Introduction to Bayesian Statistics

Part 6
Linear Mixed Models

Benjamin Rosenbaum

iDiv 2025

This lecture

- Random grouping factors

- Random effects

- Fixed & random effects
  - Random intercepts
  - Random intercepts and slopes

- Multiple random grouping factors

- Continuous correlation structures

- Summary

Assumptions for linear models

1. ~~Independent observations.~~

   ~~Systematic differences in y are because of x !~~
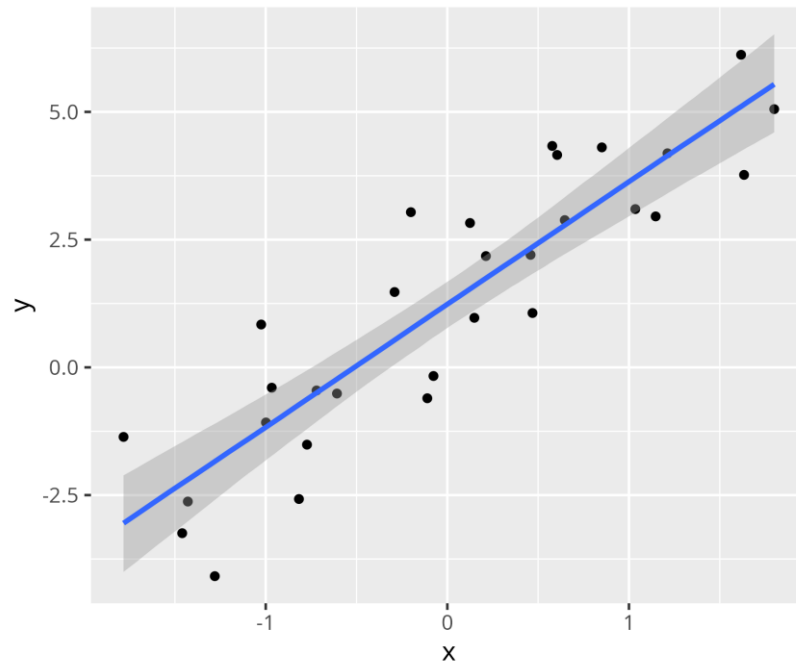
2. Trend of y follows (linear) prediction model

   $\mu(x) = a + b \cdot x$

3. Residuals follow normal distribution

   $\varepsilon \sim \text{Normal}(0, \sigma)$

4. Constant variance (standard deviation σ)

   across whole range of x

Assumptions for linear models

When are residuals not independent ?

Study design: N observations have been **sampled in groups**
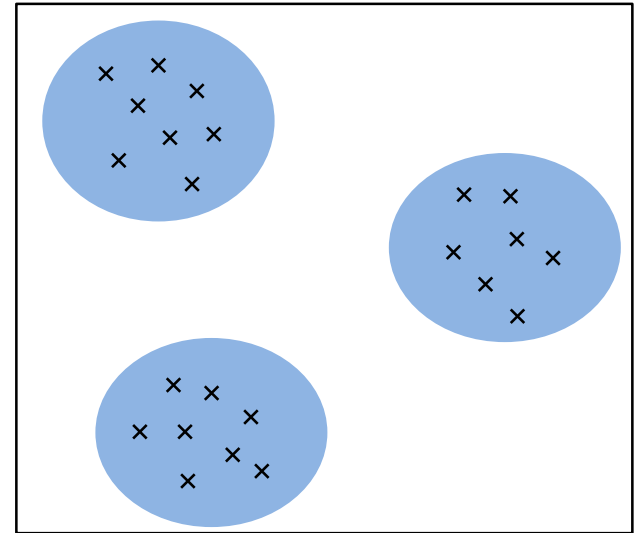
→ There is a categorical predictor with M levels $(M < N)$

→ Each level / group contains multiple observations,
   these are no independent replicates !

→ Maybe structural (or random) differences
   between (unobserved) conditions of these groups

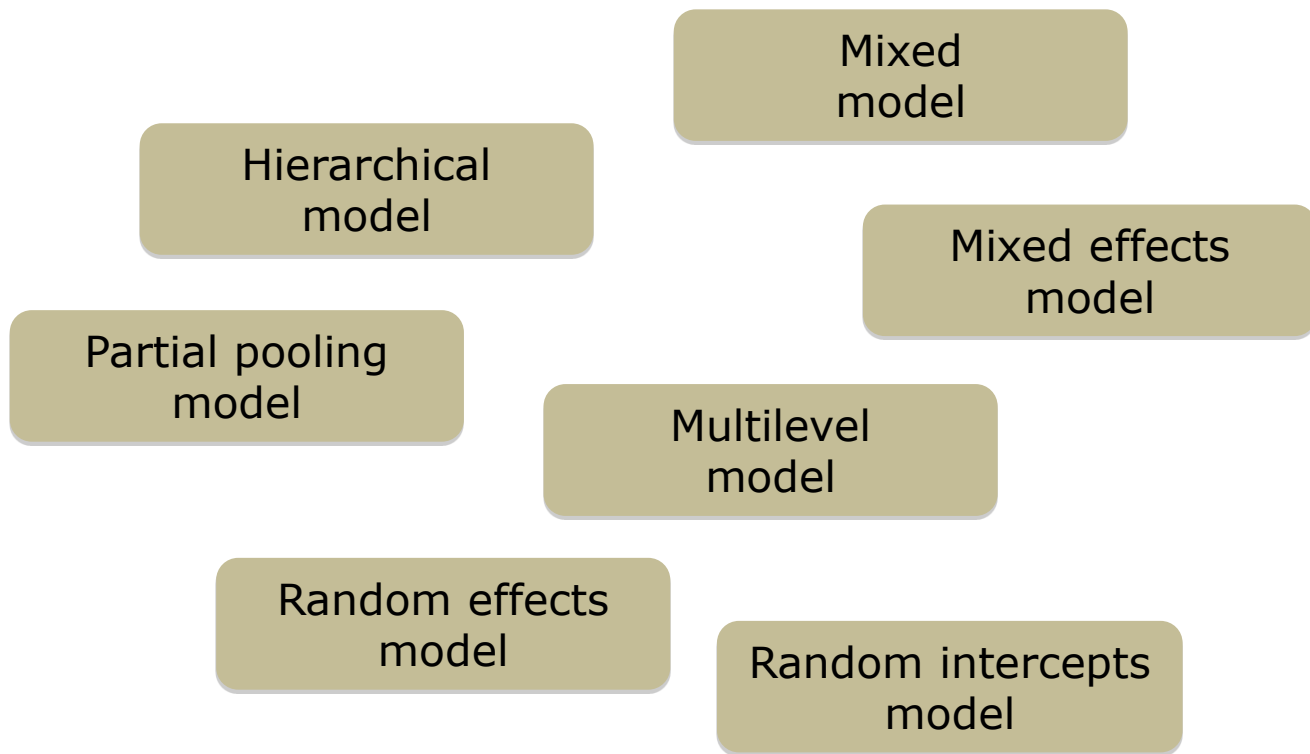→ Still want to make **inference on whole population**
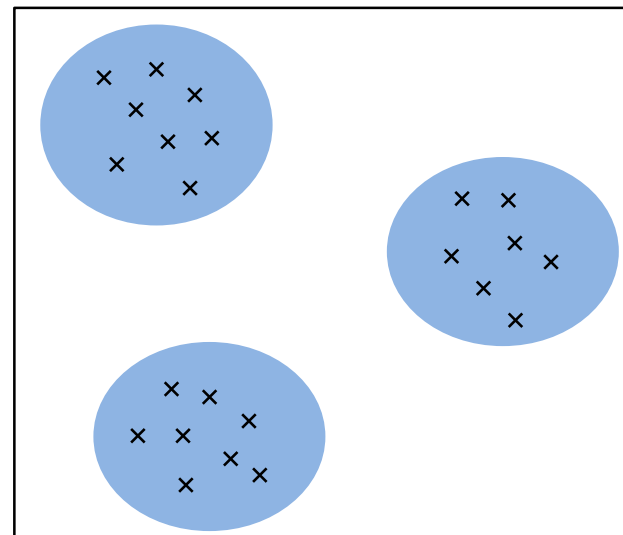
Random grouping factors

Examples:

Site: Data collected at multiple locations

Year: Data collected ~same time in different years

Observer: Data collected by different resarchers

Individual: Multiple measurements on the same test object

Unit: Multiple experimental units like chambers / tanks

Species: Data on higher level (e.g. family) contains
                 multiple species

Word salad !!

Mixed model

Hierarchical model

Mixed effects model

Partial pooling model

Multilevel model

Random effects model

Random intercepts model

Random effects

Example

- Measured weight of $N = 200$ viper snakes

- Data collected on $M = 9$ different sites across France
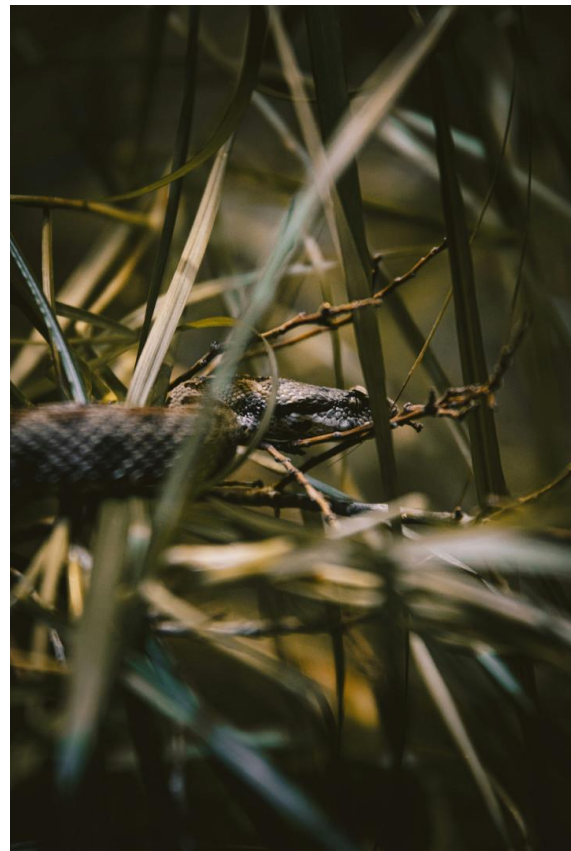
- Also measured length, but we're not using it for now

Q: What is the correct statistical model for species mean weight ?

`brm( weight ~ 1 )` **?**

→ No, observations not independent

`brm( weight ~ site )` **?**

→ No, does not estimate overall weight (just site-specific means)

(1) Complete pooling

Model: $y_i = a + \varepsilon_i$ $\qquad i = 1 \dots N$

$\qquad\quad \varepsilon_i \sim \text{Normal}(0, \sigma)$

Priors: $a \sim \text{Normal}(150, 10)$

- Ignore categorical predictor $site$

- Information is completely pooled across all levels

- Fit joint intercept a only
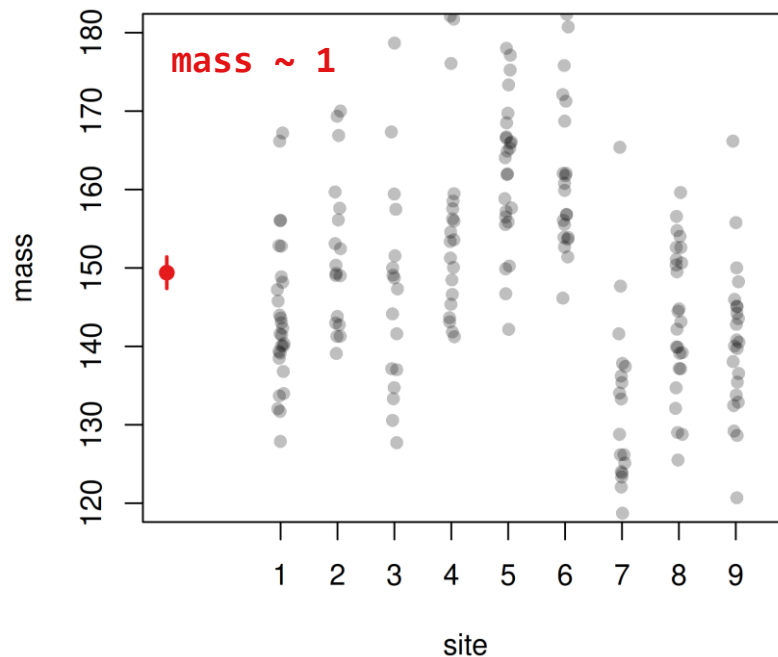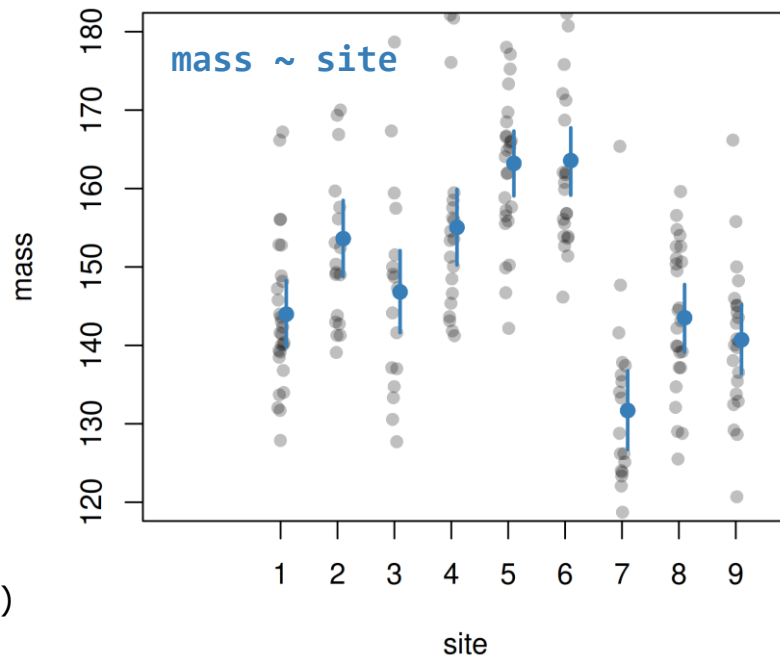
- a is given a prior

