Introduction to Bayesian Statistics

Part 4

Linear Models

Benjamin Rosenbaum

iDiv 2025

In this lecture

- What is a linear model?
- Continuous predictors (Regression)
- Categorical predictors (ANOVA)
- Categorical & continuous predictors (ANCOVA)

In-between:

- Model selection
- Post-hoc analysis

What is a linear model?

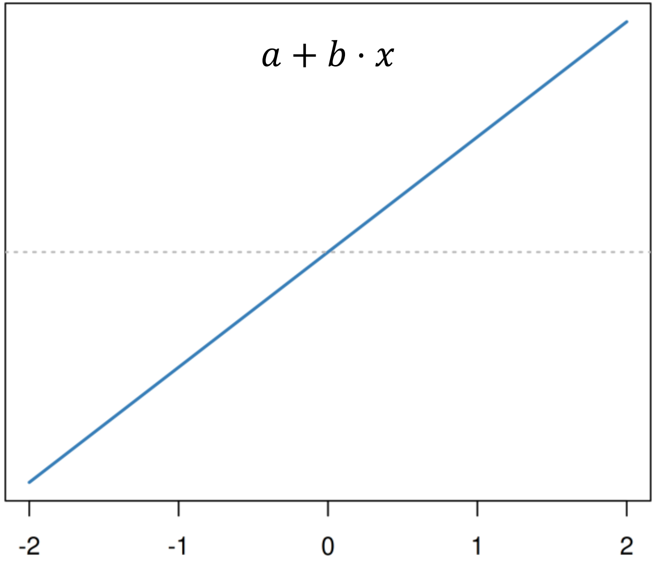
Linear functions

Linear in $\boldsymbol{x}$ (predictor)

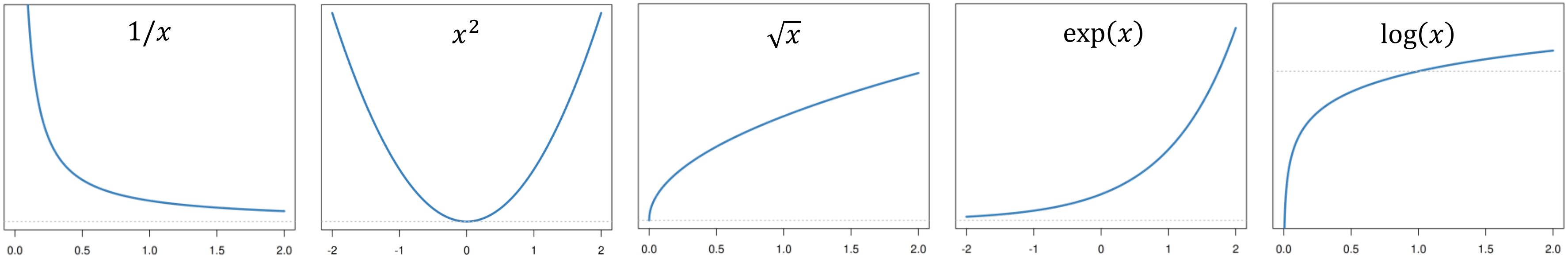
$f(x) = a + b \cdot \boldsymbol{x}$

Additive with constant a (intercept)

Multiplication only with constant b (slope)



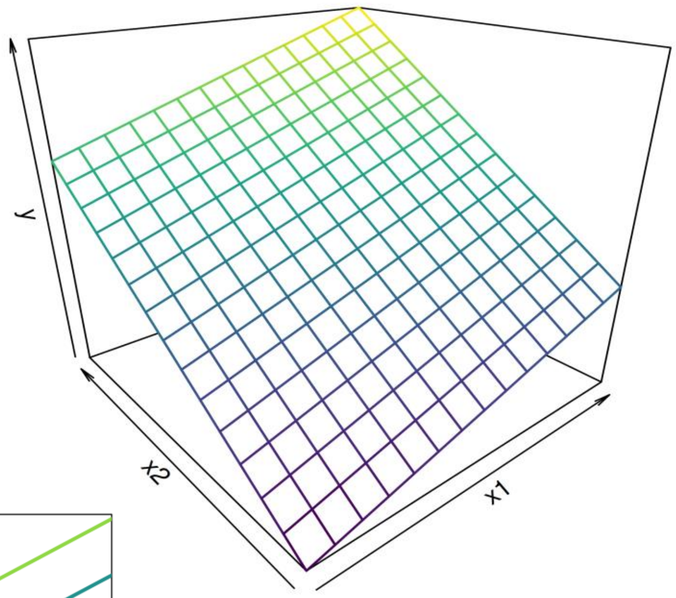
Some **nonlinear** functions:

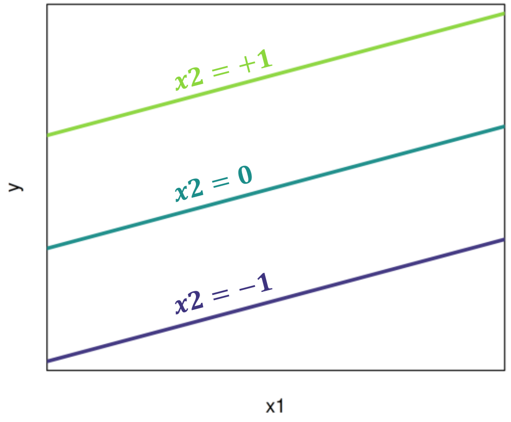


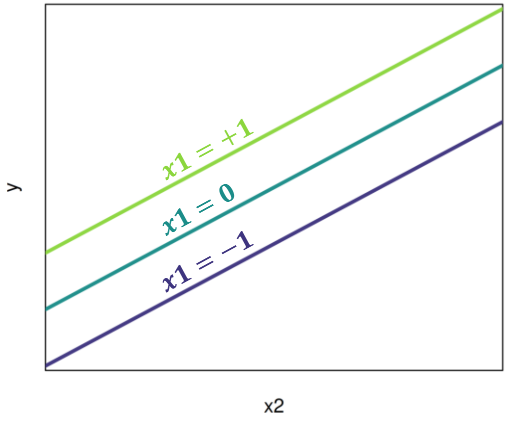
Linear functions

Extend to **multiple predictors** $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots$

$$f(x) = b_0 + b_1 \cdot \boldsymbol{x_1} + b_2 \cdot \boldsymbol{x_2}$$







Linear statistical models

Linear in b (parameters) and
Gaussian random errors ε (normally distributed)

$$y(x) = \boldsymbol{b_0} + \boldsymbol{b_1} \cdot x + \boldsymbol{\varepsilon}$$

$$y(x) = \boldsymbol{b_0} + \boldsymbol{b_1} \cdot x_1 + \boldsymbol{b_2} \cdot x_2 + \boldsymbol{\varepsilon}$$

Linear statistical models in the frequentist world:

Analytical solution (formula) for parameter estimates

Easy computation with `lm()`

Nonlinear in b, for example:

$$y(x) = \boldsymbol{b_0} + x^{\boldsymbol{b_1}} + \boldsymbol{\varepsilon}$$

$$y(x) = \boldsymbol{b_0} + \exp(\boldsymbol{b_1} \cdot x) + \boldsymbol{\varepsilon}$$

Nonlinear models in the frequentist world:

Maximum likelihood estimation (iterative algorithm)

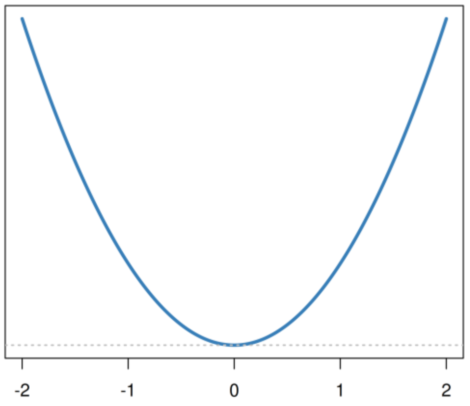
Linear statistical models

Quadratic (polynomial) relationships

$$y = \boldsymbol{b_0} + \boldsymbol{b_1} \cdot x + \boldsymbol{b_2} \cdot x^2 + \boldsymbol{\varepsilon}$$

$$= \boldsymbol{b_0} + \boldsymbol{b_1} \cdot x_1 + \boldsymbol{b_2} \cdot x_2 + \varepsilon$$

define
$\boldsymbol{x_1 = x}$
$\boldsymbol{x_2 = x^2}$

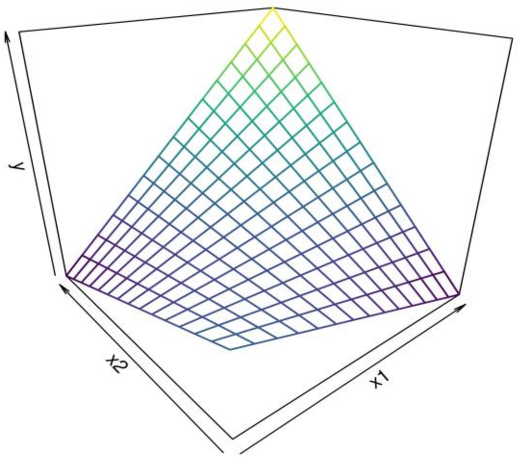


Interaction effects

$$y = \boldsymbol{b_0} + \boldsymbol{b_1} \cdot x_1 + \boldsymbol{b_2} \cdot x_2 + \boldsymbol{b_3} \cdot x_1 \cdot x_2 + \boldsymbol{\varepsilon}$$

$$= \boldsymbol{b_0} + \boldsymbol{b_1} \cdot x_1 + \boldsymbol{b_2} \cdot x_2 + \boldsymbol{b_3} \cdot x_3 + \boldsymbol{\varepsilon}$$

define $\boldsymbol{x_3 = x_1 x_2}$



→ Some nonlinear relationships can be described with linear statistical models (**linear in the parameters**)

Linear statistical models

Transformation of response variable

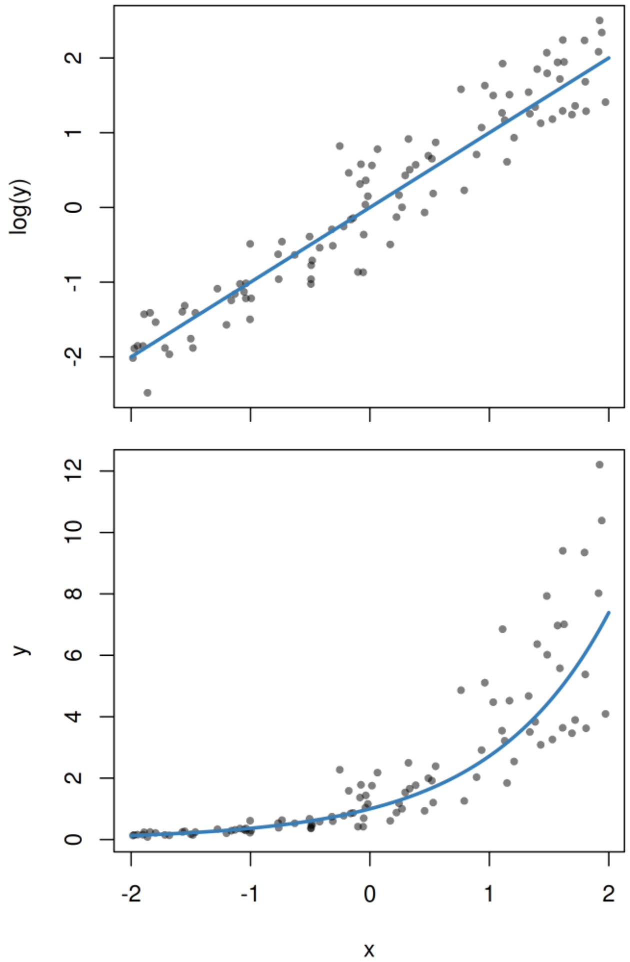
$$\log(y) = \boldsymbol{b_0} + \boldsymbol{b_1} \cdot x + \boldsymbol{\varepsilon}$$

Attention: model becomes multiplicative

When back transforming to y-scale

$$\begin{aligned} y &= \exp(b_0 + b_1 \cdot x + \varepsilon) \\ &= \exp(b_0) \cdot \exp(b_1 x) \cdot \exp(\varepsilon) \\ &= \tilde{b}_0 \cdot \exp(b_1 x) \cdot \tilde{\varepsilon} \end{aligned}$$

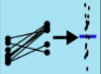
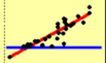
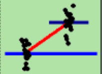
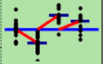
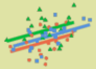
→ Sometimes statistical models can be "linearized" by transformation



Common statistical tests are linear models

Last updated: 28 June, 2019.Also check out the [Python version](#)!

