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

- 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

- \rightarrow 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
- \rightarrow Still want to make inference on whole population



Random grouping factors

Examples:

Site:Data collected at multiple locationsYear:Data collected ~same time in different yearsObserver:Data collected by different researchersIndividual:Multiple measurements on the same test objectUnit:Multiple experimental units like chambers / tanksSpecies:Data on higher level (e.g. family) contains
multiple species



Word salad !!



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) ?

 \rightarrow No, observations not independent

brm(weight ~ site) ?

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



(1) Complete pooling

Model: $y_i = a + \varepsilon_i$ $i = 1 \dots N$ $\varepsilon_i \sim \text{Normal}(0, \sigma)$ 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:
$$y_i = a_{site(i)} + \varepsilon_i$$
 $i = 1 \dots N$ $\varepsilon_i \sim Normal(0, \sigma)$ Priors: $a_j \sim Normal(150, 10)$ $j = 1 \dots M$

- 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)



site

(3) Partial pooling

Model:
$$y_i = a_{site(i)} + \varepsilon_i$$
 $i = 1 \dots N$
 $\varepsilon_i \sim \text{Normal}(0, \sigma)$
 $a_j \sim \text{Normal}(\mu_a, \sigma_a)$ $j = 1 \dots M$
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 $a_1 \dots a_M$
- 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



(3) Partial pooling

Model:
$$y_i = \mu_a + \delta_{site(i)} + \varepsilon_i$$
 $i = 1 \dots N$
 $\varepsilon_i \sim \text{Normal}(0, \sigma)$
 $\delta_j \sim \text{Normal}(0, \sigma_a)$ $j = 1 \dots M$
Priors: $\mu_a \sim \text{Normal}(150, 10)$
 $\sigma_a \sim \text{Exponential}(1/10)$

- 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



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**



13

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_i)

Hyperparameters μ_a , σ_a informed by all 9 sites

- \rightarrow Some information is shared / pooled
- \rightarrow Borrowing strength
- → For unbalanced designs this improves prediction in levels with few observations.



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 $\sigma_a \rightarrow 0$ and all $\delta_j \rightarrow 0$

But the residuals' likelihood $y_i \sim \text{Normal}(\mu_a + \delta_{site(i)}, \sigma)$ would decrease if all δ_j were close to 0 (worse model fit)