(2) No pooling

Model: $y_i = a_{site(i)} + \varepsilon_i$ $i = 1 \dots N$

$\varepsilon_i \sim \text{Normal}(0, \sigma)$

Priors: $a_j \sim \text{Normal}(150,10)$ $j = 1 \dots M$

- Include categorical predictor $site$ **(ANOVA)**

- No information is shared across levels

- Fit independent intercepts $a_1 \dots a_M$

- Each a_j is given a prior

  (here they are identical, but could also be different)

(3) Partial pooling

Model:
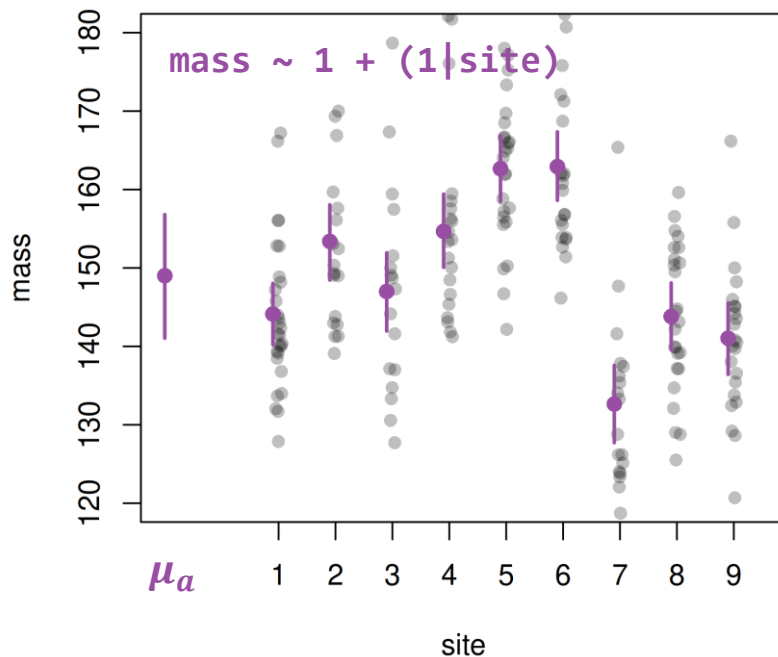$$y_i = a_{site(i)} + \varepsilon_i \qquad i = 1 \ldots N$$

$$\varepsilon_i \sim \text{Normal}(0, \sigma)$$

$$a_j \sim \text{Normal}(\mu_a, \sigma_a) \qquad j = 1 \ldots M$$

Priors:
$$\mu_a \sim \text{Normal}(150, 10)$$

$$\sigma_a \sim \text{Exponential}(1/10)$$

- Replace prior distributions by a joint distribution of the site-specific intercepts $a_1 \ldots a_M$

- Their mean μ_a and sdev σ_a are parameters and are estimated with all others by MCMC

- These "hyperparameters" μ_a, σ_a are given priors

11

(3) Partial pooling

Model:　　$y_i = \boldsymbol{\mu_a} + \boldsymbol{\delta_{site(i)}} + \varepsilon_i$　　　$i = 1 \dots N$

　　　　　$\varepsilon_i \sim \text{Normal}(0, \sigma)$

　　　　　$\boldsymbol{\delta_j \sim \text{Normal}(0, \sigma_a)}$　　　$\boldsymbol{j = 1 \dots M}$

Priors:　　$\boldsymbol{\mu_a} \sim \text{Normal}(150,10)$

　　　　　$\boldsymbol{\sigma_a} \sim \text{Exponential}(1/10)$

- Replaced $a_j = \mu_a + \delta_j$

- Effects　$\delta_j = a_j - \mu_a$　describe each site's
  deviation from overall mean μ_a

- Identical model, just rearrangement of coefficients

12
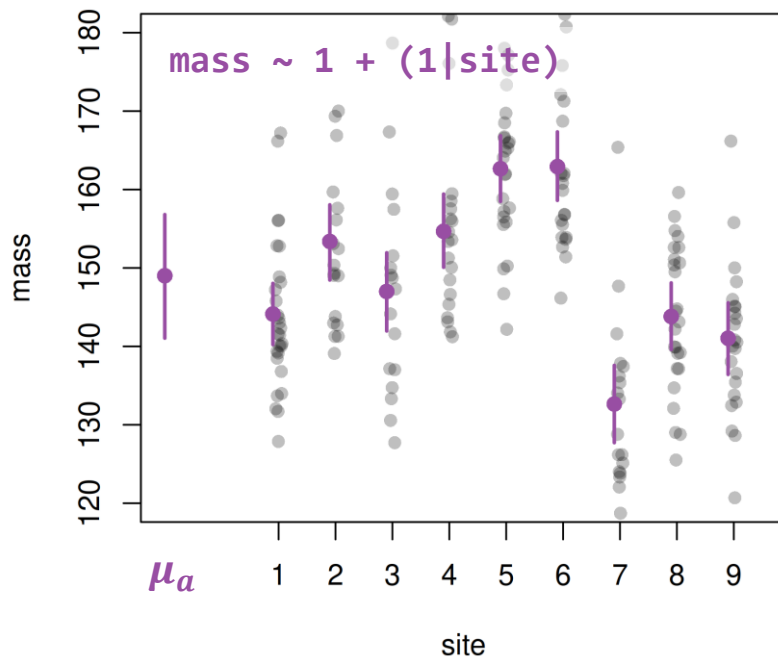
Comparison

vs. complete pooling `mass ~ 1`

Uncertainty of μ_a in **mass~1** is smaller.
Wrongly assumes $N = 200$ independent obs.

Uncertainty of μ_a in **mass~1+(1|site)** correct.
Informed by $M = 9$ sites.

vs. no pooling `mass ~ site`

Site-specific means a_j of **mass~1+(1|site)**
closer to overall mean μ_a than in **mass~site**

„Shrinkage" of parameters
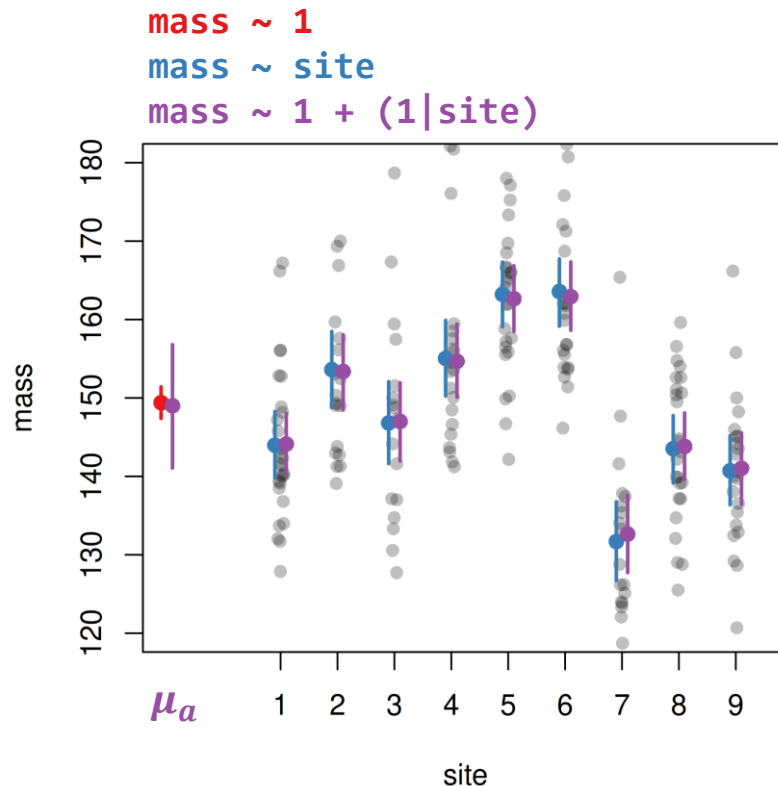

```
mass ~ 1
mass ~ site
mass ~ 1 + (1|site)
```

Random effects

Random effects $\delta_j = a_j - \mu_a$ describe each site's

deviation from overall mean μ_a with sdev σ_a

Assumption: normally distributed

(Bayesian 3d printer: could also be other distribution,

 e.g. positive lognormal distribution for a_j)

Hyperparameters μ_a, σ_a informed by all 9 sites

→ Some information is shared / pooled

→ Borrowing strength

→ For unbalanced designs this improves prediction

  in levels with few observations.
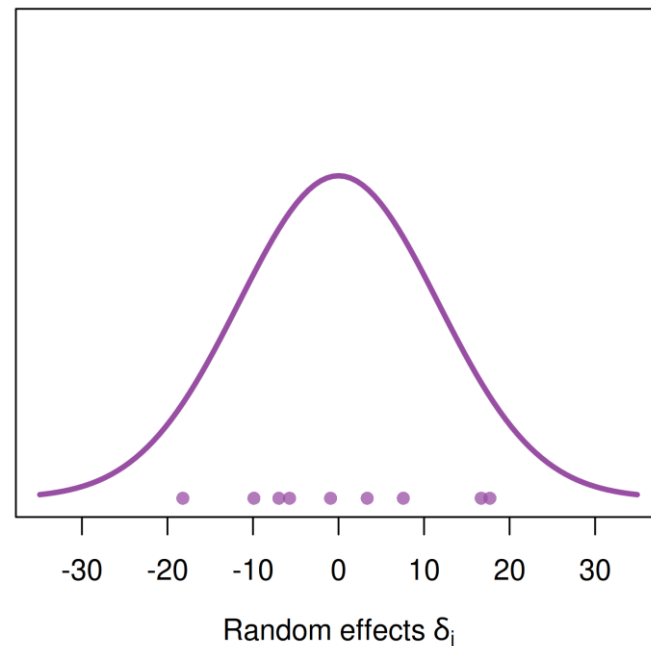
Random effects δ_j

Random effects

Random effects $\delta_j = a_j - \mu_a$ describe each site's deviation from overall mean μ_a with sdev σ_a

This part of the likelihood $\delta_j \sim \text{Normal}(0, \sigma_a)$

would be maximized for $\sigma_a \to 0$ and all $\delta_j \to 0$

But the residuals' likelihood $y_i \sim \text{Normal}(a_{site(i)}, \sigma)$

would decrease if all δ_j were close to 0

→ Partial pooling model is a compromise
  between small random effects (on site-level)
  and good model fit (on observation level)
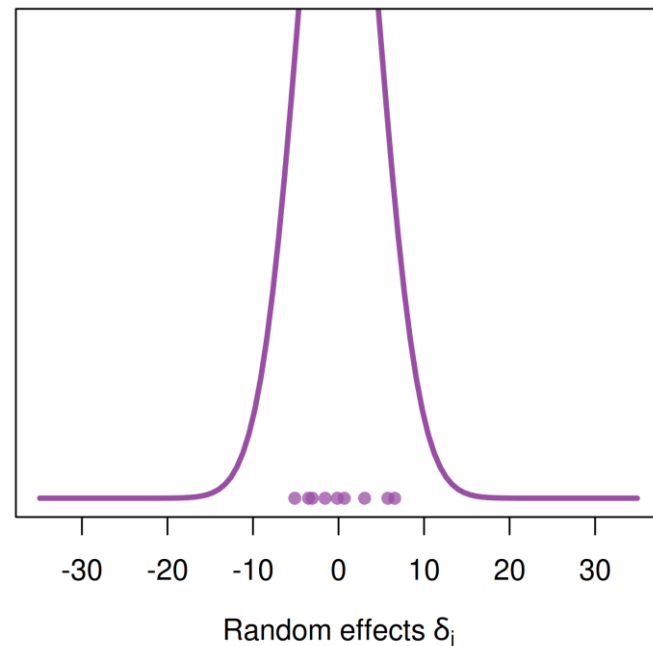
Random effects δ_j