See worked examples and more details at the accompanying notebook: https://lindeloev.github.io/tests-as-linear

	Common name	Built-in function in R	Equivalent linear model in R	Exact?	The linear model in words	Icon
Simple regression: lm(y ~ 1 + x)	**y is independent of x** P: One-sample t-test N: Wilcoxon signed-rank	t.test(y) wilcox.test(y)	lm(y ~ 1) lm(signed_rank(y) ~ 1)	✓ for N >14	One number (intercept, i.e., the mean) predicts **y**. - (Same, but it predicts the *signed rank* of **y**.)	
	P: Paired-sample t-test N: Wilcoxon matched pairs	t.test(y_1, y_2, paired=TRUE) wilcox.test(y_1, y_2, paired=TRUE)	lm(y_2 - y_1 ~ 1) lm(signed_rank(y_2 - y_1) ~ 1)	✓ for N >14	One intercept predicts the pairwise **y_2-y_1** differences. - (Same, but it predicts the *signed rank* of **y_2-y_1**.)	
	y ~ continuous x P: Pearson correlation N: Spearman correlation	cor.test(x, y, method='Pearson') cor.test(x, y, method='Spearman')	lm(y ~ 1 + x) lm(rank(y) ~ 1 + rank(x))	✓ for N >10	One intercept plus **x** multiplied by a number (slope) predicts **y**. - (Same, but with *ranked* **x** and **y**)	
	y ~ discrete x P: Two-sample t-test P: Welch's t-test N: Mann-Whitney U	t.test(y_1, y_2, var.equal=TRUE) t.test(y_1, y_2, var.equal=FALSE) wilcox.test(y_1, y_2)	lm(y ~ 1 + G_2)[A] gls(y ~ 1 + G_2, weights=...[B])[A] lm(signed_rank(y) ~ 1 + G_2)[A]	✓ ✓ for N >11	An intercept for **group 1** (plus a difference if **group 2**) predicts **y**. - (Same, but with one variance *per group* instead of one common.) - (Same, but it predicts the *signed rank* of **y**.)	
Multiple regression: lm(y ~ 1 + x_1 + x_2 +...)	P: One-way ANOVA N: Kruskal-Wallis	aov(y ~ group) kruskal.test(y ~ group)	lm(y ~ 1 + G_2 + G_3 +...+ G_N)[A] lm(rank(y) ~ 1 + G_2 + G_3 +...+ G_N)[A]	✓ for N >11	An intercept for **group 1** (plus a difference if group ≠ 1) predicts **y**. - (Same, but it predicts the *rank* of **y**.)	
	P: One-way ANCOVA	aov(y ~ group + x)	lm(y ~ 1 + G_2 + G_3 +...+ G_N + x)[A]	✓	- (Same, but plus a slope on **x**.) *Note: this is discrete AND continuous. ANCOVAs are ANOVAs with a continuous x.*	
	P: Two-way ANOVA	aov(y ~ group * sex)	lm(y ~ 1 + G_2 + G_3 +...+ G_N + S_2 + S_3 +...+ S_K + G_2*S_2 + G_3*S_3 + ... + G_N*S_K)	✓	Interaction term: changing **sex** changes the **y ~ group** parameters. *Note: **$G_{2 \text{ to } N}$** is an [indicator (0 or 1)](#) for each non-intercept levels of the **group** variable. Similarly for **$S_{2 \text{ to } K}$** for sex. The first line (with G_i) is main effect of group, the second (with S_i) for sex and the third is the **group × sex** interaction. For two levels (e.g. male/female), line 2 would just be "S_2" and line 3 would be S_2 multiplied with each G_i.*	*[Coming]*
	Counts ~ discrete x N: Chi-square test	chisq.test(groupXsex_table)	**Equivalent log-linear model** glm(y ~ 1 + G_2 + G_3 + ... + G_N + S_2 + S_3+ ... + S_K + G_2*S_2 + G_3*S_3 +...+ G_N*S_K, family=...)[A]	✓	Interaction term: (Same as Two-way ANOVA.) *Note: Run glm using the following arguments:* `glm(model, family=poisson())` *As linear-model, the Chi-square test is $log(y_i) = log(N) + log(\alpha_i) + log(\beta_j) + log(\alpha_i\beta_j)$ where α_i and β_j are proportions. See more info in [the accompanying notebook](#).*	*Same as Two-way ANOVA*
	N: Goodness of fit	chisq.test(y)	glm(y ~ 1 + G_2 + G_3 +...+ G_N, family=...)[A]	✓	(Same as One-way ANOVA and see Chi-Square note.)	*1W-ANOVA*

https://lindeloev.github.io/tests-as-linear/

Bayesian stats & linearity

Linearity actually **not that important**

MCMC does not care if deterministic model part is

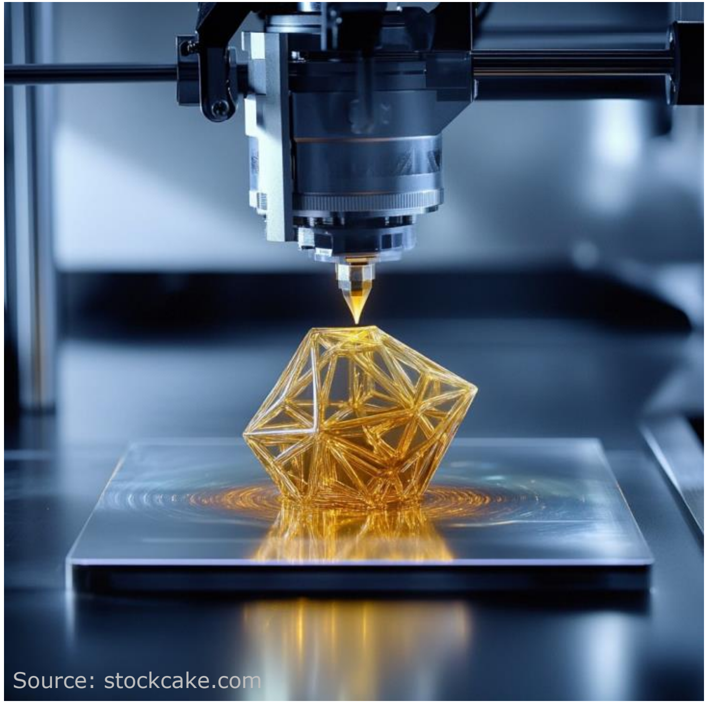
linear $\mu(x) = a + b \cdot x$

or nonlinear $\mu(x) = \frac{a \cdot x}{b + x}$

However, nonlinear (and also polynomial) models should only be considered when there is a good reason, not just because they would fit the data better.

Principle of parsimony, Occam's razor (14th century):
"Entities must not be multiplied beyond necessity"

The Bayesian 3D printer



Source: stockcake.com

Continuous predictors
(Linear regression)

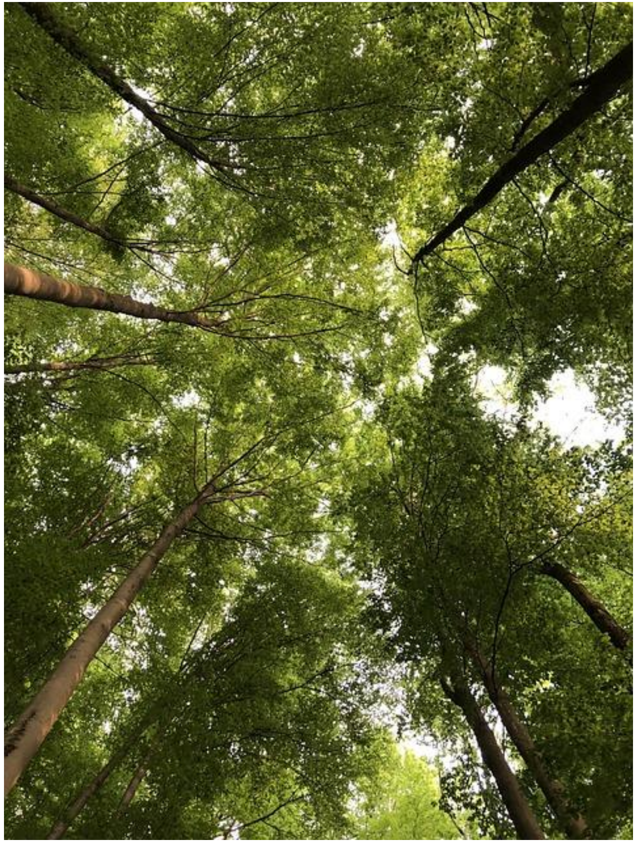
Single predictor

Example: Latitudinal gradient of plant size

Global database with:

- log10 of plant height as response
- latitude as predictor

Later: include precipitation as environmental predictor



Single predictor

Example: Latitudinal gradient of plant size

Stochastic part: $\log(height) \sim \text{Normal}(\mu, \sigma)$

Deterministic part: $\mu = b_0 + b_1 \cdot lat$

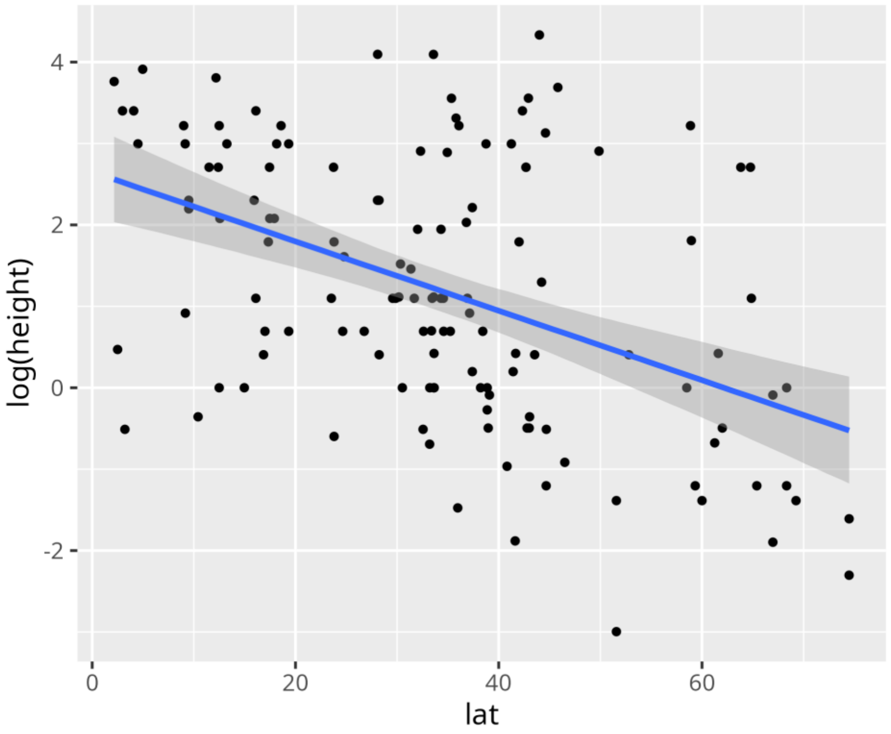
```
> brm(log(height)~lat, data=globalPlants)
```

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: log(height) ~ lat
   Data: globalPlants (Number of observations: 131)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000

Regression Coefficients:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     2.65      0.28     2.09     3.20 1.00     3634     2805
lat          -0.04      0.01    -0.06    -0.03 1.00     4126     2981

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     1.50      0.09     1.33     1.70 1.00     3419     2835

