 3 effect

deviation of block j from mean (aka "random effects")

$$b_j \sim \text{Normal}(0, \sigma_b^2)$$

variance among blocks

Block-scale stochastic model

$$e_i \sim \text{Normal}(0, \sigma_e^2)$$

variance among plots within blocks

Plot-scale stochastic model

R code: `stan_lmer(y ~ treatment + (1|block))`

Additive decomposition form

Means parameterization

observed value for plot i

treatment 2 mean

deviation of plot i from mean (aka residual deviation)

$$y_i = \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + b_{j[i]} + e_i$$

treatment 1 mean (reference or control)

treatment 3 mean

deviation of block j from mean (aka "random effects")

$$b_j \sim \text{Normal}(0, \sigma_b^2)$$

variance among blocks

Block-scale stochastic model

$$e_i \sim \text{Normal}(0, \sigma_e^2)$$

variance among plots within blocks

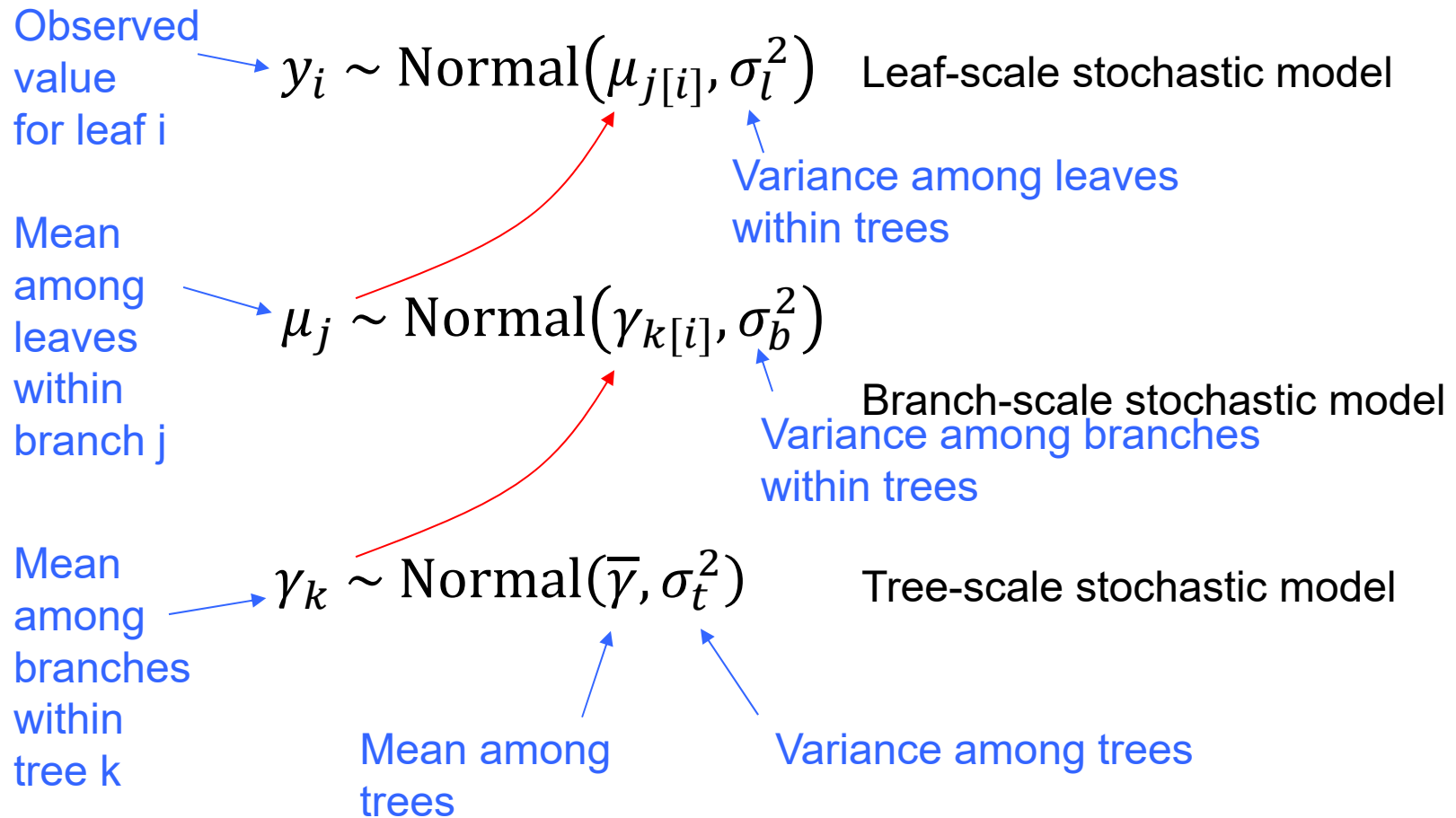
Plot-scale stochastic model

R code: `stan_lmer(y ~ -1 + treatment + (1|block))`

Multilevel designs

- 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

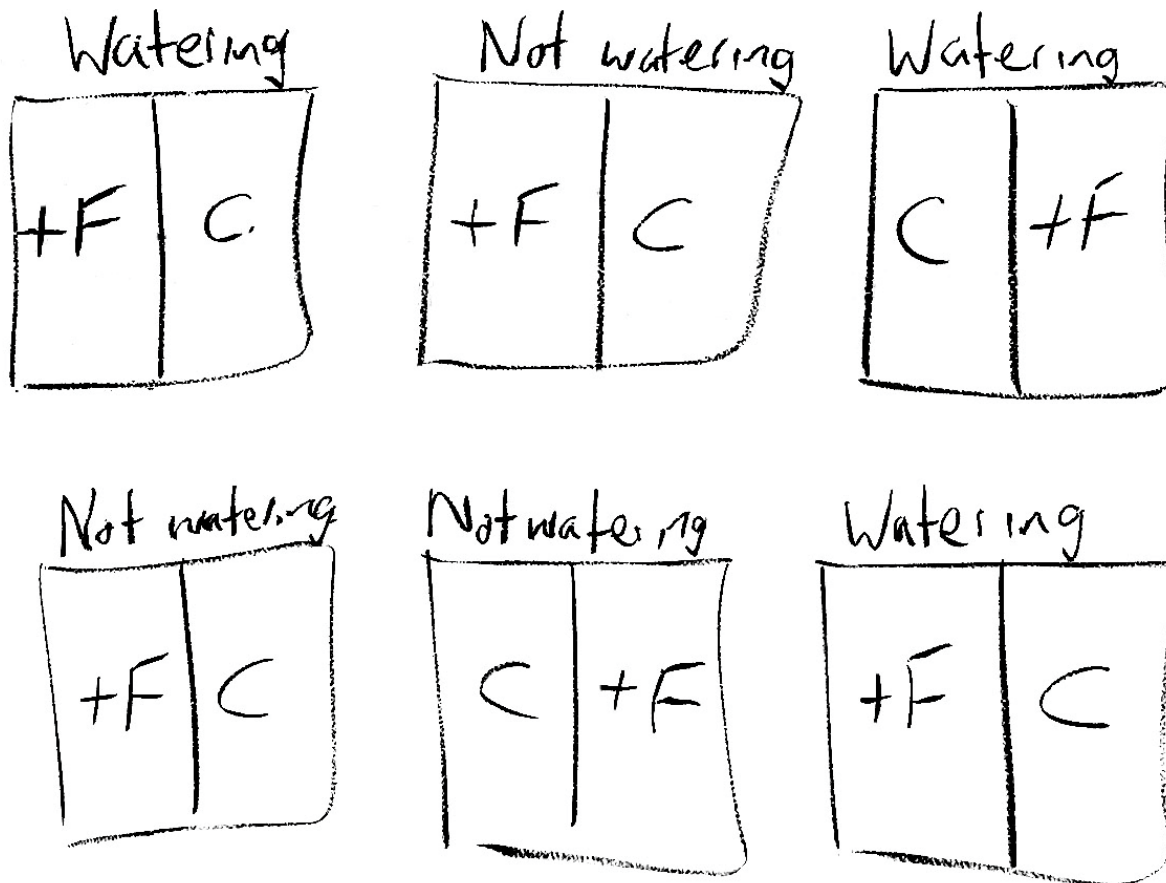
Nested random sample



R code: `stan_lmer(y ~ (1|tree) + (1|branch))`
`stan_lmer(y ~ (1|tree/branch))`

Multilevel designs

- Split plot experiment



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)