Variance partitioning

Overall variance of the data is estimated on different levels
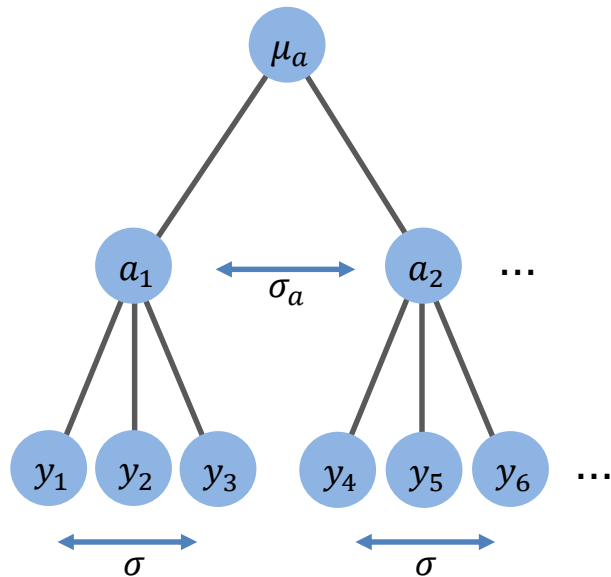
$$s^2_{\text{total}} = s^2_{\text{among}} + s^2_{\text{within}}$$

Site-level (among) σ_a

Explains site-specific deviation from overall mean μ_a

Observation-level (within) σ

Explains each data point's deviation from site-mean a_j

(All the randomness in the data which is not caused by

 site-level differences)

```
> brm(mass ~ 1 + (1|site), …)
```

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: mass ~ 1 + (1 | site)
   Data: df (Number of observations: 200)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000

Multilevel Hyperparameters:
~site (Number of levels: 9)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)    11.64      3.26     6.99    19.70 1.01      814     1353

Regression Coefficients:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept   149.01      3.91   141.05   156.81 1.00      777     1038