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```



Single predictor

Example: Latitudinal gradient of plant size

Stochastic part: $\log(height) \sim \text{Normal}(\mu, \sigma)$

Deterministic part: $\mu = b_0 + b_1 \cdot lat$

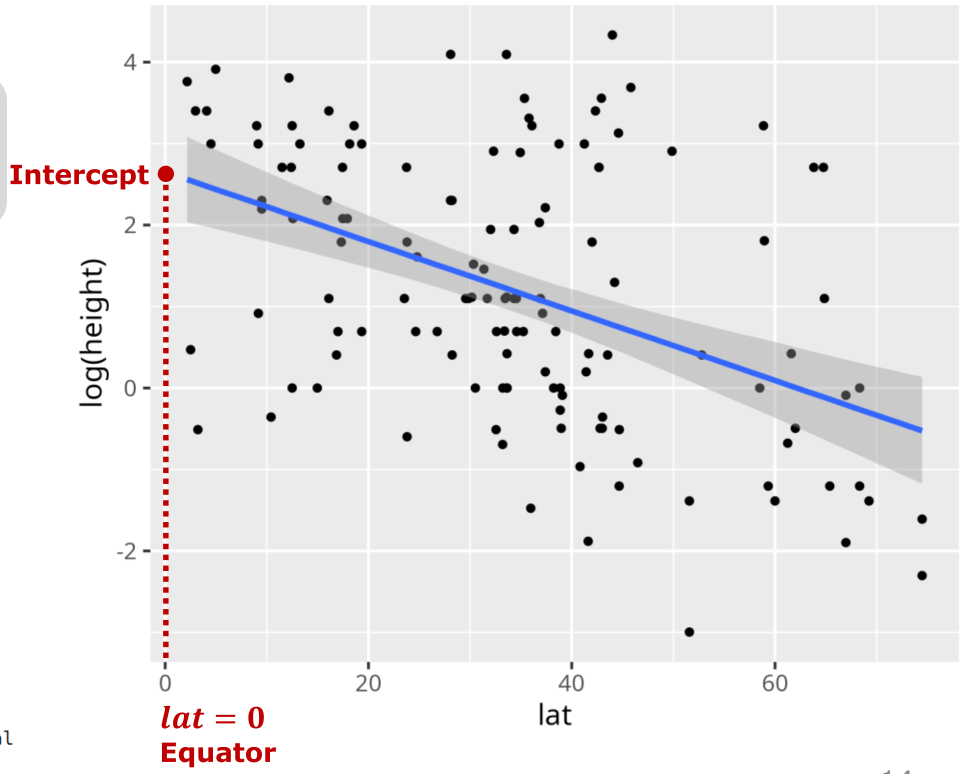
```
> brm(log(height)~lat, data=globalPlants)

 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: log(height) ~ lat
   Data: globalPlants (Number of observations: 131)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000

Regression Coefficients:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     2.65      0.28     2.09     3.20 1.00     3634     2805
lat          -0.04      0.01    -0.06    -0.03 1.00     4126     2981

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     1.50      0.09     1.33     1.70 1.00     3419     2835

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```



Single predictor

Example: Latitudinal gradient of plant size

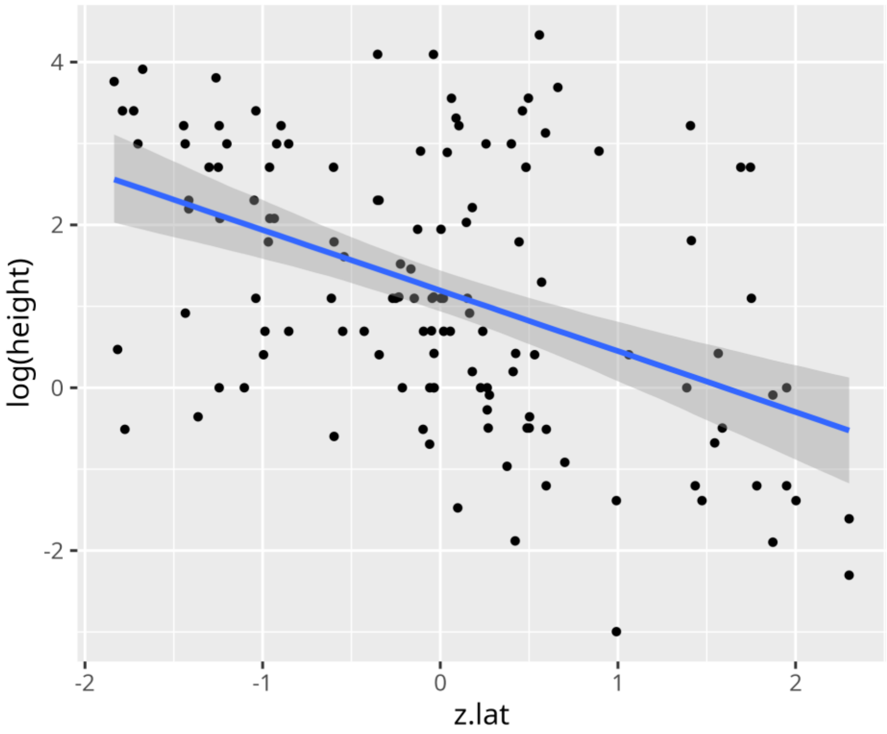
scale predictor (→ $\text{mean} = 0, \text{sdev} = 1$)

$$z.lat = \frac{lat - \text{mean}(lat)}{\text{sdev}(lat)}$$ „z-score"

Deterministic part: $\mu = b_0 + b_1 \cdot z.lat$

```
> brm(log(height)~z.lat)
```

```
Regression Coefficients:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     1.19      0.13     0.94     1.44 1.00     3799     2873
z.lat        -0.75      0.13    -1.00    -0.49 1.00     4196     2815
```



Single predictor

Example: Latitudinal gradient of plant size

scale predictor (→ $\text{mean} = 0, \text{sdev} = 1$)

$$z.lat = \frac{lat - \text{mean}(lat)}{\text{sdev}(lat)}$$ „z-score"

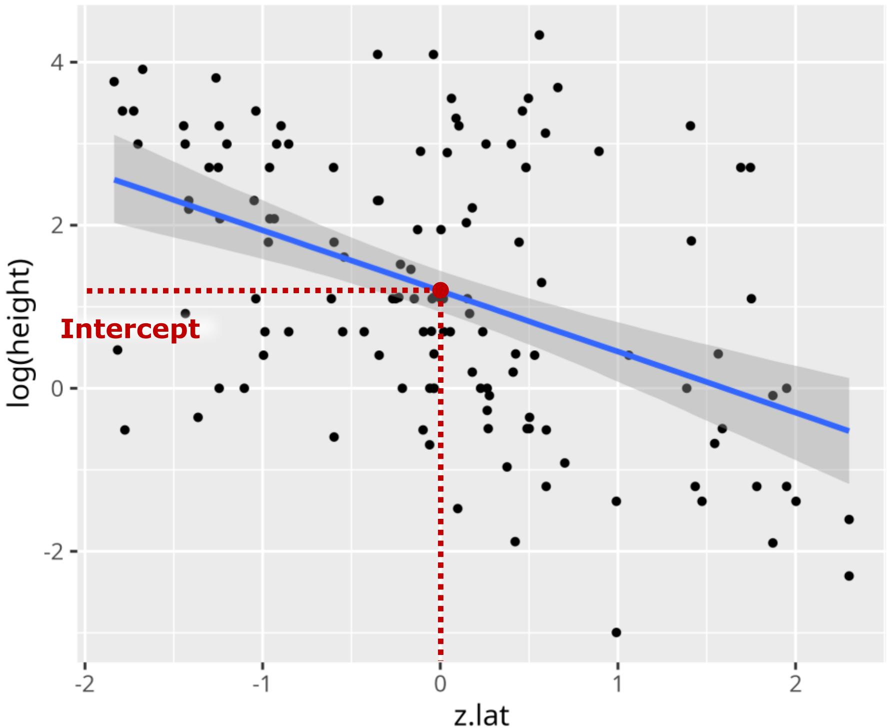
Deterministic part: $\mu = b_0 + b_1 \cdot z.lat$

```
> brm(log(height)~z.lat)
```

Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	1.19	0.13	0.94	1.44	1.00	3799	2873
z.lat	-0.75	0.13	-1.00	-0.49	1.00	4196	2815

Now intercept is predicted $\log(height)$ when predictor lat is at its average and slope is effect for 1 sdev increment of lat



$z.lat = 0$
mean latitude

Multiple predictors

Example: Latitudinal gradient of plant size

Stochastic part: $\log(height) \sim \text{Normal}(\mu, \sigma)$

Deterministic part: $\mu = b_0 + b_1 \cdot lat + \boldsymbol{b_2 \cdot rain}$

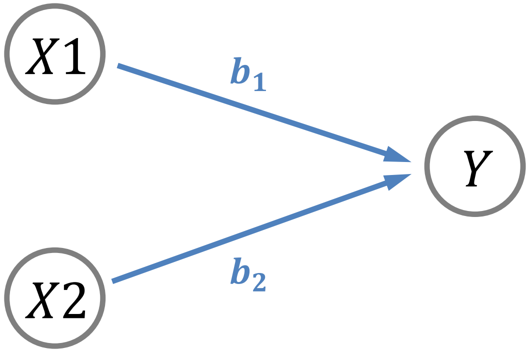
```
> brm(log(height)~z.lat+z.rain)

 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: log(height) ~ z.lat + z.rain
   Data: globalPlants (Number of observations: 131)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000

Regression Coefficients:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     1.20      0.13     0.95     1.44 1.00     3186     2563
z.lat        -0.48      0.16    -0.78    -0.19 1.00     3490     3236
z.rain        0.46      0.15     0.16     0.76 1.00     3518     3088

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     1.45      0.09     1.29     1.65 1.00     3605     2835

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```



Multiple predictors

Example: Latitudinal gradient of plant size

Stochastic part: $\log(height) \sim \text{Normal}(\mu, \sigma)$

Deterministic part: $\mu = b_0 + b_1 \cdot lat + b_2 \cdot rain$

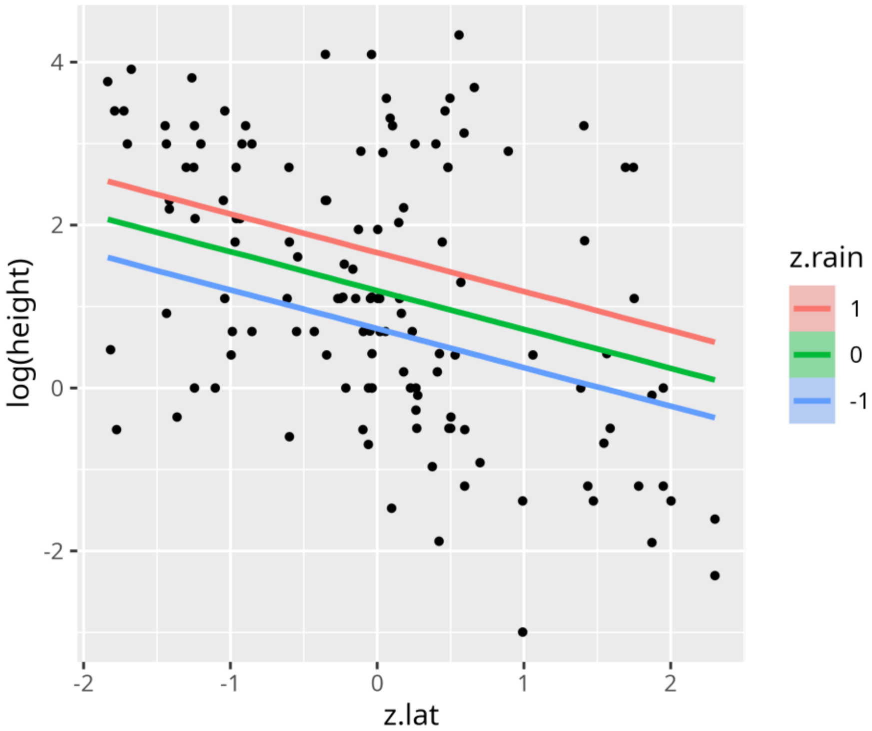
```
> brm(log(height)~z.lat+z.rain)
```

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: log(height) ~ z.lat + z.rain
   Data: globalPlants (Number of observations: 131)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000

Regression Coefficients:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     1.20      0.13     0.95     1.44 1.00     3186     2563
z.lat        -0.48      0.16    -0.78    -0.19 1.00     3490     3236
z.rain        0.46      0.15     0.16     0.76 1.00     3518     3088

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     1.45      0.09     1.29     1.65 1.00     3605     2835