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



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-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 \sim 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
\sigma_a \rightarrow sd(Intercept)
                         11.64
                                              6.99
                                     3.26
                                                       19.70 1.01
                                                                        814
                                                                                1353
        Regression Coefficients:
                  Estimate Est.Error l-95% CI u-95% CI Rhat Bulk ESS Tail ESS
\mu_a \rightarrow
       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 \rightarrow sigma
                 11.02
                             0.57
                                      9.97
                                              12.20 1.00
                                                              2527
                                                                        2566
```

Fixed and random effects

Fixed & random effects

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

 \rightarrow "Fixed effect" μ_a

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

→ "Random effect" δ_i (j = 1 ... M)

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

 \rightarrow No, observations not independent

brm(weight ~ length*site) ?

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



(1) Complete pooling

Model:
$$y_i = a + b \cdot length_i + \varepsilon_i$$
 $i = 1 \dots N$
 $\varepsilon_i \sim Normal(0, \sigma)$
Priors: $a \sim Normal(150, 10)$
 $b \sim Normal(10, 5)$

- 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



(2) No pooling

Model: $y_i = a_{site(i)} + b_{site(i)} \cdot length_i + \varepsilon_i$ $\varepsilon_i \sim Normal(0, \sigma)$

Priors: $a_j \sim \text{Normal}(150,10)$ $j = 1 \dots M$ $b_j \sim \text{Normal}(10,5)$

- 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)





(2) No pooling

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

 $\varepsilon_i \sim Normal(0, \sigma)$

Priors: $a_j \sim \text{Normal}(150,10)$ $j = 1 \dots M$ $b_j \sim \text{Normal}(10,5)$

- 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)



Model:	$y_i = \boldsymbol{a_{site(i)}} + b \cdot length_i + \varepsilon_i$	
	$\varepsilon_i \sim \text{Normal}(0, \sigma)$	
	$a_j \sim \operatorname{Normal}(\mu_a, \sigma_a) \qquad j = 1 \dots M$	
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 interceptsFixed effects: μ_a joint interceptbjoint slope



Model: $y_i = (\mu_a + \delta_{site(i)}) + b \cdot length_i + \varepsilon_i$ $\varepsilon_i \sim \text{Normal}(0, \sigma)$ $\delta_j \sim \text{Normal}(0, \sigma_a)$ $j = 1 \dots M$ 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	ioint slope



Model:	$y_i = (\boldsymbol{\mu}_a + \boldsymbol{\delta}_{site(i)}) + b \cdot length_i + \varepsilon_i$
	$\varepsilon_i \sim \text{Normal}(0, \sigma)$
	$\boldsymbol{\delta_j} \sim \mathbf{Normal}(0, \boldsymbol{\sigma_a}) \qquad j = 1 \dots M$
Priors:	$b_j \sim \text{Normal}(10,5)$
	μ _a ~ 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



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 deviationFixed effects: μ_a joint interceptbjoint slope



Model: $y_i = a_{site(i)} + b_{site(i)} \cdot length_i + \varepsilon_i$ $\varepsilon_i \sim \text{Normal}(0, \sigma)$ $a_i \sim \text{Normal}(\mu_a, \sigma_a) \qquad j = 1 \dots M$ $b_i \sim \text{Normal}(\mu_b, \sigma_b)$ Priors $\mu_a \sim \text{Normal}(150,10)$ $\mu_h \sim \text{Normal}(10,5)$ $\sigma_a \sim \text{Exponential}(1/10)$ $\sigma_h \sim \text{Exponential}(1/5)$

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)





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)



Model: $y_i = (\mu_a + \delta_{site(i)}) + (\mu_b + \gamma_{site(i)}) \cdot length_i + \varepsilon_i$ $\varepsilon_i \sim \text{Normal}(0, \sigma)$ $\delta_i \sim \text{Normal}(\mathbf{0}, \sigma_a) \qquad j = 1 \dots M$ $\gamma_i \sim \text{Normal}(0, \sigma_b)$ $\mu_a \sim \text{Normal}(150,10)$ Priors $\mu_h \sim \text{Normal}(10,5)$ $\sigma_a \sim \text{Exponential}(1/10)$ $\sigma_h \sim \text{Exponential}(1/5)$

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

3 200 -150 5 6 200 -Magentian Sector -150 -7 8 9 200 -150 length

mass ~ length + (length|site)

Conditional predictions: $a_j + b_j \cdot length$



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



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

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

$$\begin{pmatrix} \boldsymbol{\delta}_j \\ \boldsymbol{\gamma}_j \end{pmatrix} \sim \text{MVNormal} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{bmatrix} \boldsymbol{\sigma}_a^2 & \boldsymbol{\rho} \boldsymbol{\sigma}_a \boldsymbol{\sigma}_b \\ \boldsymbol{\rho} \boldsymbol{\sigma}_a \boldsymbol{\sigma}_b & \boldsymbol{\sigma}_b^2 \end{bmatrix} \right)$$
$$j = 1 \dots M$$

- Priors $\mu_a \sim \text{Normal}(150,10)$
 - $\mu_b \sim \text{Normal}(10,5)$
 - $\sigma_a \sim \text{Exponential}(1/10)$
 - $\sigma_b \sim \text{Exponential}(1/5)$
 - $\boldsymbol{\rho} \sim \text{Uniform}(0,1)$

 \leftarrow Random intercepts & slopes not independent:

Multivariate normal distributions with zero mean and 2x2 covariance matrix $\boldsymbol{\Sigma}$

 $\leftarrow \text{ Correlation coefficient } \rho$

Comparison



Uncertainty of a, b is smaller (overly confident). Wrongly assumes N = 200 independent obs.



Uncertainty of μ_a , μ_b is correct.



Regressions for each site independent.

Intercepts & slopes draws to joint means "Shrinkage"



Regressions for each site independent.

Intercepts & slopes draws to joint means "Shrinkage"

Comparison: unbalanced design



High uncertainty in sites with few data.

mass ~ length + (length|site)



Borrowing strength from other sites. \rightarrow Better predictions!



```
> 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
\sigma_a \rightarrow sd(Intercept)
                                 12.36
                                             3.45
                                                      7.39
                                                               20.97 1.00
                                                                               1269
\sigma_h \rightarrow sd(length)
                                 4.65
                                            1.57 2.52
                                                               8.72 1.00
                                                                              1154
   → cor(Intercept, length)
                                 0.05
                                            0.33
D
                                                     -0.58
                                                                0.64 1.00
                                                                               1937
      Regression Coefficients:
                 Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
\mu_a \rightarrow Intercept
                   148.95
                                3.88
                                       141.48
                                                 156.81 1.00
                                                                  1265
                                                                            2055
\mu_h \rightarrow 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
\sigma \rightarrow 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: $y_i = \mu_a + \delta_{site(i)} + \gamma_{campaign(i)} + \varepsilon_i$ $\varepsilon_i \sim \text{Normal}(0, \sigma)$ $\delta_i \sim \text{Normal}(0, \sigma_{\delta}) \qquad j = 1 \dots M$ $\gamma_k \sim \text{Normal}(0, \sigma_{\gamma}) \qquad k = 1 \dots L$ Priors: $\mu_a \sim \text{Normal}(150,10)$ $\sigma_{\delta} \sim \text{Exponential}(1/10)$ $\sigma_{\gamma} \sim \text{Exponential}(1/10)$

mass $\sim 1 + (1|site) + (1|campaign)$

- \leftarrow Site-effect
- ← Campaign-effect
- ← Overall mean is the only fixed effect in this basic example

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 $\,\mu_a$

caused by site (σ_{δ}) and campaign (σ_{γ})