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma    11.02      0.57     9.97    12.20 1.00     2527     2566
```

$\sigma_a \rightarrow$ sd(Intercept)

$\mu_a \rightarrow$ Intercept

$\sigma \rightarrow$ sigma

Fixed and random effects
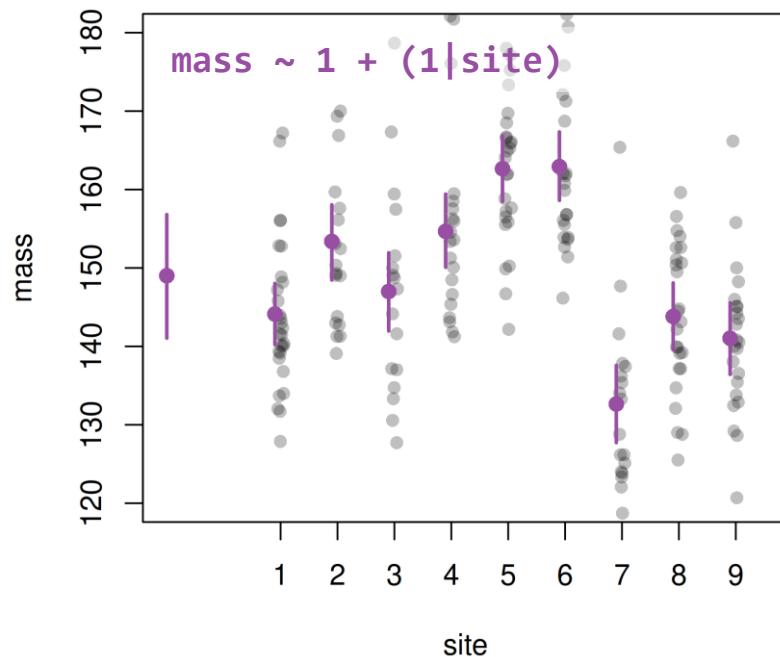
Fixed & random effects

In previous example, we were interested in a correct statistical model for the overall intercept

→ „Fixed effect" μ_a

Variation among sites was not our focus, but we wanted to account for non-independence of observations

→ „Random effect" δ_j $(j = 1 \dots M)$

Fitted an intercept-only model with random intercepts (site)

mass ~ 1 + (1|site)

Example

- Measured weight of $N = 200$ viper snakes

- Data collected on $M = 9$ different sites across France

- **Also measured length (continuous predictor)**
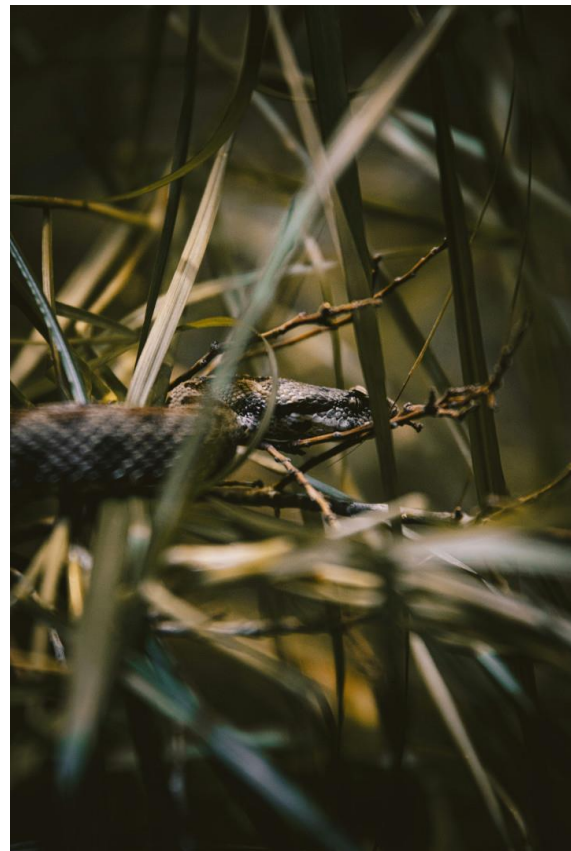
Q: Correct statistical regression model for weight vs. length?

`brm( weight ~ length )` **?**

→ No, observations not independent

`brm( weight ~ length*site )` **?**

→ No, does not estimate overall slope (just site-specific regressions)

(1) Complete pooling
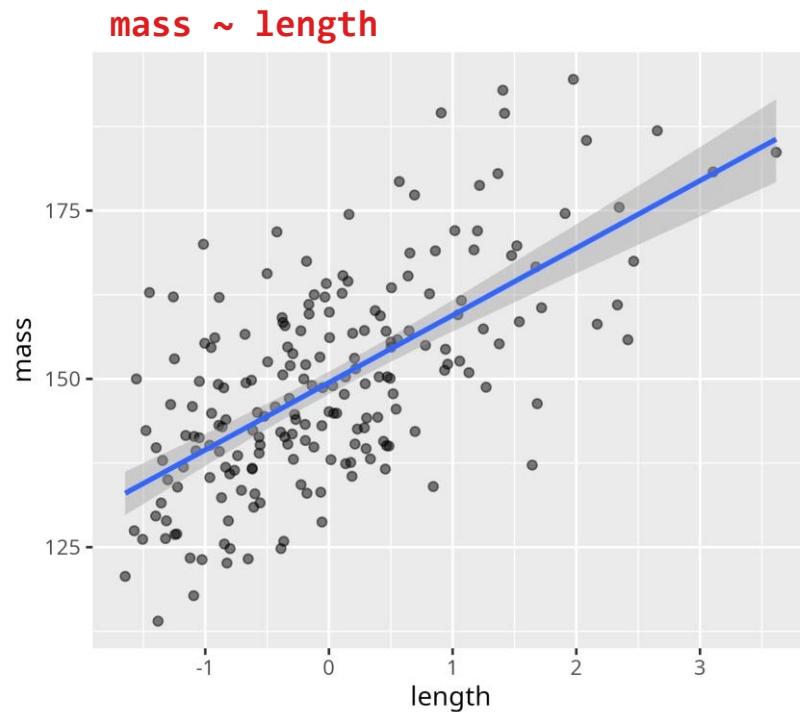
Model: $y_i = a + b \cdot length_i + \varepsilon_i \qquad i = 1 \ldots N$

$\varepsilon_i \sim \text{Normal}(0, \sigma)$

Priors: $a \sim \text{Normal}(150, 10)$

$b \sim \text{Normal}(10, 5)$

- Ignore categorical predictor $site$

- Information is completely pooled across all levels

- Fit joint intercept a & slope b only

- a, b are given a prior

mass ~ length

21

(2) No pooling

mass ~ length*site

Model: $y_i = a_{site(i)} + b_{site(i)} \cdot length_i + \varepsilon_i$

$\varepsilon_i \sim \text{Normal}(0, \sigma)$

Priors: $a_j \sim \text{Normal}(150,10)$ $\qquad j = 1 \dots M$

$b_j \sim \text{Normal}(10,5)$

- Include categorical predictor $site$ **(ANCOVA)**

- No information is shared across levels

- Fit independent intercepts $a_1 \dots a_M$ & slopes $b_1 \dots b_M$

- Each a_j & b_j is given a prior

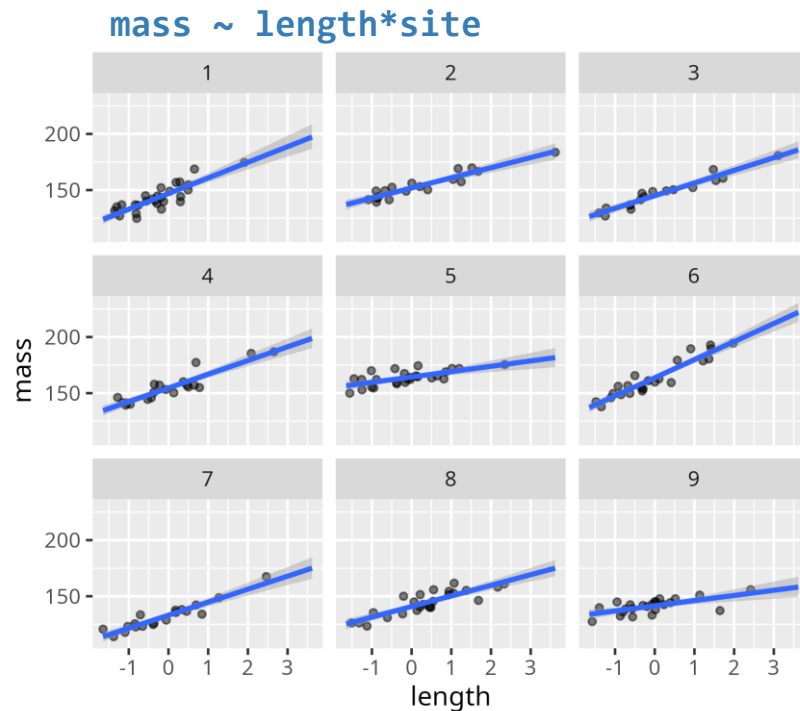 (here they are identical, but could also be different)

22

(2) No pooling

Model:
$$y_i = a_{site(i)} + b_{site(i)} \cdot length_i + \varepsilon_i$$

$$\varepsilon_i \sim \text{Normal}(0, \sigma)$$

Priors:
$$a_j \sim \text{Normal}(150,10) \qquad j = 1 \dots M$$

$$b_j \sim \text{Normal}(10,5)$$

- Include categorical predictor $site$ **(ANCOVA)**

- No information is shared across levels

- Fit independent intercepts $a_1 \dots a_M$ & slopes $b_1 \dots b_M$

- Each a_j & b_j is given a prior

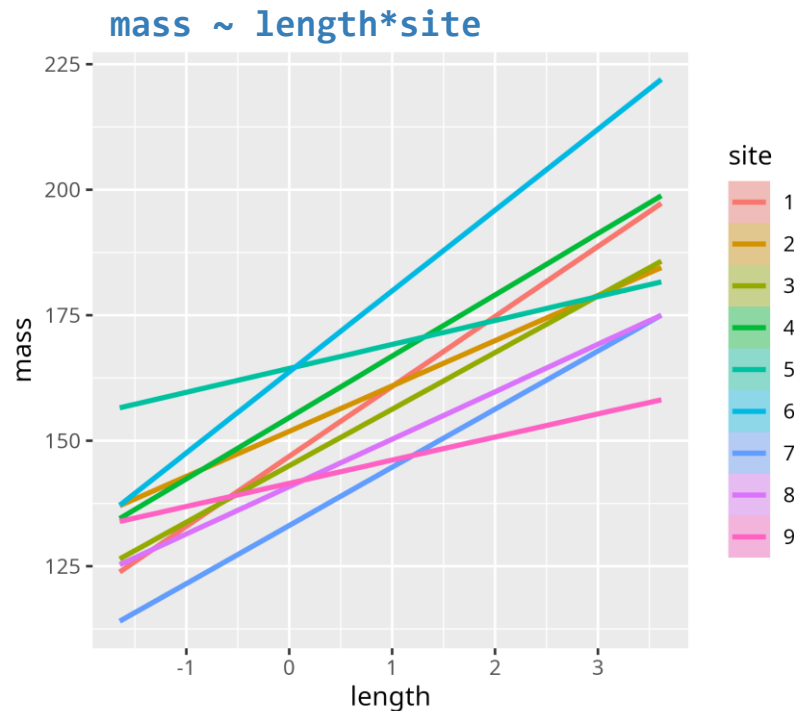 (here they are identical, but could also be different)

mass ~ length*site

23

(3) Partial pooling: Random intercepts

Model:
$$y_i = a_{site(i)} + b \cdot length_i + \varepsilon_i$$

$$\varepsilon_i \sim \text{Normal}(0, \sigma)$$

$$a_j \sim \text{Normal}(\mu_a, \sigma_a) \qquad j = 1 \dots M$$

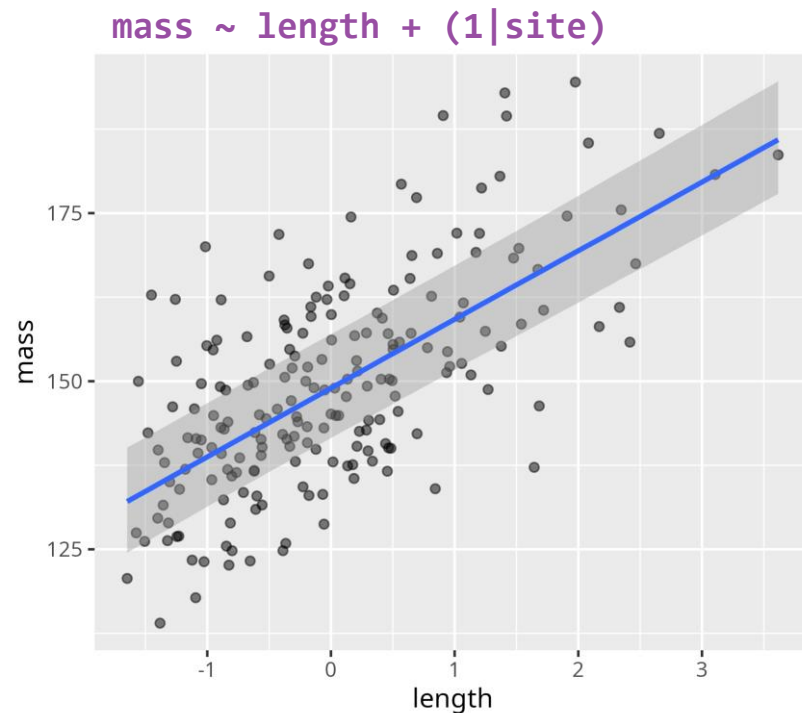Priors:
$$b_j \sim \text{Normal}(10,5)$$

$$\mu_a \sim \text{Normal}(150,10)$$

$$\sigma_a \sim \text{Exponential}(1/10)$$

Random effects: $a_1 \dots a_M$ site-specific intercepts

Fixed effects: μ_a joint intercept

   b joint slope

`mass ~ length + (1|site)`

Marginal predictions: $\mu_a + b \cdot length$

(3) Partial pooling: Random intercepts

Model: $y_i = (\boldsymbol{\mu_a} + \boldsymbol{\delta_{site(i)}}) + b \cdot length_i + \varepsilon_i$

$\varepsilon_i \sim \text{Normal}(0, \sigma)$

$\boldsymbol{\delta_j \sim \text{Normal}(0, \sigma_a)} \qquad j = 1 \dots M$

Priors: $b_j \sim \text{Normal}(10,5)$

$\boldsymbol{\mu_a} \sim \text{Normal}(150,10)$