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```



Multiple predictors

Example: Latitudinal gradient of plant size

Stochastic part: $\log(height) \sim \text{Normal}(\mu, \sigma)$

Deterministic part: $\mu = b_0 + b_1 \cdot lat + b_2 \cdot rain$

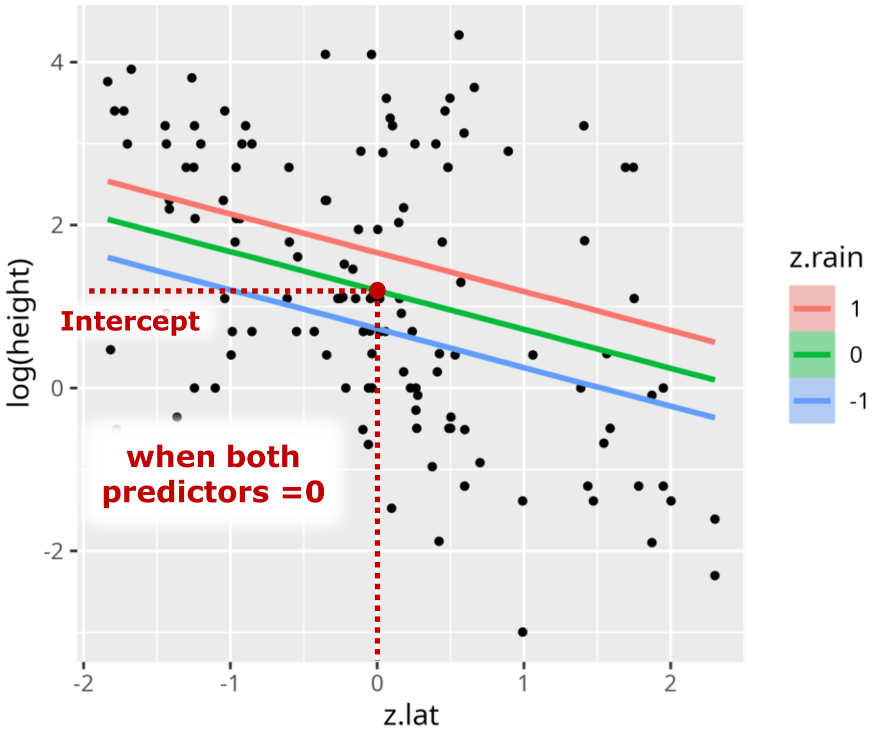
```
> brm(log(height)~z.lat+z.rain)

 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: log(height) ~ z.lat + z.rain
   Data: globalPlants (Number of observations: 131)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000

Regression Coefficients:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     1.20      0.13     0.95     1.44 1.00     3186     2563
z.lat        -0.48      0.16    -0.78    -0.19 1.00     3490     3236
z.rain        0.46      0.15     0.16     0.76 1.00     3518     3088

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     1.45      0.09     1.29     1.65 1.00     3605     2835

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```



Multiple predictors

Example: Latitudinal gradient of plant size

Stochastic part: $\log(height) \sim \text{Normal}(\mu, \sigma)$

Deterministic part: $\mu = b_0 + b_1 \cdot lat + b_2 \cdot rain$

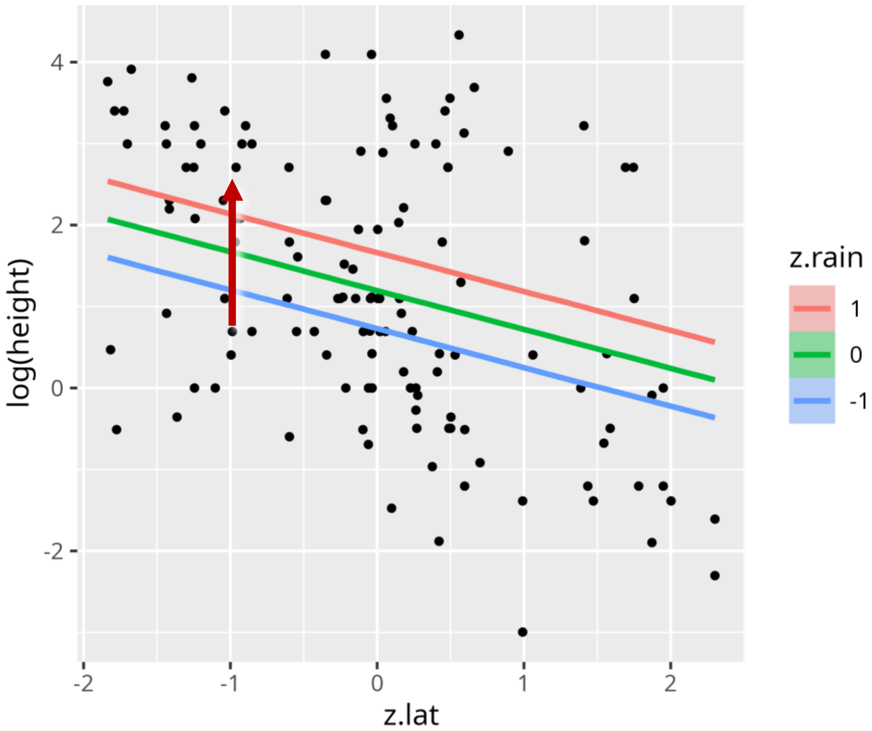
$$\mu = (b_0 + \mathbf{b_2} \cdot rain) + b_1 \cdot lat$$

Intercept depends on ***rain***

Slope constant

2^{nd} variable shifts intercept by b_2

Simpler interpretation when using scaled variables $z.lat$ and $z.rain$



Multiple predictors

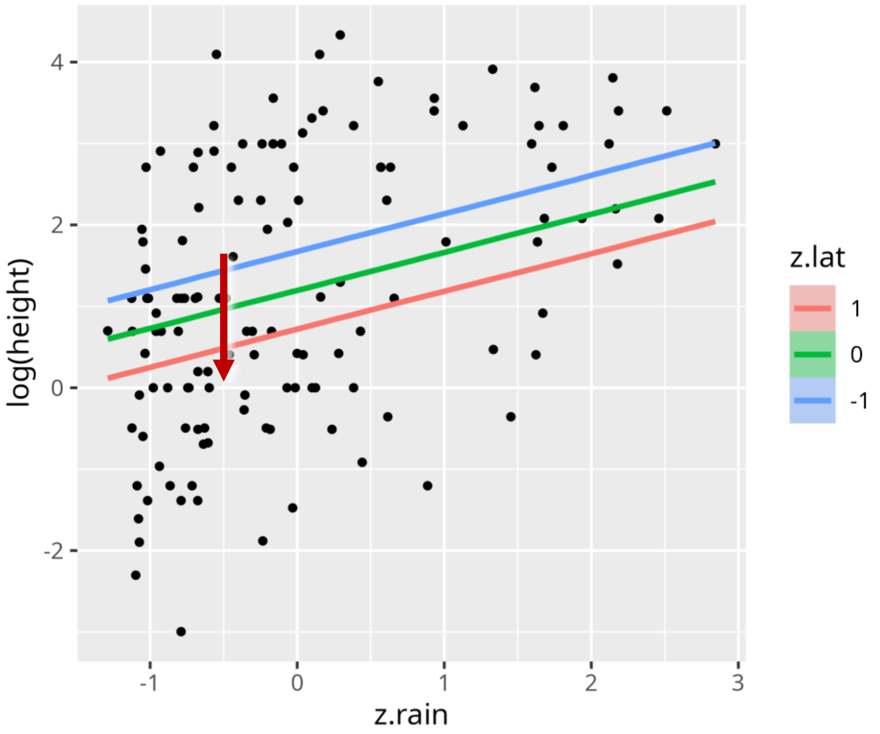
Now let‘s look at it from the perspective the 2nd predictor $rain$

$$\mu = (b_0 + \boldsymbol{b_1} \cdot lat) + b_2 \cdot rain$$

Intercept depends on $\boldsymbol{lat}$

Slope constant

1st variable shifts intercept by b_1



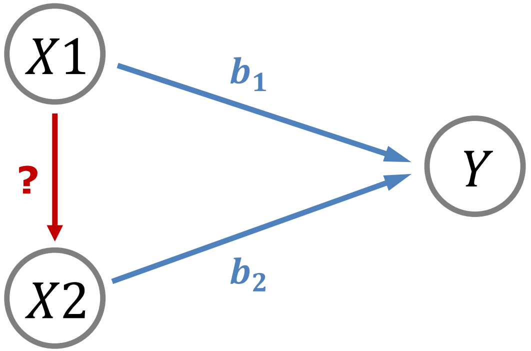
Multiple predictors: multicollinearity

Example: Latitudinal gradient of plant size

Stochastic part: $\log(height) \sim \text{Normal}(\mu, \sigma)$

Deterministic part: $\mu = b_0 + b_1 \cdot lat + b_2 \cdot rain$

What if predictor variables are correlated ?
Here lat influences $rain$!



- A bit of multicollinearity is OK.
- But be aware of interpretation of effects!
- $\boldsymbol{b_1}$ is effect $\boldsymbol{x_1} \rightarrow \boldsymbol{y}$, while $\boldsymbol{x_2}$ held constant!
- Slopes describe direct (isolated) effects only, not total effect
- Often the problem when dealing with observational data instead controlled experiments.

→ Cinelli, Forney & Pearl (2024). A crash course in good and bad controls

Multiple predictors & interaction

Example: Latitudinal gradient of plant size

$$\log(height) \sim \text{Normal}(\mu, \sigma)$$
$$\mu = b_0 + b_1 \cdot lat + b_2 \cdot rain + \boldsymbol{b_3 \cdot lat \cdot rain}$$

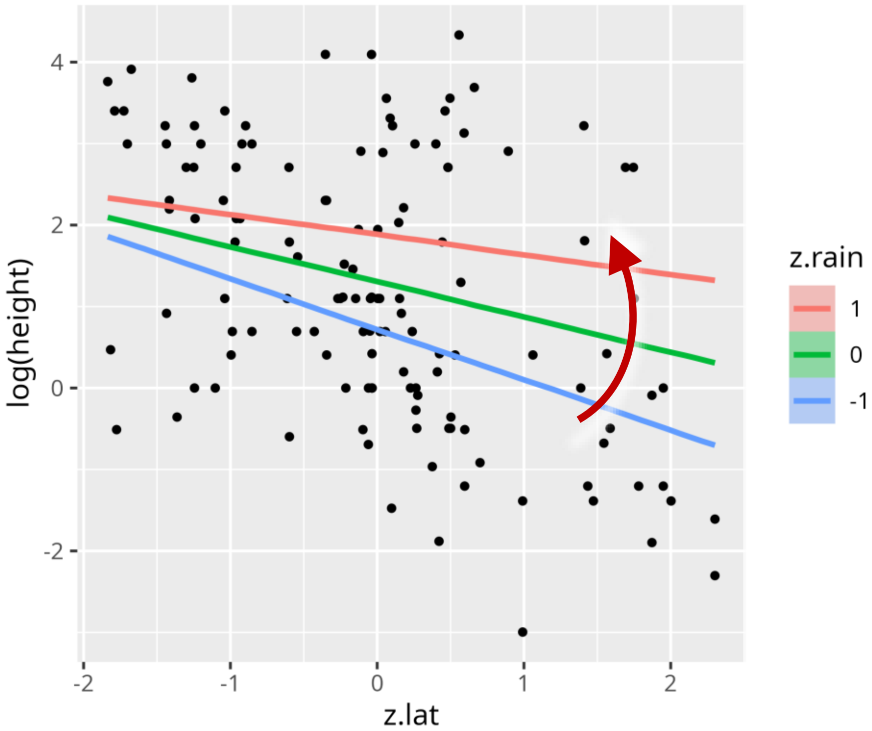
```
> brm(log(height)~z.lat*z.rain)

 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: log(height) ~ z.lat * z.rain
   Data: globalPlants (Number of observations: 131)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000

Regression Coefficients:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept        1.30      0.15     1.00     1.59 1.00     4804     2827
z.lat           -0.43      0.16    -0.74    -0.11 1.00     3606     3169
z.rain           0.58      0.18     0.23     0.95 1.00     3153     3169
z.lat:z.rain     0.19      0.14    -0.09     0.47 1.00     4010     2982

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     1.45      0.09     1.28     1.64 1.00     4778     2801

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```



Multiple predictors & interaction

Example: Latitudinal gradient of plant size

$$\log(height) \sim \text{Normal}(\mu, \sigma)$$
$$\mu = b_0 + b_1 \cdot lat + b_2 \cdot rain + \boldsymbol{b_3 \cdot lat \cdot rain}$$

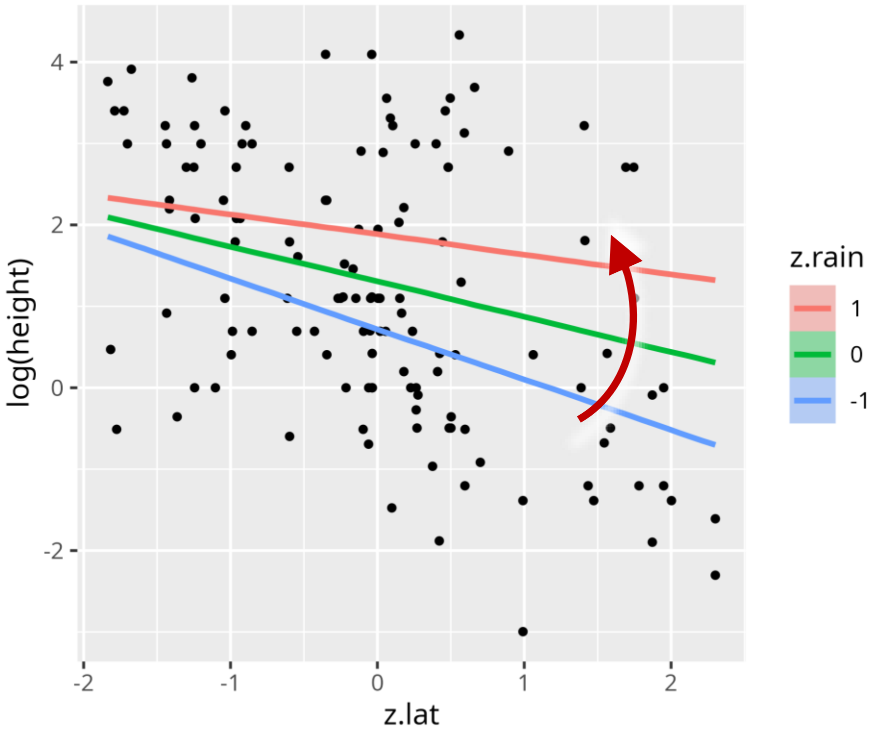
$$\mu = (b_0 + b_2 \cdot rain) + (b_1 + \boldsymbol{b_3} \cdot rain) \cdot lat$$