Split plot

Effects parameterization

Treatments at 2 scales

$$y_i \sim \text{Normal}(\mu_i, \sigma_e^2)$$

Sub-plot-scale stochastic model

$$\mu_i = \alpha_{j[i]} + \beta_1 x_{1,i} + \beta_3 x_{1,i} x_{2,j[i]}$$

interaction

fertilizer

$$\alpha_j \sim \text{Normal}(\mu_\alpha, \sigma_\alpha^2)$$

Plot-scale stochastic model

$$\mu_\alpha = \beta_0 + \beta_2 x_{2,j}$$

control

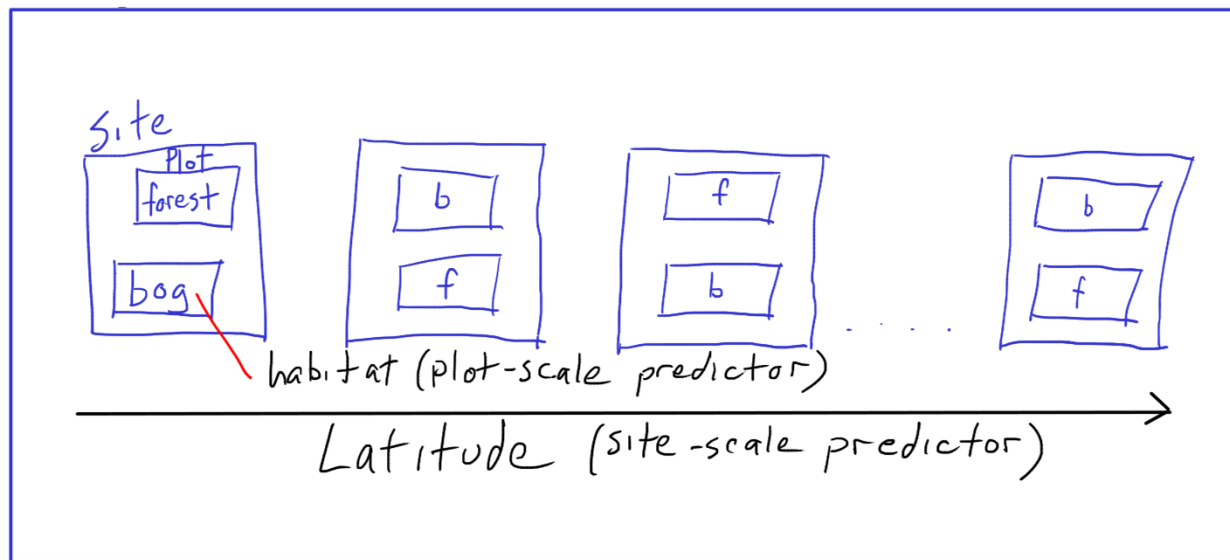
watering

(no fertilizer or water)

R code: `stan_lmer(y ~ watering * fertilizer + (1|plot))`

Multilevel designs

- Split plot (ants sampling)



Sites (aka plots) are split into plots (aka sub-plots).

Latitude is at large scale (site), habitat is at small scale (plot).

Pro: travel simpler, control large scale var

Con: replication of large scale factor is reduced (22)

Con: penalty for model complexity (need a grouping variable)

Split plot – ants GLM

Effects parameterization

Predictors at 2 scales

$$y_i \sim \text{Poisson}(\mu_i)$$

Plot-scale stochastic model

$$\ln(\mu_i) = \alpha_{j[i]} + \beta_1 x_{1,i} + \beta_3 x_{1,i} x_{2,j[i]} + e_i$$

interaction

forest (habitat)

overdispersion

$$\alpha_j \sim \text{Normal}(\mu_\alpha, \sigma_\alpha^2)$$

Site-scale stochastic model

$$\mu_\alpha = \beta_0 + \beta_2 x_{2,j}$$

bog (intercept)

latitude

$$e_i \sim \text{Normal}(0, \sigma_e^2)$$

R code: `stan_lmer(y ~ habitat * latitude + (1|site/unit))`