$\boldsymbol{\sigma_a} \sim \text{Exponential}(1/10)$

Random effects: $\delta_1 \dots \delta_M$ intercepts deviation

Fixed effects: μ_a joint intercept

b joint slope

mass ~ length + (1|site)

Marginal predictions: $\mu_a + b \cdot length$

(3) Partial pooling: Random intercepts

Model:
$$y_i = (\mu_a + \delta_{site(i)}) + b \cdot length_i + \varepsilon_i$$

$$\varepsilon_i \sim \text{Normal}(0, \sigma)$$

$$\delta_j \sim \text{Normal}(0, \sigma_a) \qquad j = 1 \dots M$$
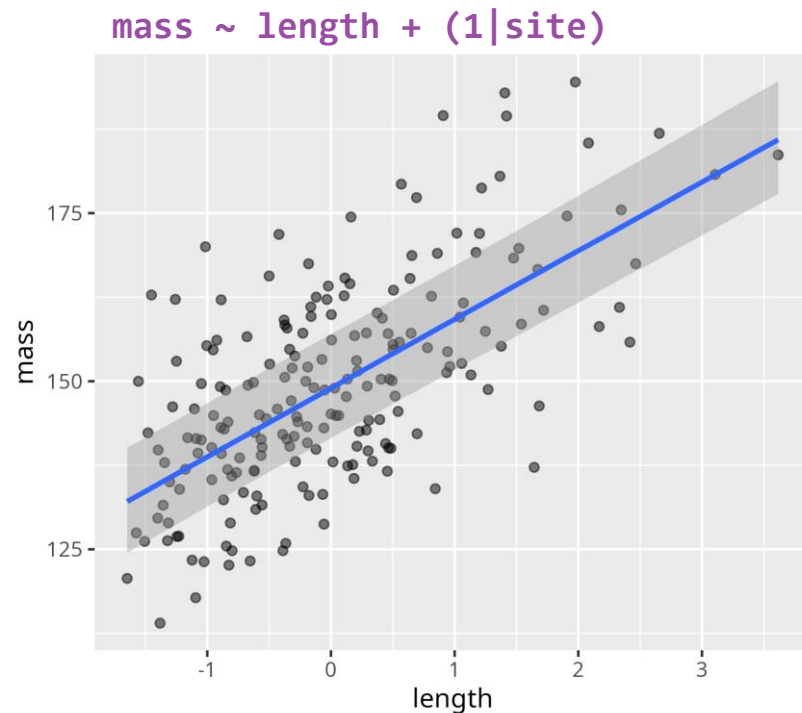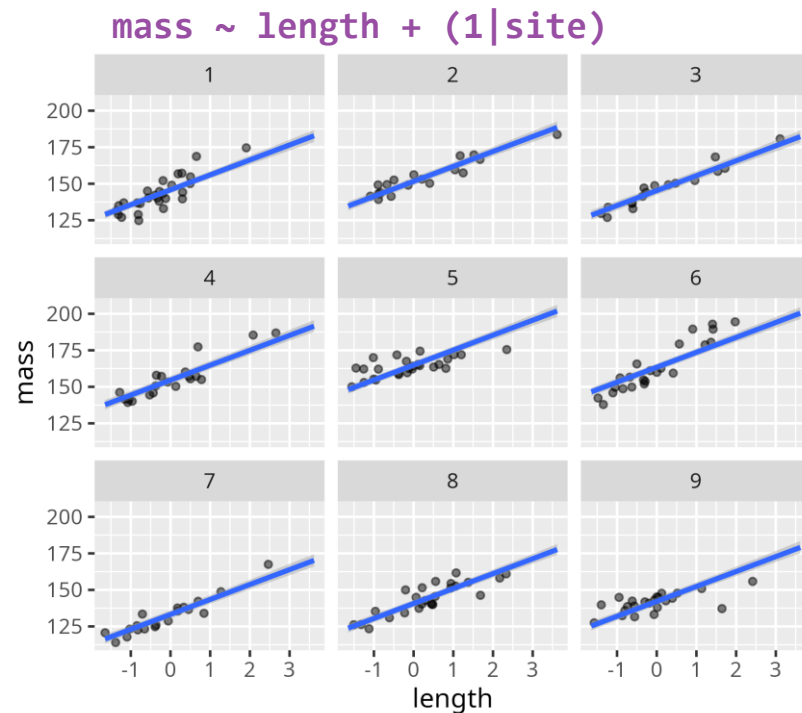
Priors:
$$b_j \sim \text{Normal}(10,5)$$

$$\mu_a \sim \text{Normal}(150,10)$$

$$\sigma_a \sim \text{Exponential}(1/10)$$

Random effects: $\delta_1 \dots \delta_M$ intercepts deviation

Fixed effects: μ_a joint intercept

b joint slope

mass ~ length + (1|site)

Conditional predictions: $a_j + b \cdot length$

26

(3) Partial pooling: Random intercepts

Model: $y_i = (\boldsymbol{\mu_a} + \boldsymbol{\delta_{site(i)}}) + b \cdot length_i + \varepsilon_i$

$\varepsilon_i \sim \text{Normal}(0, \sigma)$

$\boldsymbol{\delta_j} \sim \textbf{Normal}(\textbf{0}, \boldsymbol{\sigma_a}) \qquad j = 1 \dots M$

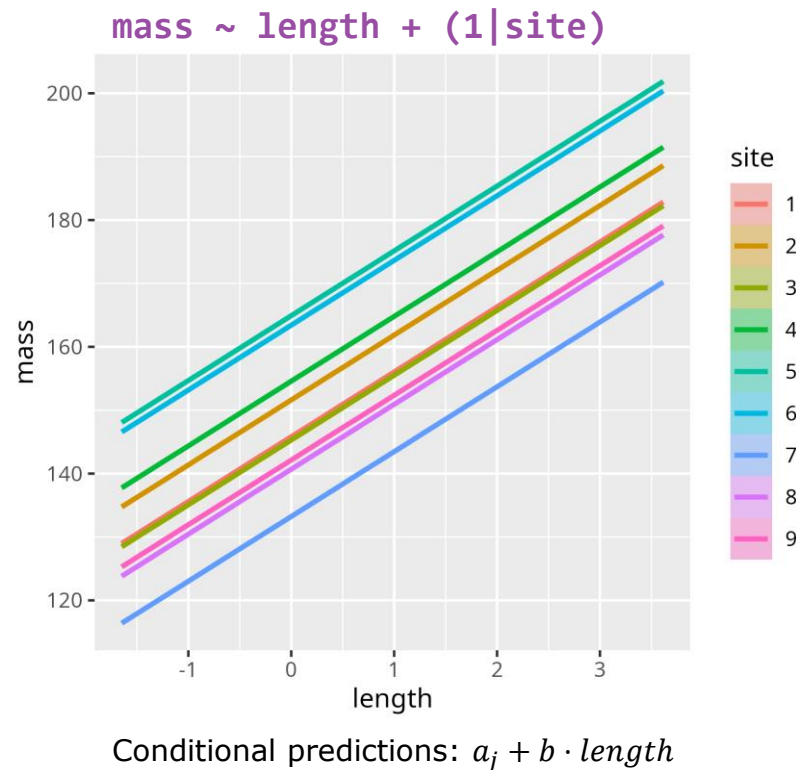Priors: $b_j \sim \text{Normal}(10,5)$

$\boldsymbol{\mu_a} \sim \text{Normal}(150,10)$

$\boldsymbol{\sigma_a} \sim \text{Exponential}(1/10)$

Random effects: $\delta_1 \dots \delta_M$ intercepts deviation

Fixed effects: μ_a joint intercept

b joint slope

mass ~ length + (1|site)

Conditional predictions: $a_j + b \cdot length$

27

(4) Partial pooling: Random slopes & intercepts

Model:
$$y_i = a_{site(i)} + b_{site(i)} \cdot length_i + \varepsilon_i$$

$$\varepsilon_i \sim \text{Normal}(0, \sigma)$$

$$a_j \sim \textbf{Normal}(\boldsymbol{\mu_a}, \boldsymbol{\sigma_a}) \qquad j = 1 \dots M$$

$$b_j \sim \textbf{Normal}(\boldsymbol{\mu_b}, \boldsymbol{\sigma_b})$$

Priors
$$\boldsymbol{\mu_a} \sim \text{Normal}(150,10)$$

$$\boldsymbol{\mu_b} \sim \text{Normal}(10,5)$$

$$\boldsymbol{\sigma_a} \sim \text{Exponential}(1/10)$$

$$\boldsymbol{\sigma_b} \sim \text{Exponential}(1/5)$$

Random effects: $a_1 \dots a_M, \ b_1 \dots b_M$
Fixed effects: μ_a, μ_b joint intercept & slope

`mass ~ length + (length|site)`

Marginal predictions: $\mu_a + \mu_b \cdot length$

(4) Partial pooling: Random slopes & intercepts

Model:
$$y_i = (\mu_a + \delta_{site(i)}) + (\mu_b + \gamma_{site(i)}) \cdot length_i + \varepsilon_i$$

$$\varepsilon_i \sim \text{Normal}(0, \sigma)$$

$$\boldsymbol{\delta_j \sim \text{Normal}(0, \sigma_a)} \qquad j = 1 \dots M$$

$$\boldsymbol{\gamma_j \sim \text{Normal}(0, \sigma_b)}$$

Priors
$$\boldsymbol{\mu_a} \sim \text{Normal}(150,10)$$

$$\boldsymbol{\mu_b} \sim \text{Normal}(10,5)$$

$$\boldsymbol{\sigma_a} \sim \text{Exponential}(1/10)$$

$$\boldsymbol{\sigma_b} \sim \text{Exponential}(1/5)$$

Random effects: $\delta_1 \dots \delta_M, \ \gamma_1 \dots \gamma_M$
Fixed effects: μ_a, μ_b joint intercept & slope

mass ~ length + (length|site)

Marginal predictions: $\mu_a + \mu_b \cdot length$

(4) Partial pooling: Random slopes & intercepts

Model:
$$y_i = (\mu_a + \delta_{site(i)}) + (\mu_b + \gamma_{site(i)}) \cdot length_i + \varepsilon_i$$

$$\varepsilon_i \sim \text{Normal}(0, \sigma)$$

$$\boldsymbol{\delta_j \sim \text{Normal}(0, \sigma_a)} \qquad j = 1 \dots M$$

$$\boldsymbol{\gamma_j \sim \text{Normal}(0, \sigma_b)}$$

Priors
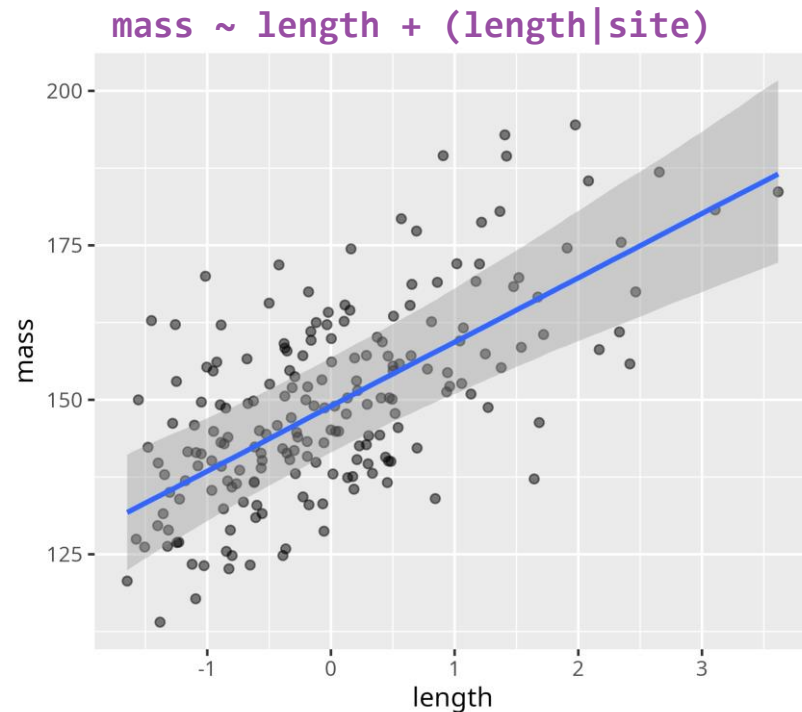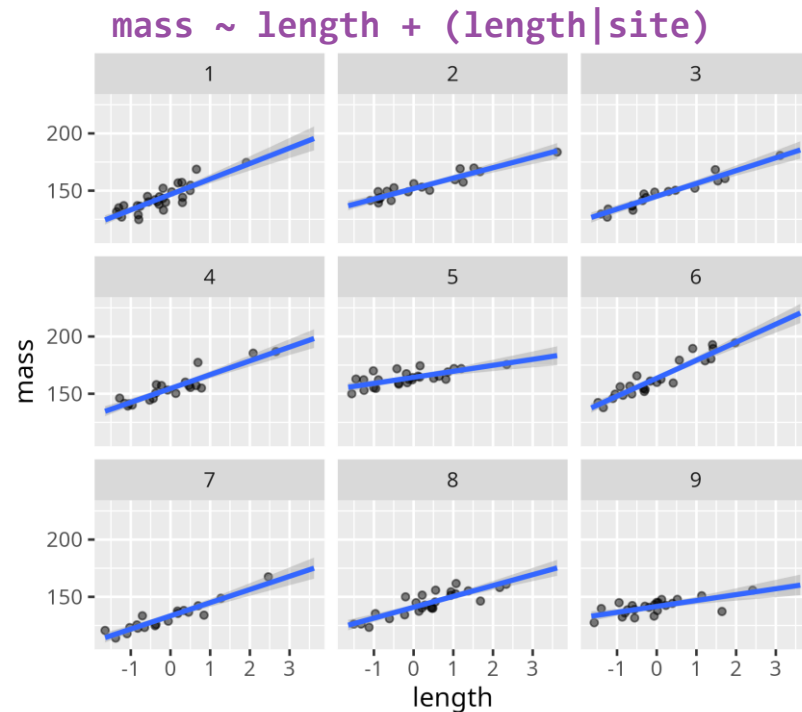
$$\mu_a \sim \text{Normal}(150,10)$$

$$\mu_b \sim \text{Normal}(10,5)$$

$$\sigma_a \sim \text{Exponential}(1/10)$$