Intercept depends on ***rain***

Slope also depends on ***rain***

2nd variable shifts intercept by b_2

shifts slope by b_3



Multiple predictors & interaction

Now let‘s look at it from the perspective the 2nd predictor $rain$

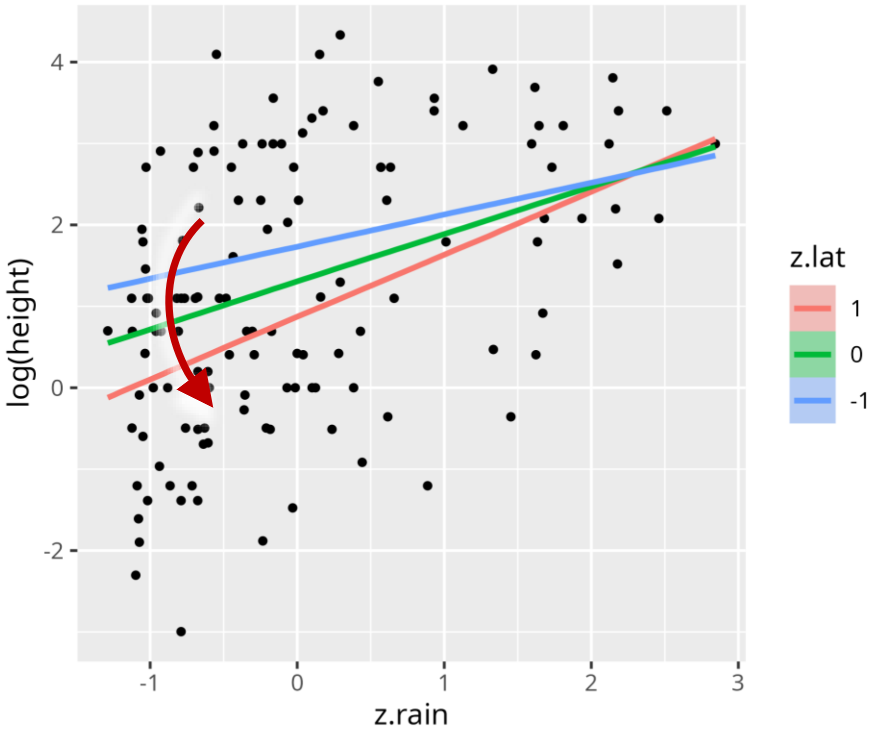
$$\mu = (b_0 + b_1 \cdot lat) + (b_2 + \boldsymbol{b_3} \cdot lat) \cdot rain$$

Intercept depends on ***lat***

Slope also depends on ***lat***

2nd variable shifts intercept by b_2

shifts slope by b_3



Multiple predictors & interaction

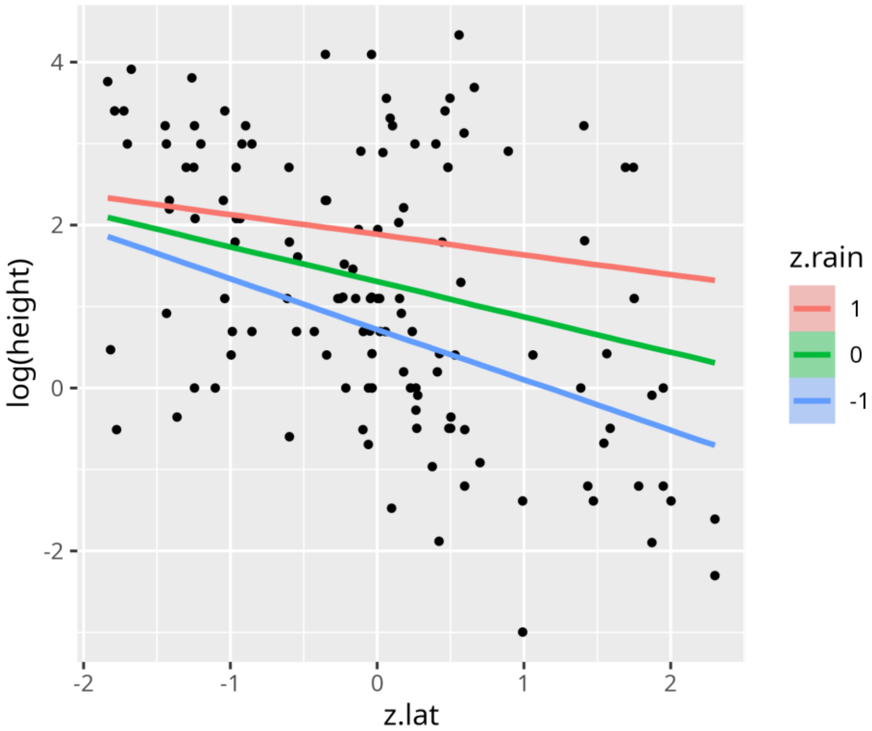
$\log(height) \sim \text{Normal}(\mu, \sigma)$

$\mu = b_0 + b_1 \cdot lat + b_2 \cdot rain + b_3 \cdot lat \cdot rain$

```
Regression Coefficients:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept        1.30      0.15     1.00     1.59 1.00     4804     2827
z.lat           -0.43      0.16    -0.74    -0.11 1.00     3606     3169
z.rain           0.58      0.18     0.23     0.95 1.00     3153     3169
z.lat:z.rain     0.19      0.14    -0.09     0.47 1.00     4010     2982
```

„Main effects" describe slope of a predictor, when other predictors = 0!

Much simpler interpretation when using **scaled** variables $z.lat$ and $z.rain$



Bayesian stats & linear regression

- Simple solutions for violation of model assumptions
  - Outliers → student-t distribution for residuals (heavier tails)
  - Non-constant residual sdev? → distributional models $\sigma(x) = \sigma_0 + \sigma_1 x$
  - Spatially / temporally autocorrelated residuals
- Simple comparison of intercepts & slopes („post-hoc analysis")
- Regularization of effect sizes with priors
- Unbiased estimates even for small datasets
- Multivariate extensions (fit multiple responses at once)
- …

Model selection

Frequentist F-tests

F-tests for (nested) linear models

Compare sums-of-squares of residuals

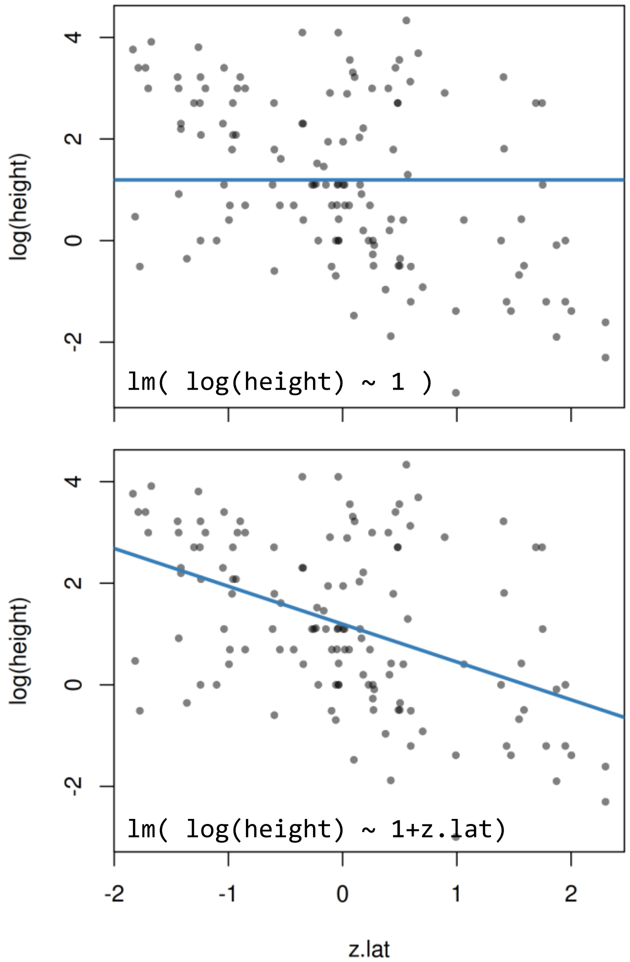
→ Connected to $\mathbf{R^2}$ values (amount of explained variation)

$\mathbf{R^2}$ always increases when adding predictors

H0: Both models perform equally

F-test checks if increase is „significant" or just random

P<0.05 → reject H0 and accept more complex model



Frequentist F-tests

```
lm( log(height) ~ z.lat*z.rain )
```

```
> summary( lm1 )
Call:
lm(formula = log(height) ~ z.lat * z.rain, data = globalPlants)

Residuals:
    Min      1Q  Median      3Q     Max
-3.2619 -0.9048  0.0017  1.0176  3.0977

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.3031     0.1490   8.745 1.14e-14 ***
z.lat         -0.4298     0.1581  -2.719  0.00746 **
z.rain         0.5855     0.1790   3.272  0.00138 **
z.lat:z.rain   0.1898     0.1411   1.345  0.18107
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.435 on 127 degrees of freedom
Multiple R-squared:  0.2653,    Adjusted R-squared:  0.2479
F-statistic: 15.29 on 3 and 127 DF,  p-value: 1.5e-08
```

Don‘t use p-values of main effects when there are higher-order effects (here interaction) → **Use anova-table**

~ z.lat*z.rain **versus** ~ 1

→ Always tests full model against intercept-only

Frequentist F-tests

```
lm( log(height) ~ z.lat*z.rain )
```

```
> summary( lm1 )
Call:
lm(formula = log(height) ~ z.lat * z.rain, data = globalPlants)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.2619 -0.9048  0.0017  1.0176  3.0977 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)    1.3031     0.1490   8.745 1.14e-14 ***
z.lat         -0.4298     0.1581  -2.719  0.00746 ** 
z.rain         0.5855     0.1790   3.272  0.00138 ** 
z.lat:z.rain   0.1898     0.1411   1.345  0.18107    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.435 on 127 degrees of freedom
Multiple R-squared:  0.2653,	Adjusted R-squared:  0.2479 
F-statistic: 15.29 on 3 and 127 DF,  p-value: 1.5e-08
```

~ z.lat*z.rain **versus** ~ 1

→ Always tests full model against intercept-only

```
> anova( lm1 )
Analysis of Variance Table

Response: log(height)
              Df  Sum Sq Mean Sq F value   Pr(>F)    
1: z.lat         1  72.083  72.083 35.0077 2.86e-08 ***
2: z.rain        1  18.612  18.612  9.0388 0.003186 ** 
3: z.lat:z.rain  1   3.724   3.724  1.8086 0.181073    
Residuals    127 261.502   2.059                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

1: ~ z.lat **versus** ~ 1

2: ~ z.lat+z.rain **versus** ~ z.lat

3: ~ z.lat*z.rain **versus** ~ z.lat+z.rain

→ Incremently tests more complex models

Frequentist AIC

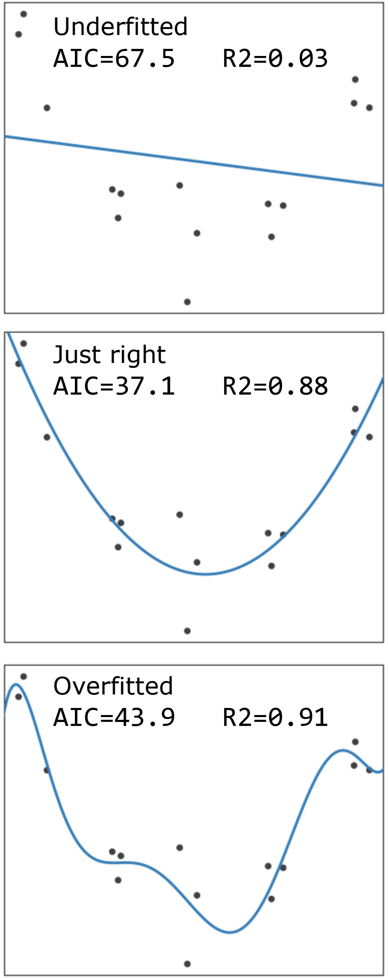
„Akaike information criterion" more flexible than F-tests

$$\text{AIC} = -2 \cdot \log(L) + 2 \cdot k$$

- Computed from likelihood L (remember: *maximum* likelihood)
  Model with higher likelihood-value fits the data better
- Adds a penalty term for model complexity k (number of parameters)

  → Model with lower AIC is better

„Principle of parsimony"



Bayesian LOO

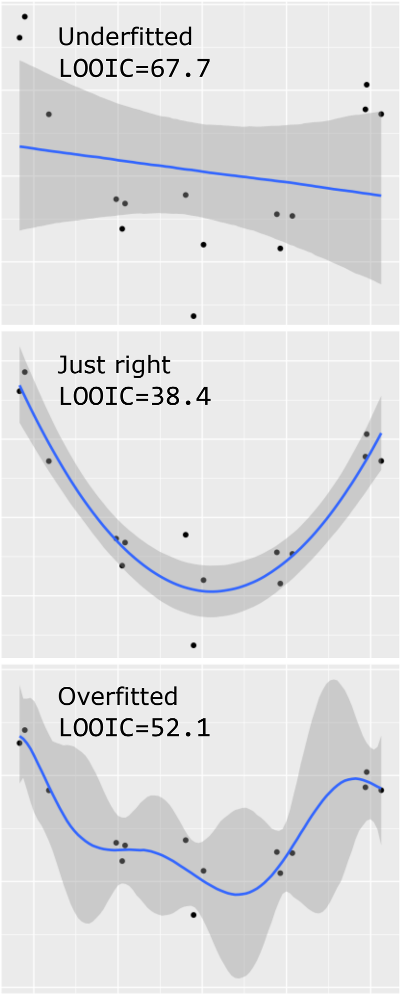
“Leave-one-out cross-validation”

elpd = expected log predictive density

- Computed from likelihood & posterior
- Includes parameter uncertainty & penalizes model complexity
- Estimates for how well the model would predict for a new dataset

  → Model with higher elpd = better

$\text{LOOIC} = -2 \cdot \text{elpd}$ (lower values = better) just for convenience for people used to AIC



Bayesian LOO

```
fit2 = brm(log(height)~z.lat+z.rain)
fit3 = brm(log(height)~z.lat*z.rain)
```

```
> LOO(fit3)

Computed from 4000 by 131 log-likelihood matrix.