Observation-level (within) σ

Explains each data point's deviation from

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

(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) Estimate Est.Error l-95% CI u-95% CI Rhat Bulk ESS Tail ESS $\sigma_v \rightarrow sd(Intercept)$ 1.39 1.33 0.05 4.94 1.00 1983 2233 ~site (Number of levels: 9) Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS $\sigma_{\delta} \rightarrow \text{sd(Intercept)}$ 11.38 3.22 6.89 19.27 1.01 1085 2004 Regression Coefficients: Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk ESS Tail ESS $\mu_a \rightarrow$ Intercept 149.14 3.86 141.37 156.68 1.00 828 1274 Further Distributional Parameters: Estimate Est.Error l-95% CI u-95% CI Rhat Bulk ESS Tail ESS 11.04 0.58 9.99 12.20 1.00 3320 $\sigma \rightarrow$ sigma 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: $y_i = \mu_a + \gamma_{region(i)} + \delta_{region:site(i)} + \varepsilon_i$ mass
mass $\varepsilon_i \sim Normal(0, \sigma)$ $\gamma_k \sim Normal(0, \sigma_k)$ $k = 1 \dots L$ \leftarrow Reg $\delta_j \sim Normal(0, \sigma_\delta)$ $j = 1 \dots M$ \leftarrow SitePriors: $\mu_a \sim Normal(150, 10)$ $\sigma_\gamma \sim Exponential(1/10)$

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

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

← Regional-effect← Site-effect in region

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

Variance partitioning

Overall variance of the data is estimated on different levels

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

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 \sim 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
\sigma_{\gamma} \rightarrow \text{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
\sigma_{\delta} \rightarrow \text{sd(Intercept)} 6.01 2.60 2.57 12.65 1.00
                                                                    1003
                                                                              1395
      Regression Coefficients:
                 Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk ESS Tail ESS
\mu_a \rightarrow 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
                           0.57 10.01
                                                            3750
               11.05
                                            12.21 1.00
                                                                     2543
\sigma \rightarrow sigma
```

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	L F1/F2)		
	Factor 2			
Factor 1	a	b	с	d
A	х	Х		
В			Х	Х
Partially crossed of	lesign (1	F1)+(1	F2)	
	Factor 2			
Factor 1	a	b	с	d
A	х	Х		
В	Х	Х		
С			Х	X
D			Х	X
Fully crossed desi	gn (1	L F1)+(1	F2)	
	Factor 2			
Factor 1	а	b	с	d
A	Х	Х	Х	X
В	Х	Х	Х	X
С	Х	Х	Х	X
D	Х	Х	Х	X

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)

 \rightarrow 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.



 \rightarrow Specify spatial autocorrelation model.



Spatial autocorrelation model

Model: $y_i = \mu_a + \varepsilon_i$ $\varepsilon \sim \text{MVNormal}(0, \sigma^2 R)$ Correlation: $R_{ij} = \exp(-\text{dist}(i, j)/\theta)$ Priors: $\mu_a \sim \text{Normal}(150, 10)$ $\sigma \sim \text{Exponential}(1/10)$ $\theta \sim \text{Exponential}(1/100)$

- N-dimensional vector of residuals $oldsymbol{arepsilon}$: multivariate normal distribution
- \leftarrow with $N \times N$ correlation matrix **R** & variance σ^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.
- \rightarrow Code in **Stan**, or use alternative models.

Fixed residual correlation

- Model: $y_i = \mu_a + \varepsilon_i$ $\varepsilon \sim \text{MVNormal}(0, \sigma^2 R)$ Correlation: R_{ij} provided by userPriors: $\mu_a \sim \text{Normal}(150, 10)$ $\sigma \sim \text{Exponential}(1/10)$
- Not ideal, since correlation range θ is not estimated jointly with the other parameters.
- But it's a fast and efficient solution.

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

More "adventures in covariance" (McElreath)

brms has a couple of options:

- Fixed correlation model
- Gaussian process regression
- Generalized additive models (GAM) \rightarrow 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.

 \rightarrow **INLA** is designed for that. Very efficient !



Random or fixed effects ?

Classical / oldschool interpretation

You are interested in overall means / overall effects, but must account for non-independence in data?	You are interested in differences in means / differences in effects between sites?	
ightarrow site as a random grouping factor	\rightarrow site as fixed effect	
~ 1 + (1 site) random intercepts	~ site ANOVA	
~ x + (1 site) random intercepts	~ x + site ANCOVA (diff. in intercepts)	
~ x + (x site) random intercepts & slopes	~ 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 Ime4 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?
 - \rightarrow 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 Ime4.
- 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



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*. <u>https://doi.org/10.1093/acprof:oso/9780199672547.003.0014</u>

Fieberg, J. (2024). Statistics 4 Ecologists. <u>https://statistics4ecologists-v2.netlify.app/</u> [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. <u>https://doi.org/10.7717/peerj.4794</u>

Inchausti, P. (2023). Statistical Modeling With R: a dual frequentist and Bayesian approach for life scientists. *Oxford University Press*. [Chapter 13]

Kery, M. & Kellner, F. (2024): Applied Statistical Modelling for Ecologists. *Elsevier*. [Chapter 10]

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

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