$$\sigma_b \sim \text{Exponential}(1/5)$$

Random effects: $\delta_1 \dots \delta_M, \ \gamma_1 \dots \gamma_M$
Fixed effects: μ_a, μ_b joint intercept & slope

mass ~ length + (length|site)

Conditional predictions: $a_j + b_j \cdot length$

(4) Partial pooling: Random slopes & intercepts

Model: $y_i = (\mu_a + \delta_{site(i)}) + (\mu_b + \gamma_{site(i)}) \cdot length_i + \varepsilon_i$

$\varepsilon_i \sim \text{Normal}(0, \sigma)$

$\boldsymbol{\delta_j \sim \text{Normal}(0, \sigma_a)} \qquad j = 1 \dots M$

$\boldsymbol{\gamma_j \sim \text{Normal}(0, \sigma_b)}$

Priors $\quad \boldsymbol{\mu_a} \sim \text{Normal}(150, 10)$

$\boldsymbol{\mu_b} \sim \text{Normal}(10, 5)$

$\boldsymbol{\sigma_a} \sim \text{Exponential}(1/10)$

$\boldsymbol{\sigma_b} \sim \text{Exponential}(1/5)$

Random effects: $\delta_1 \dots \delta_M, \ \gamma_1 \dots \gamma_M$
Fixed effects: μ_a, μ_b joint intercept & slope

mass ~ length + (length|site)

Conditional predictions: $a_j + b_j \cdot length$

(4) Partial pooling: Random slopes & intercepts

Model:
$$y_i = (\boldsymbol{\mu_a} + \boldsymbol{\delta_{site(i)}}) + (\boldsymbol{\mu_b} + \boldsymbol{\gamma_{site(i)}}) \cdot length_i + \varepsilon_i$$

$$\varepsilon_i \sim \text{Normal}(0, \sigma)$$

$$\begin{pmatrix} \boldsymbol{\delta_j} \\ \boldsymbol{\gamma_j} \end{pmatrix} \sim \textbf{MVNormal}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{bmatrix} \sigma_a^2 & \rho\sigma_a\sigma_b \\ \rho\sigma_a\sigma_b & \sigma_b^2 \end{bmatrix} \right)$$

← Random intercepts & slopes not independent:

Multivariate normal distributions with
zero mean and 2x2 covariance matrix Σ

$$j = 1 \dots M$$

Priors

$$\boldsymbol{\mu_a} \sim \text{Normal}(150,10)$$

$$\boldsymbol{\mu_b} \sim \text{Normal}(10,5)$$

$$\boldsymbol{\sigma_a} \sim \text{Exponential}(1/10)$$

$$\boldsymbol{\sigma_b} \sim \text{Exponential}(1/5)$$

$$\boldsymbol{\rho} \sim \text{Uniform}(0,1)$$

← Correlation coefficient ρ

Comparison

mass ~ length

mass ~ length + (length|site)
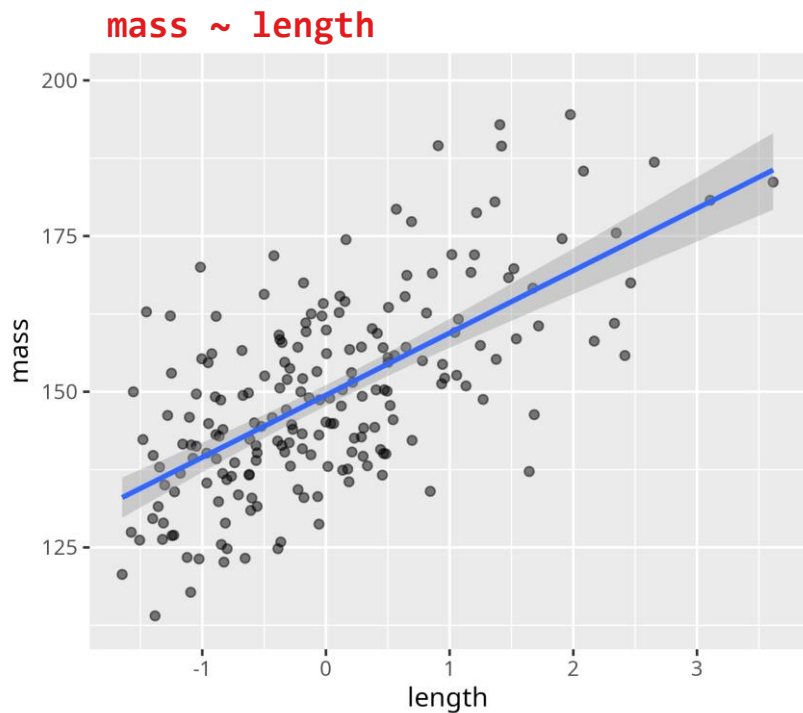
Uncertainty of a, b is smaller (overly confident).
Wrongly assumes $N = 200$ independent obs.
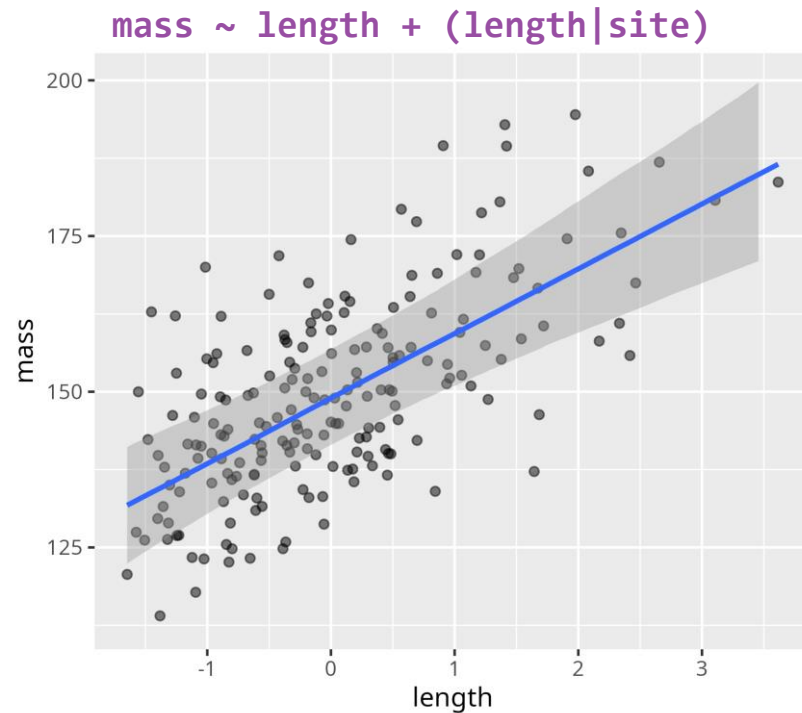
Uncertainty of μ_a, μ_b is correct.

Comparison

Here, difference is very weak
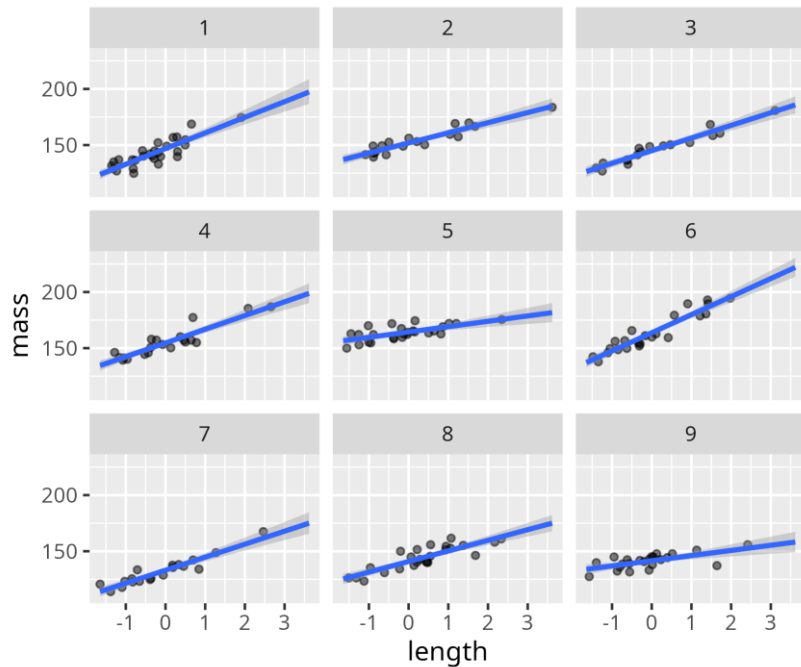
mass ~ length*site

mass ~ length + (length|site)

Regressions for each site independent.
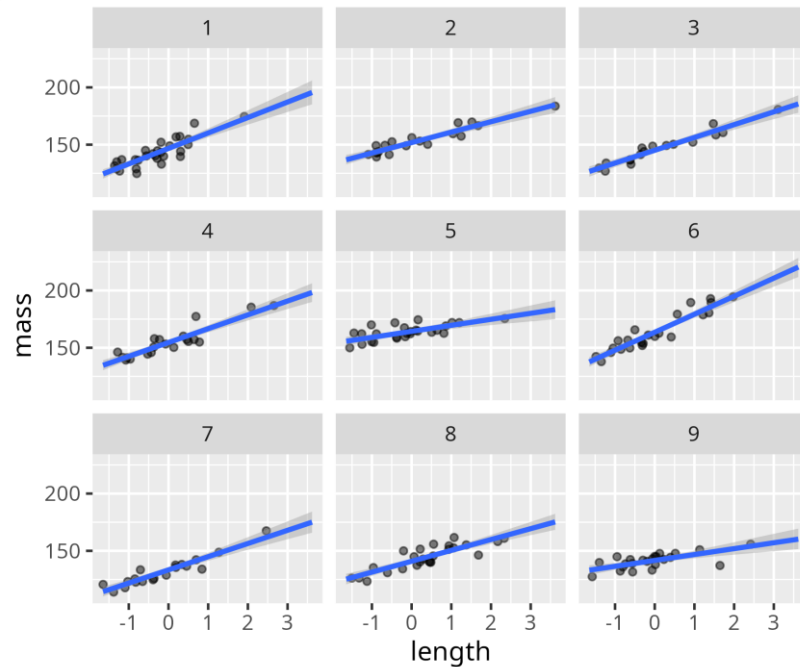
Intercepts & slopes draws to joint means „Shrinkage"

Comparison

Here, difference is very weak

mass ~ length*site

mass ~ length + (length|site)

Regressions for each site independent.

Intercepts & slopes draws to joint means „Shrinkage"

35

Comparison: unbalanced design
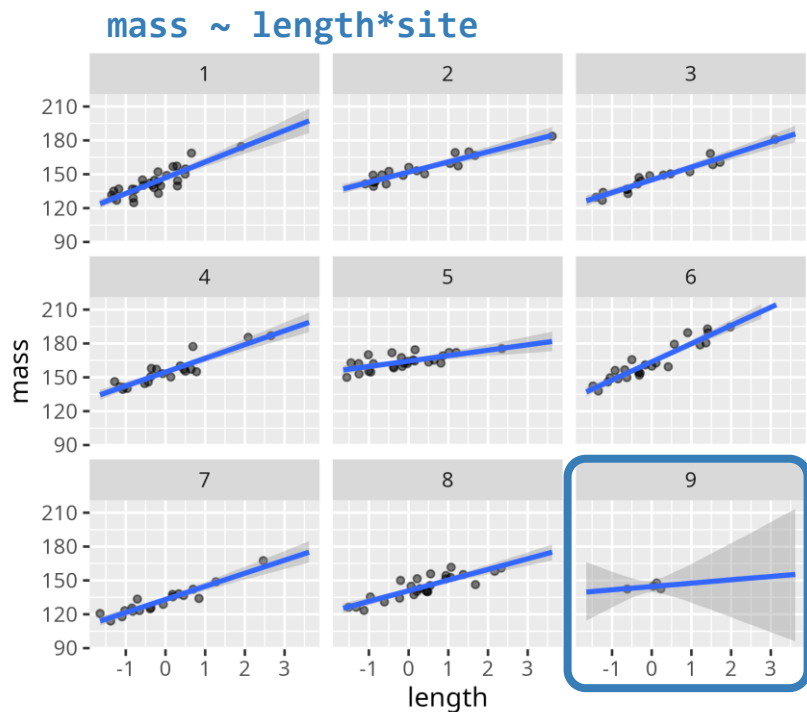
mass ~ length*site

High uncertainty in sites with few data.

mass ~ length + (length|site)
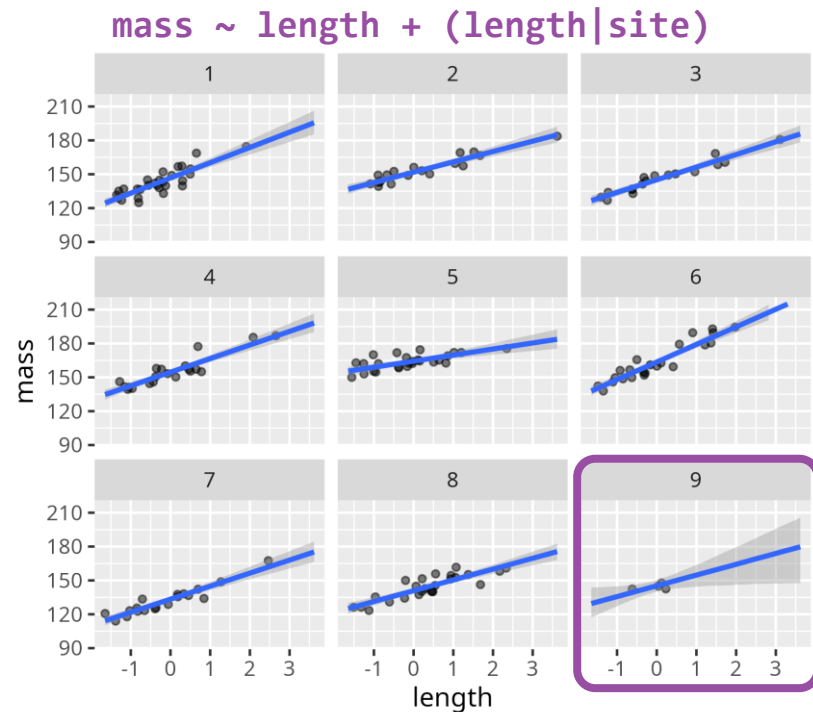
Borrowing strength from other sites.
→ Better predictions!
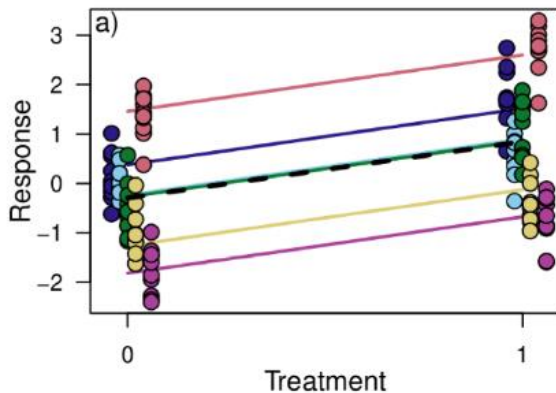
Random intercepts   **Random intercepts & „slopes"**

Categorical predictor
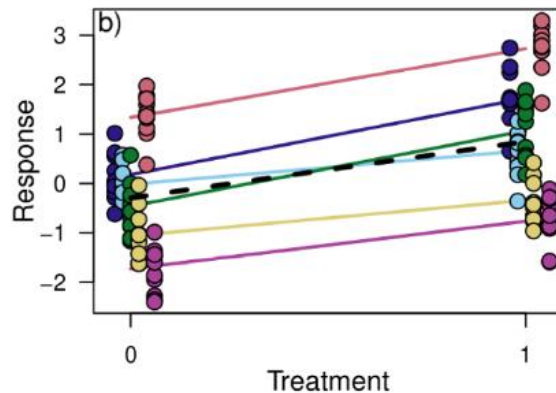
Continuous predictor
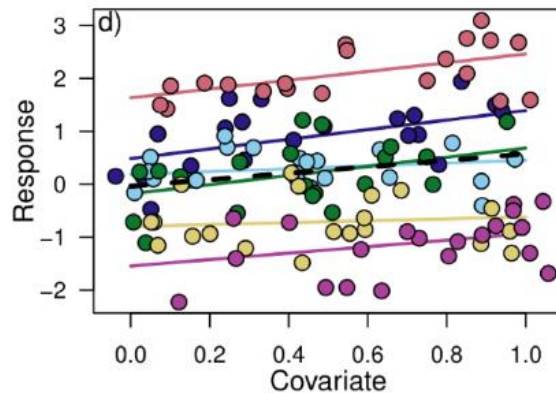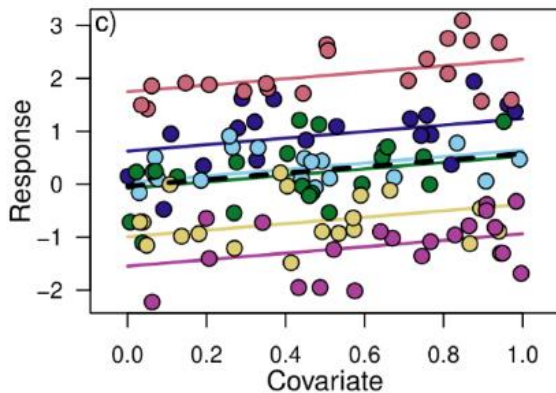
Harrison et al (2020) *PeerJ*

37

> brm(mass ~ length + (length|site), …)

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: mass ~ length + (length | site)
   Data: df (Number of observations: 200)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000
```

```
Multilevel Hyperparameters:
~site (Number of levels: 9)
                        Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
```

		Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS
σ_a →	sd(Intercept)	12.36	3.45	7.39	20.97	1.00	1269
σ_b →	sd(length)	4.65	1.57	2.52	8.72	1.00	1154
ρ →	cor(Intercept,length)	0.05	0.33	-0.58	0.64	1.00	1937

```
Regression Coefficients:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
```

		Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
μ_a →	Intercept	148.95	3.88	141.48	156.81	1.00	1265	2055
μ_b →	length	10.43	1.69	7.11	13.81	1.00	1318	1461

```
Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
```

		Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
σ →	sigma	5.28	0.27	4.79	5.86	1.00	3411	2993

Multiple grouping factors

Example

- Measured weight of $N = 200$ viper snakes

- Data collected on $M = 9$ different sites across France

- **Collection was done in $L = 5$ sampling campaigns**

Can be fully crossed (each campaign visited all sites)

or partially crossed (not all campaigns visited all sites).

→ Specify **additive (independent)** random effects
for grouping factor site and grouping factor campaign
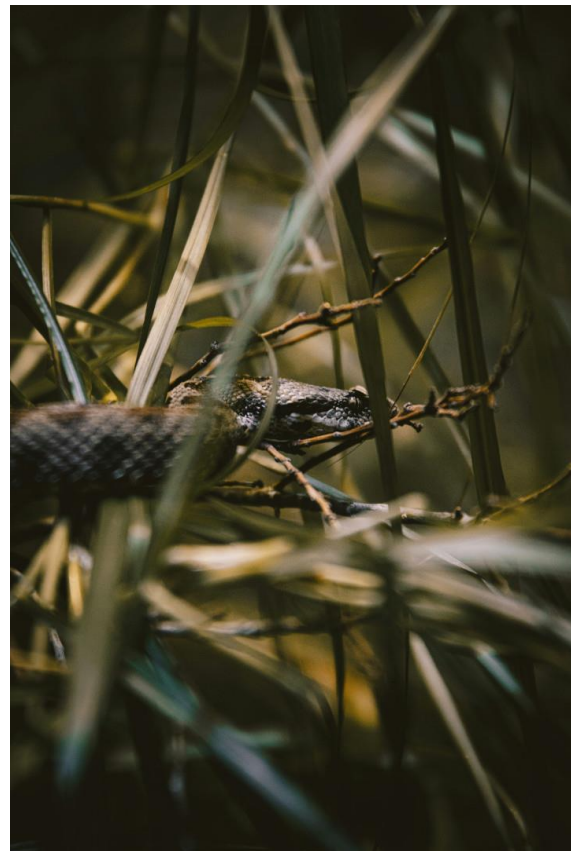
Crossed random effects

Model: $y_i = \boldsymbol{\mu_a} + \boldsymbol{\delta_{site(i)}} + \boldsymbol{\gamma_{campaign(i)}} + \varepsilon_i$

$\varepsilon_i \sim \text{Normal}(0, \sigma)$

$\boldsymbol{\delta_j \sim \text{Normal}(0, \sigma_\delta)}$ $\boldsymbol{j = 1 \ldots M}$

$\boldsymbol{\gamma_k \sim \text{Normal}(0, \sigma_\gamma)}$ $\boldsymbol{k = 1 \ldots L}$

Priors: $\boldsymbol{\mu_a} \sim \text{Normal}(150,10)$

$\boldsymbol{\sigma_\delta} \sim \text{Exponential}(1/10)$

$\boldsymbol{\sigma_\gamma} \sim \text{Exponential}(1/10)$

`mass ~ 1 + (1|site) + (1|campaign)`

← Site-effect

← Campaign-effect

← Overall mean is the only fixed effect
   in this basic example

41

Variance partitioning

Overall variance of the data is estimated on different levels

$$s_{\text{total}}^2 = s_{\text{among}}^2 + s_{\text{within}}^2$$

Site-campaign-level (among) $\sigma_\delta + \sigma_\gamma$

Explains additive deviations from overall mean μ_a

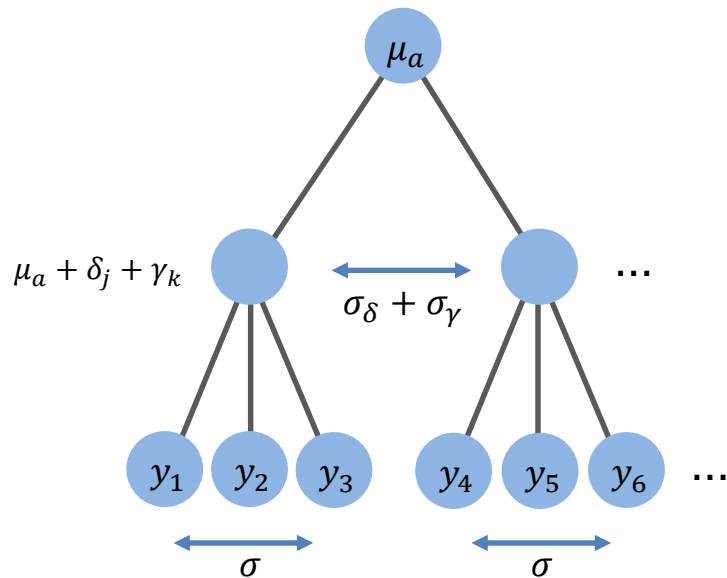caused by site (σ_δ) and campaign (σ_γ)

Observation-level (within) σ

Explains each data point's deviation from

site-campaign-mean $\mu_a + \delta_j + \gamma_k$

(All the randomness in the data which is not caused by site-level differences)

```
> brm(mass ~ 1 + (1|site) + (1|campaign), …)
```

```
  Family: gaussian
    Links: mu = identity; sigma = identity
 Formula: mass ~ 1 + (1 | site) + (1 | campaign)
    Data: df (Number of observations: 200)
   Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
          total post-warmup draws = 4000
```

```
Multilevel Hyperparameters:
~campaign (Number of levels: 5)
```

$\sigma_\gamma \rightarrow$

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	1.39	1.33	0.05	4.94	1.00	1983	2233

```
~site (Number of levels: 9)
```

$\sigma_\delta \rightarrow$

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	11.38	3.22	6.89	19.27	1.01	1085	2004

```
Regression Coefficients:
```

$\mu_a \rightarrow$

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	149.14	3.86	141.37	156.68	1.00	828	1274

```
Further Distributional Parameters:
```

$\sigma \rightarrow$

|	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	11.04	0.58	9.99	12.20	1.00	3320	2678

Example

- Measured weight of $N = 200$ viper snakes

- Data collected on $M = 9$ different sites across France

- **Sites belong to $L = 3$ distinct mountain regions**

Every site uniquely belongs to one region.

At least one region has 2 or more sites.

→ Specify **nested** random effects
   for factor site in factor region
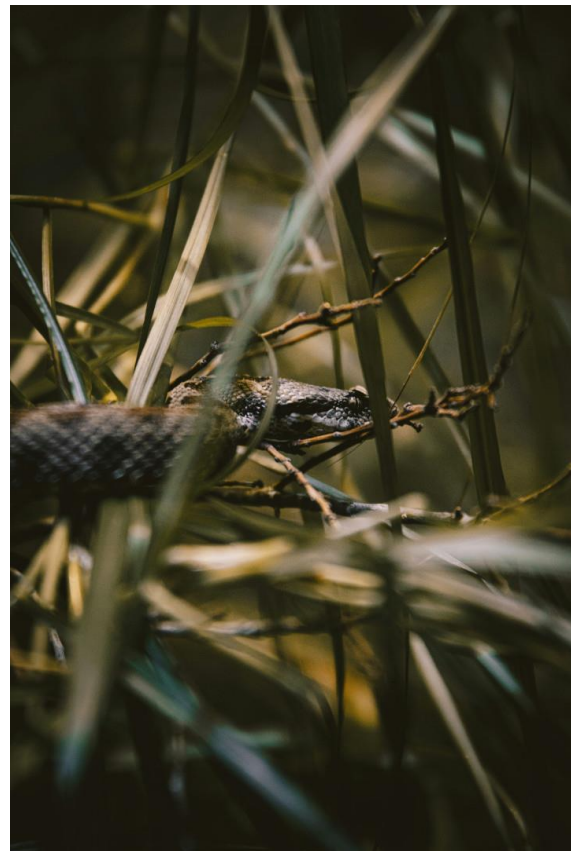
Nested random effects

Model: $y_i = \mu_a + \gamma_{region(i)} + \delta_{region:site(i)} + \varepsilon_i$

$\varepsilon_i \sim \text{Normal}(0, \sigma)$

$\gamma_k \sim \text{Normal}(0, \sigma_k) \qquad k = 1 \dots L$ ← Regional-effect

$\delta_j \sim \text{Normal}(0, \sigma_\delta) \qquad j = 1 \dots M$ ← Site-effect in region

Priors: $\mu_a \sim \text{Normal}(150, 10)$

$\sigma_\gamma \sim \text{Exponential}(1/10)$

$\sigma_\delta \sim \text{Exponential}(1/10)$

```
mass ~ 1 + (1|region/site)
mass ~ 1 + (1|region) + (1|region:site)
```

Automatically uses unique region:site combinations with *M* levels

46

Variance partitioning

Overall variance of the data is estimated on different levels

$$s_{\text{total}}^2 = s_{\text{among } regions}^2 + s_{\text{among } sites}^2 + s_{\text{within}}^2$$

Regional-level σ_γ
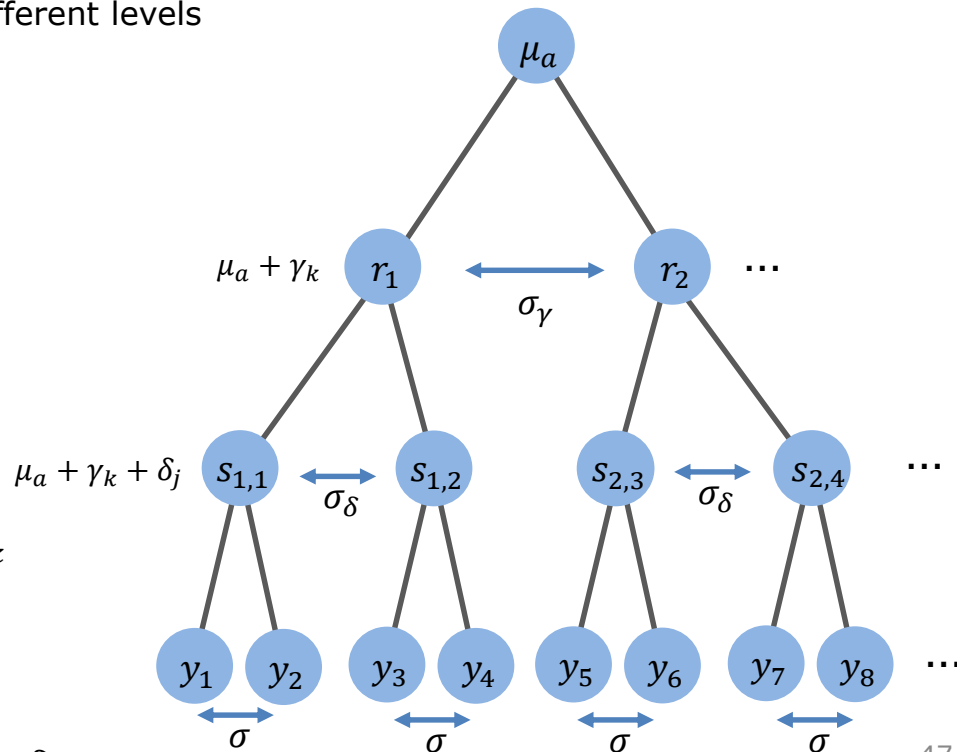
Explains deviations from overall mean μ_a

Site-level σ_δ

Explains deviations from regional mean $\mu_a + \gamma_k$

Observation-level (within) σ

Data point deviations from site-mean $\mu_a + \gamma_k + \delta_j$

```
> brm(mass ~ 1 + (1|region/site), …)
> brm(mass ~ 1 + (1|region) + (1|region:site), …)
```

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: mass ~ 1 + (1 | region/site)
   Data: df (Number of observations: 200)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000
```

```
Multilevel Hyperparameters:
~region (Number of levels: 3)
               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
```
σ_γ →
```
sd(Intercept)    12.21      6.76     2.81    30.38 1.00     1155     1073
```

```
~region:site (Number of levels: 9)
               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
```
σ_δ →
```
sd(Intercept)     6.01      2.60     2.57    12.65 1.00     1003     1395
```

```
Regression Coefficients:
            Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
```
μ_a →
```
Intercept    148.99      6.45   135.11   161.86 1.00     1594     1365
```