         Estimate   SE
elpd_loo   -235.9  6.9
p_loo         3.5  0.4
looic       471.8 13.9
------
MCSE of elpd_loo is 0.0.
MCSE and ESS estimates assume MCMC draws (r_eff in [0.7, 1.2]).

All Pareto k estimates are good (k < 0.7).
See help('pareto-k-diagnostic') for details.
```

elpd: larger values are better

p: effective number of parameters

looic=-2*elpd

Estimates come with standard error

```
> LOO(fit2, fit3)

Model comparisons:
     elpd_diff se_diff
fit3  0.0       0.0
fit2 -3.5       2.4
```

Best elpd shown on top

Difference in elpd is associated with uncertainty

When **elpd_diff>2*se_diff** (approximately), you can be sure the model is better.

Here, both models perform equally under uncertainty, so we would choose the less complex one (fit2)

→ Use model comparison with LOO (similar to AIC) in the Bayesian framework

1 Categorical predictor

1 predictor with 2 levels

Example: bird species richness vs. landscape type

Observed bird species richness in different habitats

Each habitat categorized by landscape type:
- Agriculture
- Urban
- Bauxite
- Forest

Start with subset Agriculture / Urban first

Later, we also include area as a predictor

1 predictor with 2 levels

Example: bird species richness vs. landscape type

Stochastic part: $S \sim \text{Normal}(\mu, \sigma)$

Deterministic part: $\mu = \mu(landscape)$

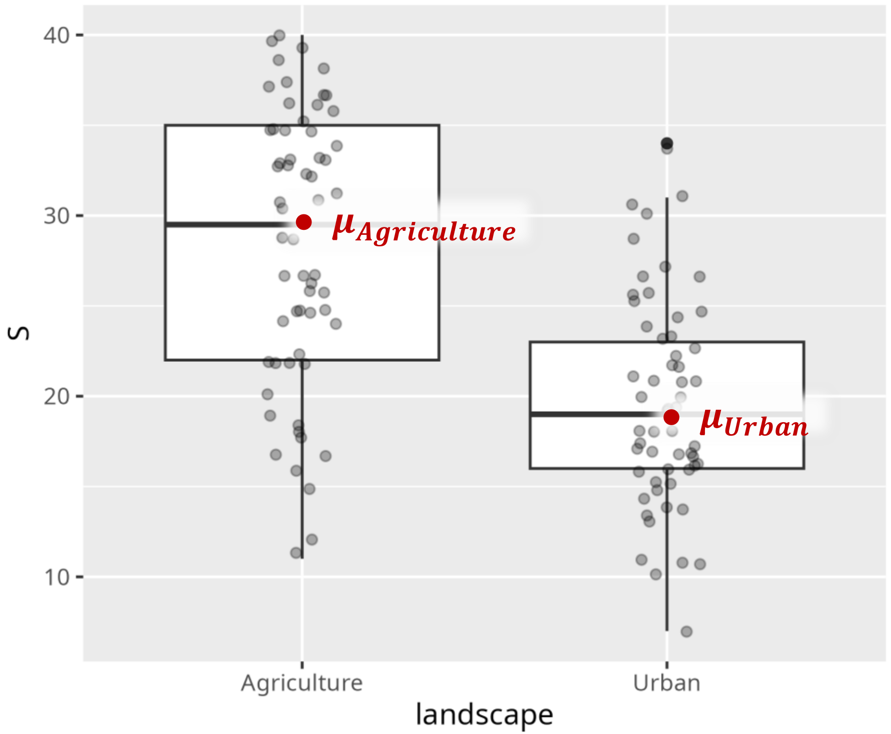
Each datapoint is a landscape patch

Categorical predictor: $landscape$ (2 levels)

3 parameters: $\mu_{Agriculture}, \ \mu_{Urban}, \ \sigma$

Estimate and compare group-level means

Frequentist method: t-test
Does not compare distributions (overlap).
Compares their **means** !



Dummy coding

Deterministic part: $\mu = b_0 + b_1 \cdot x_{Urban}$

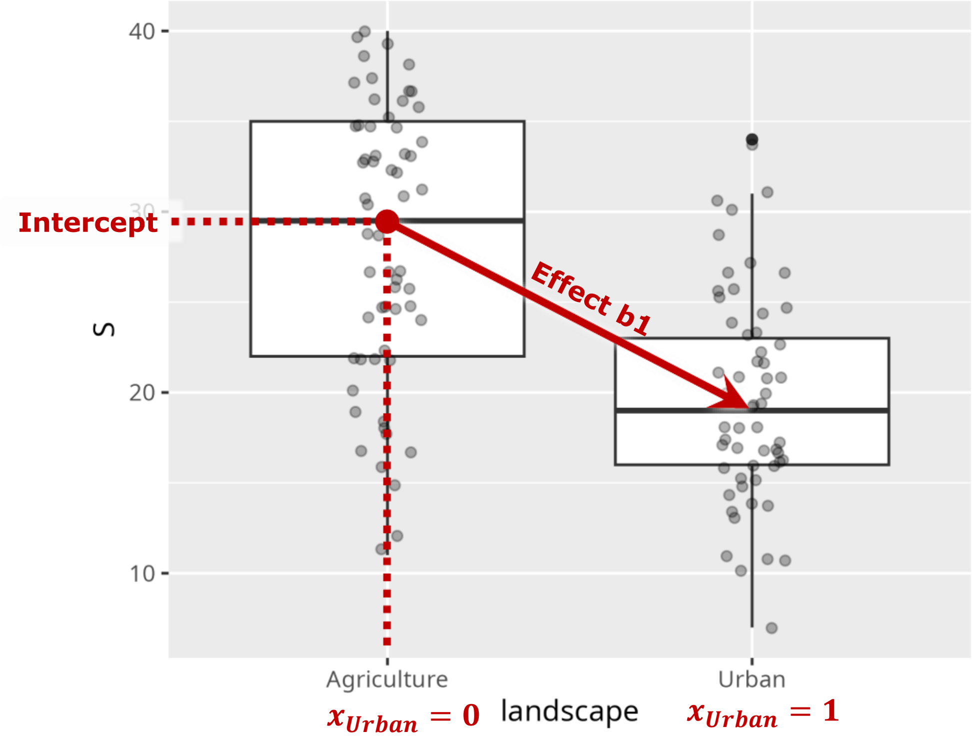
$landscape = Agriculture$ is reference level

$$x_{Urban} = \begin{cases} 1 & landscape = Urban \\ 0 & \text{otherwise} \end{cases}$$

$$\mu_{Agriculture} = b_0 + b_1 \cdot 0 = b_0$$

$$\mu_{Urban} = b_0 + b_1 \cdot 1 = b_0 + b_1$$

→ Linear model with „intercept" b_0 & „effect" b_1



Model fitting

brms uses **dummy-coding** (effect-coding) as default

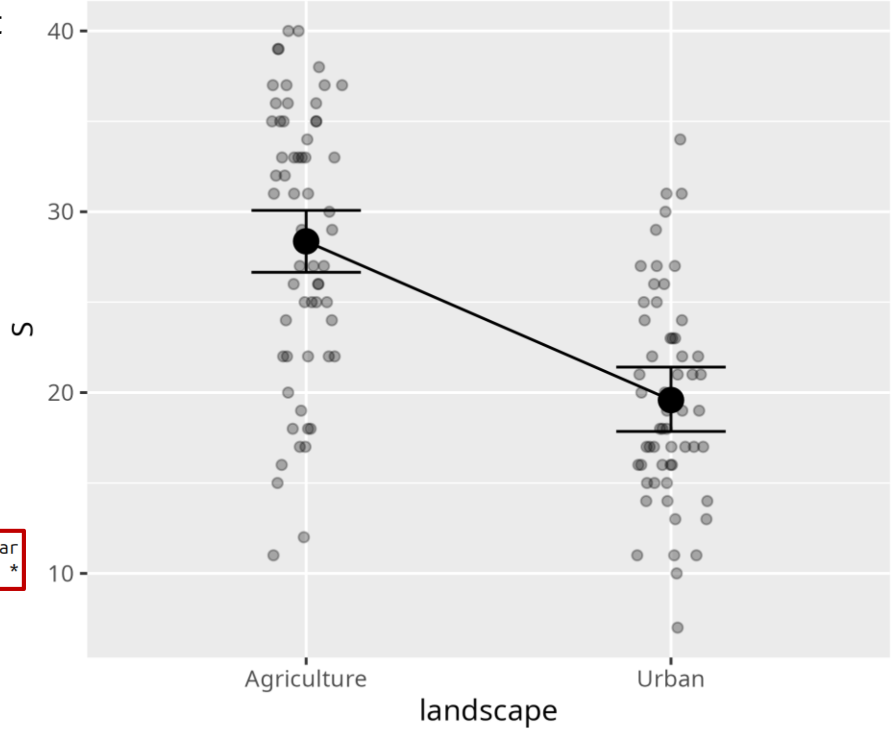
```
> brm( S ~ landscape )
```

```
Regression Coefficients:
               Estimate Est.Error l-95% CI u-95% CI
Intercept         28.37      0.88    26.65    30.07
landscapeUrban    -8.77      1.31   -11.28    -6.14
```

Q: Is there a difference in means?
→ Look at posterior distribution of effect

```
> hypothesis(fit_a1, "landscapeUrban<0")
Hypothesis Tests for class b:
          Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Star
1 (landscapeUrban) < 0    -8.77      1.31   -10.89    -6.59        Inf         1    *
```

$$P(\mu_{Urban} < \mu_{Agriculture}) = 1$$



Model fitting

But we could enforce **mean-coding**

```
> brm( S ~ landscape-1 )
```

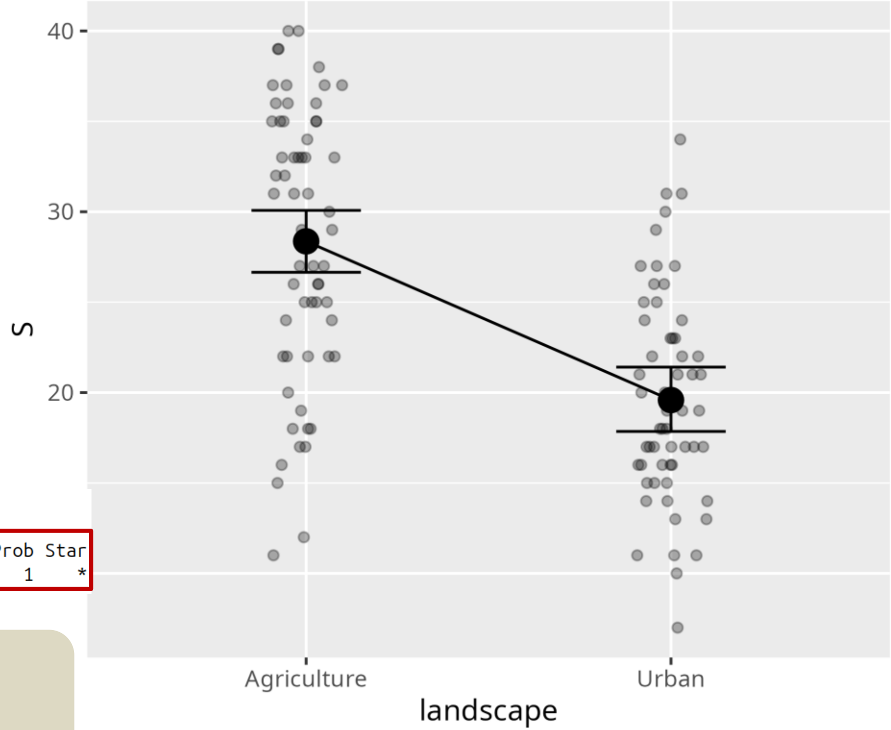
```
Regression Coefficients:
                     Estimate Est.Error l-95% CI u-95% CI
landscapeAgriculture    28.38      0.88    26.60    30.14
landscapeUrban          19.61      0.90    17.86    21.42
```

Q: Is there a difference in means?
→ Look at posterior distribution of difference