```
Further Distributional Parameters:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
```
σ →
```
sigma      11.05      0.57    10.01    12.21 1.00     3750     2543
```

Nested or crossed?

Nested designs

(1|F1/F2) just short for (1|F1)+(1|F1:F2)

If nested factor **F2** is uniquely labelled

(levels of **F2** and **F1:F2** identical)

this is same as crossed (1|F1)+(1|F2)

Crossed designs

No need to be fully crossed for additive model.

(1|F1)+(1|F2) works for partially crossed, too.

Table 1. Schematic illustration of crossed and nested designs

Nested design   (1|F1/F2)

Factor 1	a	b	c	d
A	X	X		
B			X	X

(header: Factor 2)

Partially crossed design   (1|F1)+(1|F2)

Factor 1	a	b	c	d
A	X	X		
B	X	X		
C			X	X
D			X	X

(header: Factor 2)

Fully crossed design   (1|F1)+(1|F2)

Factor 1	a	b	c	d
A	X	X	X	X
B	X	X	X	X
C	X	X	X	X
D	X	X	X	X

(header: Factor 2)

Schnielzeth & Nakagawa (2013) *MEE*

Continuous correlation structures

Example

- Measured weight of $N = 200$ viper snakes

- Data collected on $M = 9$ different sites in the same region

Instead of using site as categorical grouping predictor,

we could use location (longitude & latitude) as continuous

„grouping" predictor.

„Everything is related to everything else,

but near things are more related than distant things." (Tobler)

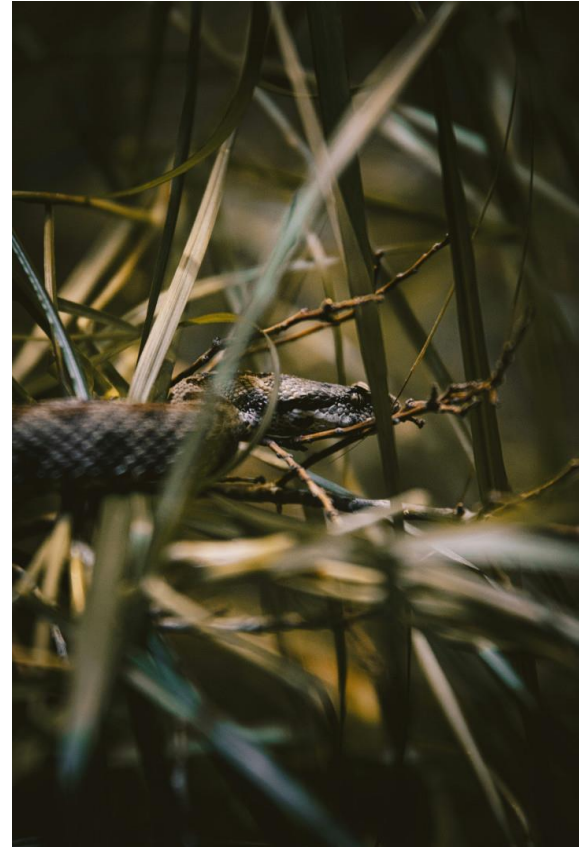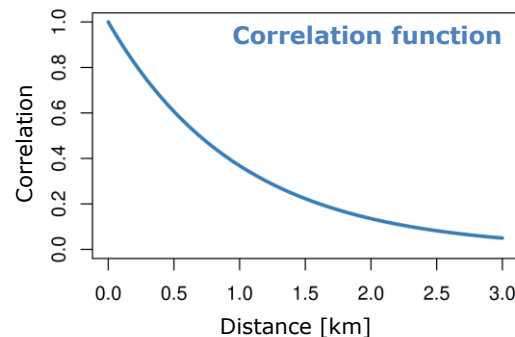→ Specify spatial autocorrelation model.

Example

- Measured weight of $N = 200$ viper snakes

- Data collected on $M = 9$ different sites in the same region

Instead of using site as categorical grouping predictor,

we could use location (longitude & latitude) as continuous

„grouping" predictor.

„Everything is related to everything else,

but near things are more related than distant things." (Tobler)
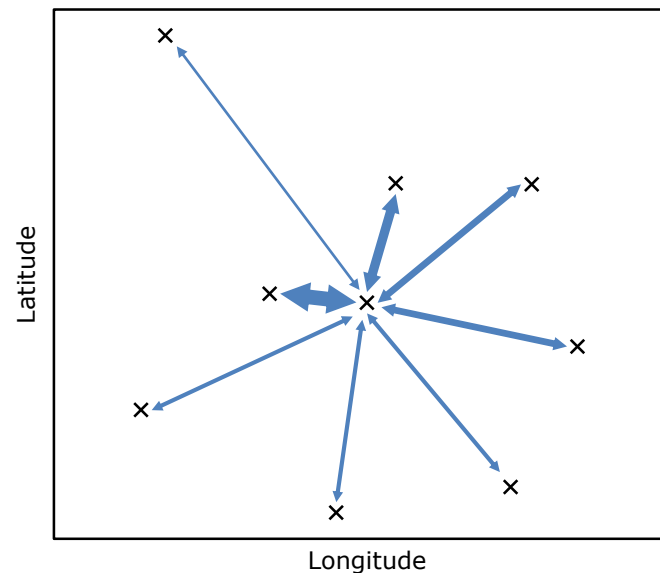
→ Specify spatial autocorrelation model.

Longitude / Latitude plot

Correlation function

Spatial autocorrelation model

Model: $y_i = \mu_a + \boldsymbol{\varepsilon_i}$

$\boldsymbol{\varepsilon} \sim \mathbf{MVNormal}(\mathbf{0}, \boldsymbol{\sigma^2 R})$

Correlation: $\boldsymbol{R_{ij}} = \exp(-\mathbf{dist}(\boldsymbol{i,j})/\boldsymbol{\theta})$

Priors: $\mu_a \sim \text{Normal}(150,10)$

$\boldsymbol{\sigma} \sim \text{Exponential}(1/10)$

$\boldsymbol{\theta} \sim \text{Exponential}(1/100)$

N-dimensional vector of residuals $\boldsymbol{\varepsilon}$:
multivariate normal distribution

← with $N \times N$ correlation matrix $\boldsymbol{R}$ & variance $\boldsymbol{\sigma^2}$

← Correlation decreases with
spatial distance of datapoints i, j

← Range parameter of spatial autocorrelation

- Model for spatial autocorrelation of **residuals**.

- Unfortunately not (yet?) implemented in brms.

→ Code in **Stan**, or use alternative models.

Fixed residual correlation

Model: $\quad y_i = \mu_a + \boldsymbol{\varepsilon_i}$

$\quad\quad\quad\quad \boldsymbol{\varepsilon} \sim \textbf{MVNormal}(\mathbf{0}, \boldsymbol{\sigma^2 R})$

Correlation: $\boldsymbol{R_{ij}}$ **provided by user**

Priors: $\quad \mu_a \sim \text{Normal}(150,10)$

$\quad\quad\quad\quad \boldsymbol{\sigma} \sim \text{Exponential}(1/10)$

```
brm( y ~ 1 + fcor(R),
     data  = … ,
     data2 = list(R=R) )
```

← Choose range θ yourself and compute, e.g.,
  $R_{ij} = \exp(-\text{dist}(i,j)/\theta)$ from data.
  Obtain θ from variogram.

- Not ideal, since correlation range θ is not estimated jointly with the other parameters.
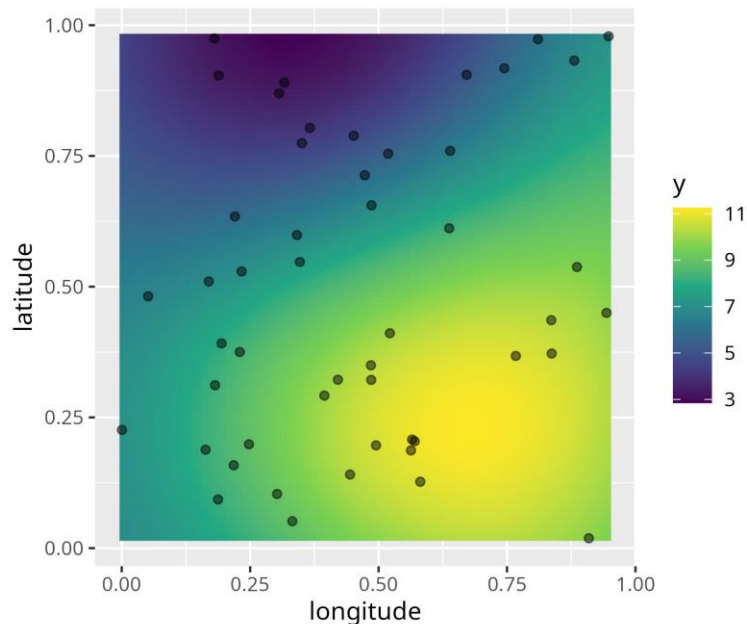- But it's a fast and efficient solution.

More „adventures in covariance" (McElreath)

brms has a couple of options:

- Fixed correlation model
- Gaussian process regression
- Generalized additive models (GAM)
  → but see mgcv package

- Adjacency matrix $R_{ij} = \begin{cases} 1 & i,j \text{ neighbours} \\ 0 & \text{otherwise} \end{cases}$

- ...

You can model autocorrelations

- spatial
- temporal
- spatio-temporal
- phylogenetic

Many of these models can get **very slow** even for moderate N.

→ **INLA** is designed for that. Very efficient !

Summary

Random or fixed effects ?

Classical / oldschool interpretation

You are interested in overall means / overall effects, but must account for non-independence in data?

→ **site as a random grouping factor**

~ 1 + (1|site)   random intercepts

~ x + (1|site)   random intercepts

~ x + (x|site)   random intercepts & slopes

You are interested in differences in means / differences in effects between sites?

→ **site as fixed effect**

~ site   ANOVA

~ x + site   ANCOVA (diff. in intercepts)

~ x * site   ANCOVA (diff. in intercepts & slopes)

Modern / relaxed interpretation

Can use random factor models for both to overcome difficulties like unbalanced / heterogeneous data

What's Bayesian about it ?

- Sometimes lme4 just does not converge !

- Correct quantification of uncertainties *on all levels*.

- Works for all numbers of groups or numbers of obs.
  Already 2-level grouping factor is OK.

- Interpretation "easy" through levels of priors

- Frequentist assumption: Random effects drawn from
  a larger population of other, unmeasured groups

- Bayes: no such limitation

- Non-normal distribution of random effects?
  → Bayes (Stan) can handle that!

Summary

- Bayesian stats does not free you from model assumptions !

- If residuals are not independent, you must account for it.

- Otherwise: Fixed effects wrongly confident, CIs too narrow

- Thinking hard about your data will reveal
  random effects structure.

- Formula notation just as in lme4.

- Combine any kind of (generalized) linear (or nonlinear)
  model with discrete (random) grouping factors and/or
  continuous correlation structure …

- … in one R-package !!

The Bayesian 3D printer
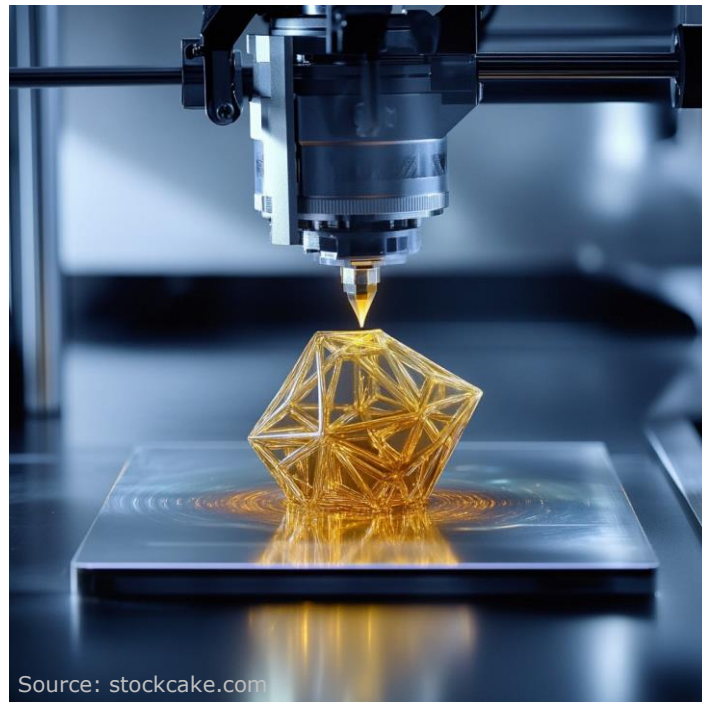
Source: stockcake.com

Further reading

Bolker, B. M. (2015). Linear and generalized linear mixed models. In Fox, G. A. , Negrete-Yankelevich, S. and Sosa, V. J. (eds). Ecological Statistics: Contemporary theory and application. *Oxford Academic*. https://doi.org/10.1093/acprof:oso/9780199672547.003.0014

Fieberg, J. (2024). Statistics 4 Ecologists. https://statistics4ecologists-v2.netlify.app/ [Chapter 18]

Gelman, A., Hill, J. (2006). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press http://www.stat.columbia.edu/~gelman/arm/

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E. D., Robinson, B. S., Hodgson, D. J., & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, 6, e4794. https://doi.org/10.7717/peerj.4794

Marshall, D. J. (2024). Principles of experimental design for ecology and evolution. *Ecology Letters*, 27(4), e14400. https://doi.org/10.1111/ele.14400

McElreath, R. (2020). Statistical Rethinking: A Bayesian Course with Examples in R and STAN (2nd ed.). *Chapman and Hall/CRC*. https://doi.org/10.1201/9780429029608 [Chapter 13]

Silk, M. J., Harrison, X. A., & Hodgson, D. J. (2020). Perils and pitfalls of mixed-effects regression models in biology. *PeerJ*, 8, e9522. https://doi.org/10.7717/peerj.9522

Warton, D. (2022). Eco-Stats: Data Analysis in Ecology. *Springer (Methods in Statistical Ecology)* [Chapter 6]