```
> hypothesis(fit_a1_1, "landscapeUrban<landscapeAgriculture")
Hypothesis Tests for class b:
                Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Star
1 (landscapeUrban)-... < 0    -8.77      1.26   -10.86    -6.72        Inf         1    *
```

$$P(\mu_{Urban} < \mu_{Agriculture}) = 1$$

dummy-coding
and mean-coding
are the same model !



1 predictor with *K* levels

Example: bird species richness vs. landscape type

Stochastic part: $S \sim \text{Normal}(\mu, \sigma)$

Deterministic part: $\mu = \mu(landscape)$

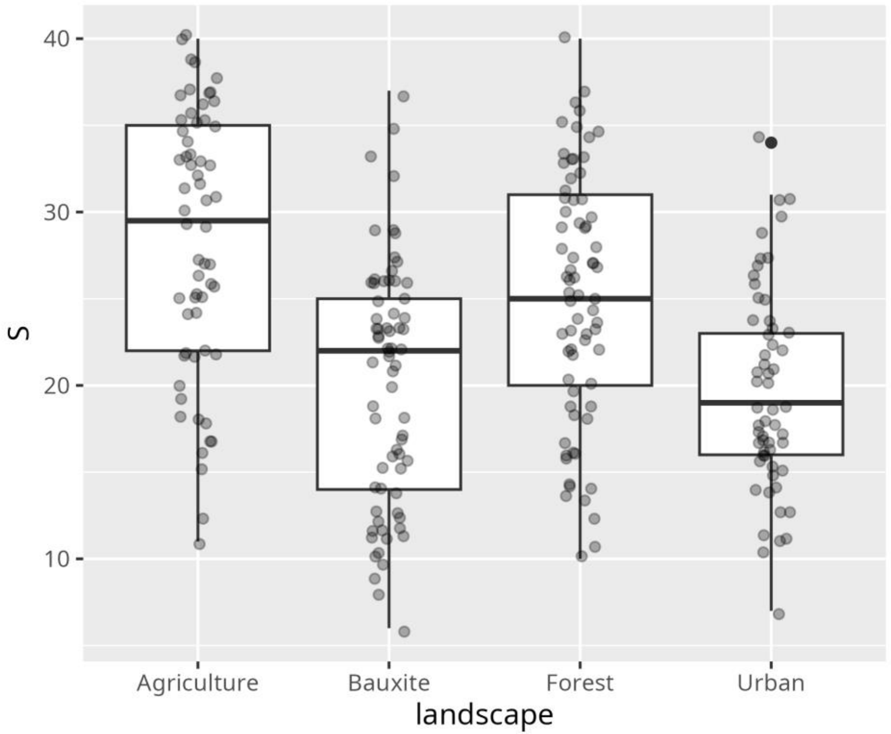
Each datapoint is a landscape patch

Categorical predictor: $landscape$ (4 levels)

K+1 parameters: $\mu_{Agriculture}$, $\mu_{Bauxite}$, μ_{Forest}, μ_{Urban}, σ

Estimate and compare group-level means

Frequentist method: F-test (ANOVA)
Test model against intercept-only model



Dummy coding with *K* levels

$\mu = b_0 + b_1 \cdot x_{Bauxite} + b_2 \cdot x_{Forest} \; b_3 \cdot x_{Urban}$

$landscape = Agriculture$ is reference level

K-1 dummy variables:

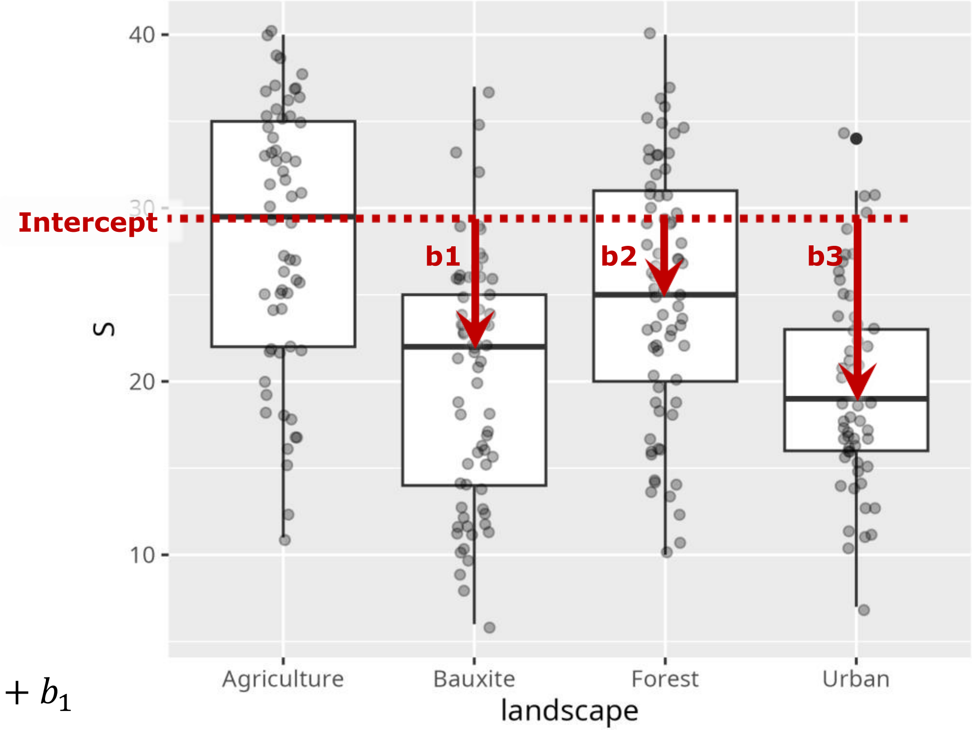
$$x_{Bauxite} = \begin{cases} 1 & landscape = Bauxite \\ 0 & \text{otherwise} \end{cases}$$

$$x_{Forest} = \begin{cases} 1 & landscape = Forest \\ 0 & \text{otherwise} \end{cases}$$

$$x_{Urban} = \begin{cases} 1 & landscape = Urban \\ 0 & \text{otherwise} \end{cases}$$

$$\mu_{Agriculture} = b_0 + b_1 \cdot 0 + b_2 \cdot 0 + b_3 \cdot 0 = b_0$$
$$\mu_{Bauxite} = b_0 + b_1 \cdot 1 + b_2 \cdot 0 + b_3 \cdot 0 = b_0 + b_1$$

etc …



Model fitting with *K* levels

```
> brm( S ~ landscape )

Regression Coefficients:
                   Estimate Est.Error l-95% CI u-95% CI
Intercept             28.38      0.88    26.58    30.05
landscapeBauxite      -8.33      1.22   -10.63    -5.84
landscapeForest       -3.35      1.21    -5.67    -0.97
landscapeUrban        -8.78      1.29   -11.31    -6.25
```

Q: Is there a difference in means?

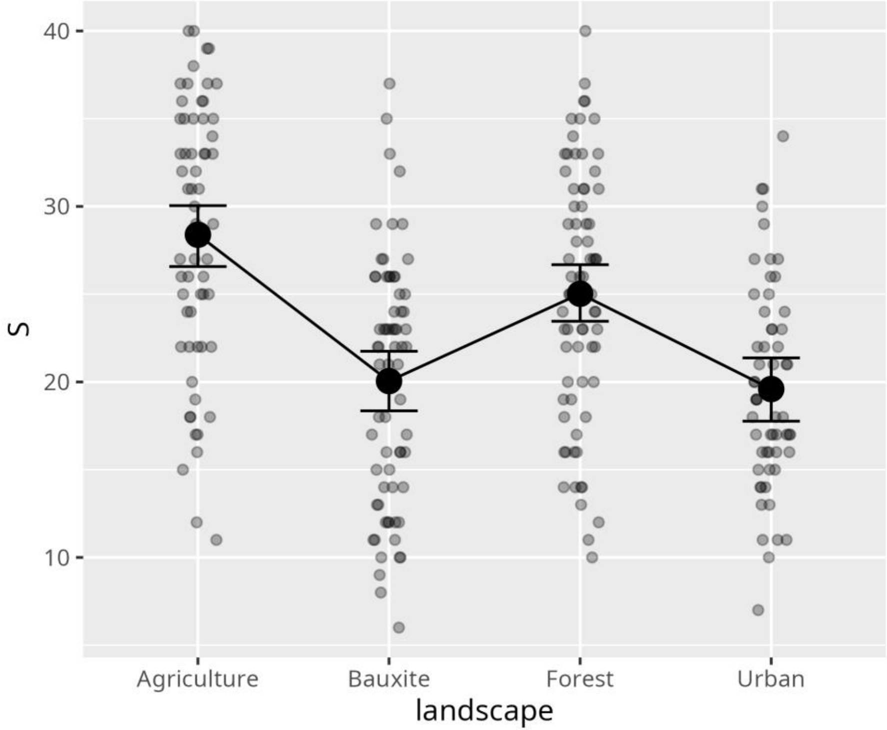
→ Individual effects don't give an overall answer

→ Compare against intercept-only model (similar to frequentist F-test)

```
> LOO(fit_landscape, fit_intercept)

Model comparisons:
              elpd_diff se_diff
fit_landscape   0.0       0.0
fit_intercept -27.1       7.4
```

Yes, **S~landscape** is a better model than **S~1**



2 Categorical predictors

2 predictors with *K* & *L* levels

Example: S vs. landscape type & area size

Additive model `S ~ landscape + area`

Accounts for difference in area size

Area effect is the same for all landscape levels

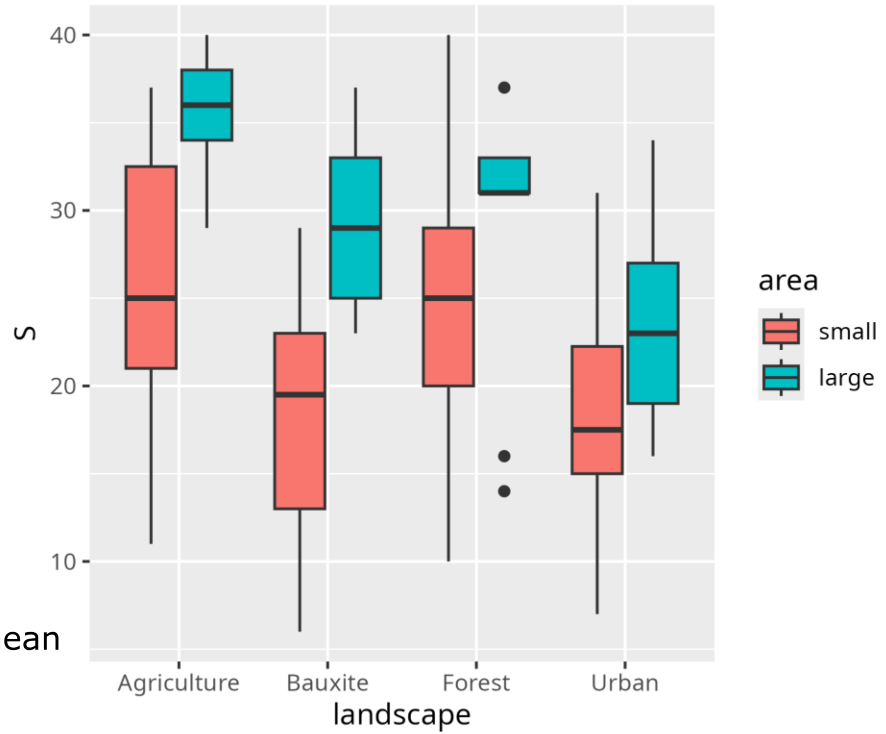
$K + L - 1$ parameters (+1 for sdev)

Factorial model `S ~ landscape * area`

Area effect changes over landscape levels

Each landscape:size combination is fitted with own mean

$K \cdot L$ parameters (+1 for sdev)



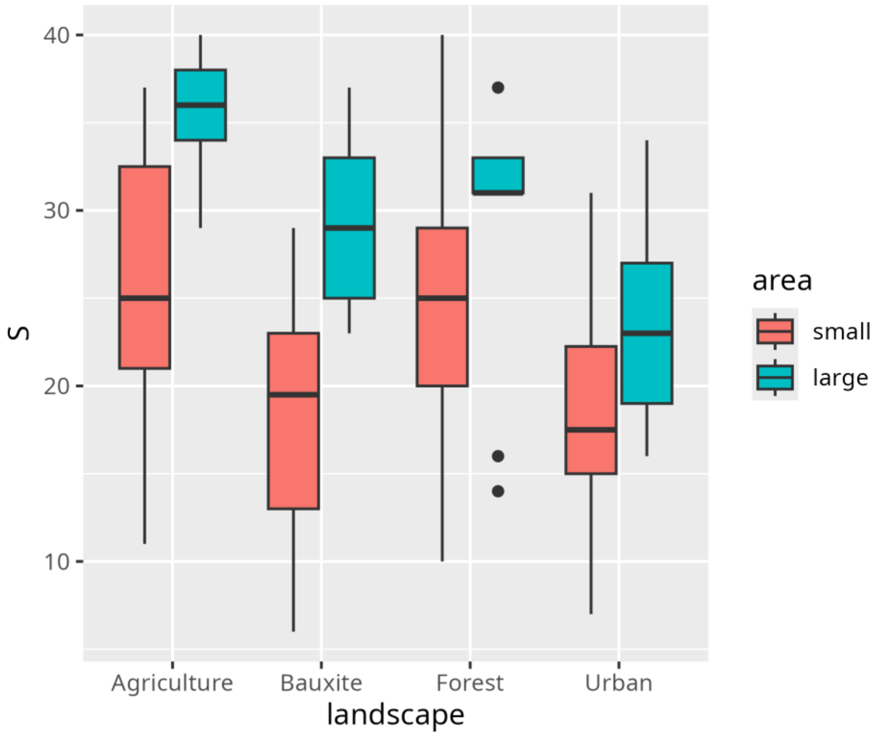
Additive model

$$\mu = b_0 +$$
$$b_1 \cdot x_{Bauxite} + b_2 \cdot x_{Forest}\ b_3 \cdot x_{Urban} +$$
$$\mathbf{b_4 \cdot x_{large}}$$

$landscape = Agriculture,\ area = small$ is reference level

1 intercept
K-1 dummy variables for landscape
L-1 dummy variables for area

=K+L-1 variables
less than K*L level combinations
→ Will not fit independent group-level means



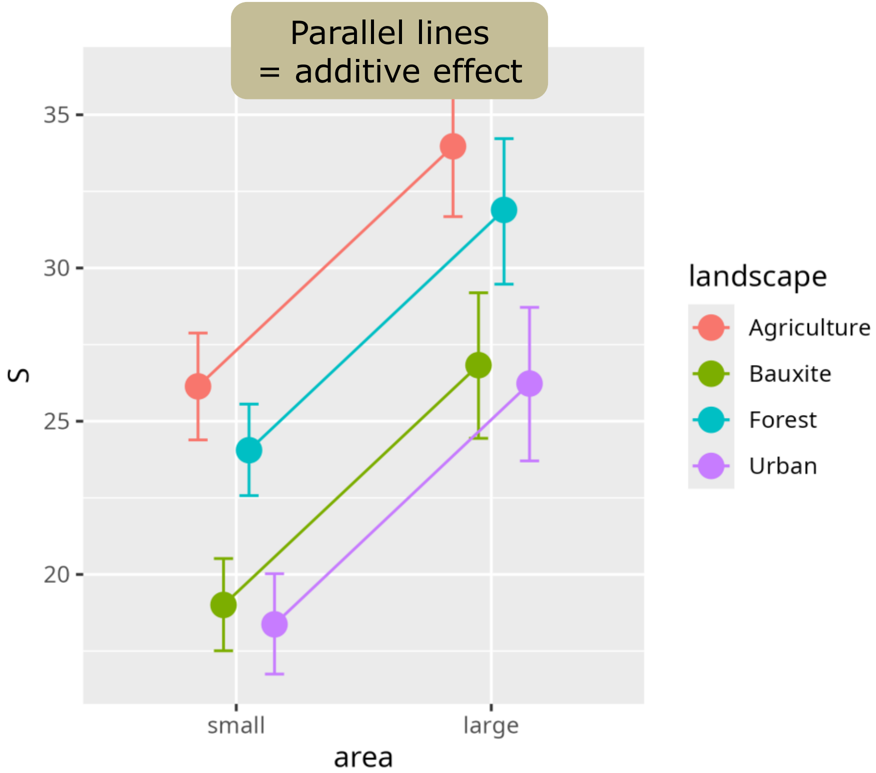
Additive model

```
> brm( S ~ landscape + area )
```

```
Regression Coefficients:
                 Estimate Est.Error l-95% CI u-95% CI
Intercept           26.14      0.89    24.39    27.88
landscapeBauxite    -7.13      1.15    -9.36    -4.84
landscapeForest     -2.07      1.14    -4.25     0.16
landscapeUrban      -7.75      1.20   -10.10    -5.46
arealarge            7.82      1.10     5.62     9.98
```

Q: Is there an additional effect of patch size?

→ Strong effect of area
7.82 more species in large patches (on avg.)



Additive model

```
> brm( S ~ landscape + area )
```

```
Regression Coefficients:
                  Estimate Est.Error l-95% CI u-95% CI
Intercept            26.[illegible]                    [illegible].88
landscapeBauxite     -7[illegible]                      [illegible].84
landscapeForest      -2[illegible]                      [illegible].16
landscapeUrban       -7[illegible]                      [illegible].46
arealarge             7.82      1.10     5.62     9.98
```

Summary table not helpful for predictors with >2 levels

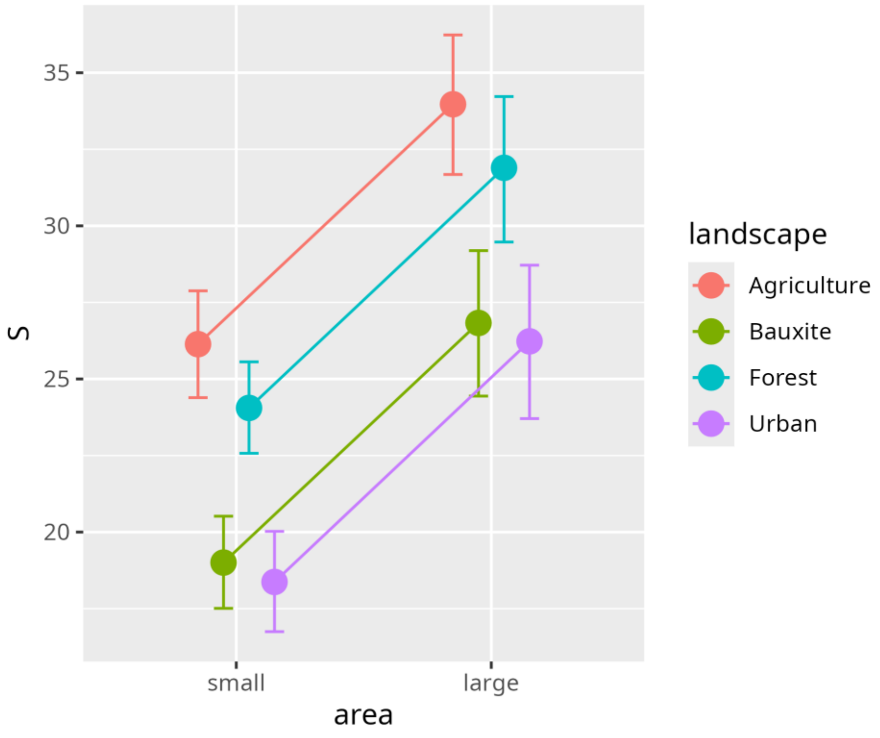
Q: Is there an additional effect of patch size?

→ Model comparison

```
> LOO(fit_additive, fit_landscape)
```

```
Model comparisons:
              elpd_diff se_diff
fit_additive    0.0       0.0
fit_landscape -23.7       6.9
```

Yes, **S~landscape+area** is a better model than **S~landscape**



Additive model

```
> brm( S ~ landscape + area )
```

Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	26.[illegible]	[illegible]	[illegible]	[illegible].88
landscapeBauxite	-7[illegible]	[illegible]	[illegible]	[illegible].84
landscapeForest	-2[illegible]	[illegible]	[illegible]	[illegible].16
landscapeUrban	-7[illegible]	[illegible]	[illegible]	[illegible].46
arealarge	7.82	1.10	5.62	9.98

Summary table not helpful for predictors with >2 levels

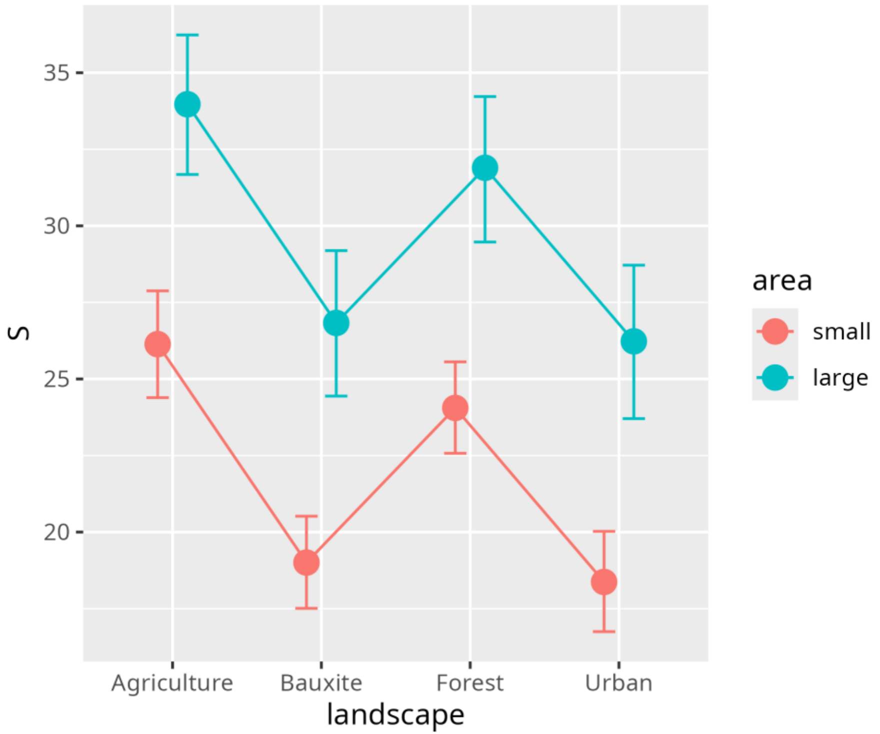
Q: Are there differences between landscapes, when controlling for area size?

```
> LOO(fit_additive, fit_area)
```

Model comparisons:

	elpd_diff	se_diff
fit_additive	0.0	0.0
fit_area	-26.9	7.1

Yes, **S~landscape+area** is a better model than **S~area**



Factorial model

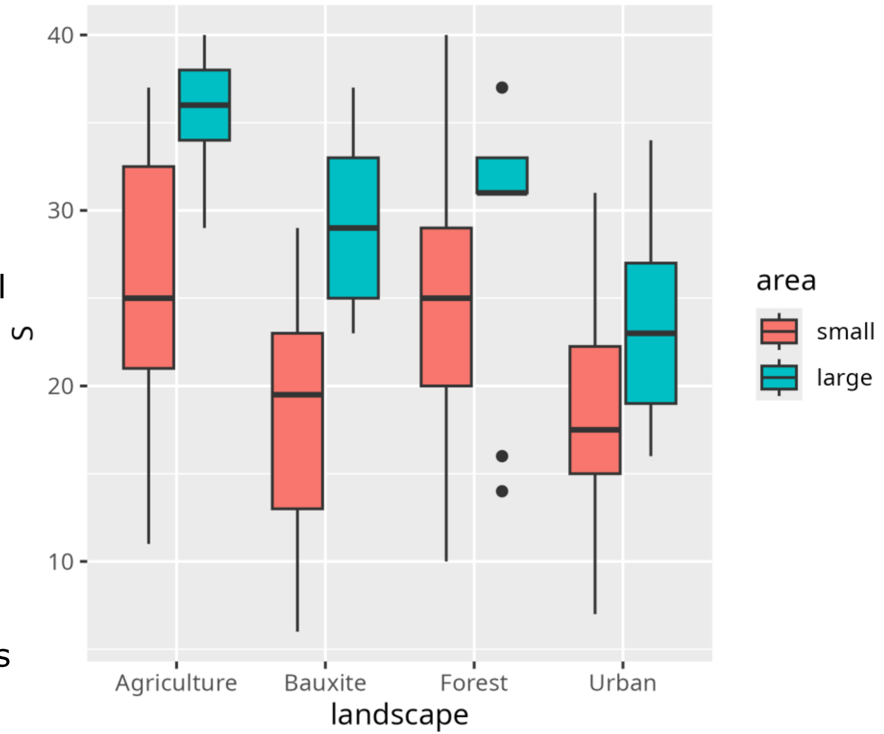
$$\mu = b_0 + b_1 \cdot x_{Bauxite} + b_2 \cdot x_{Forest}\ b_3 \cdot x_{Urban} + b_4 \cdot x_{large} + \boldsymbol{b_5 \cdot x_{Bauxite,large} + b_6 \cdot x_{Forest,large} + b_7 \cdot x_{Urban,large}}$$

$landscape = Agriculture,\ area = small$ is reference level

1 intercept
K-1 dummy variables for landscape
L-1 dummy variables for area
(K-1)*(L-1) dummy variables for landscape:area

= K*L variables in total
→ Fitting independent means to all level combinations

$$\mu = \mu(landscape, area)$$



Factorial model

```
> brm( S ~ landscape * area )
```

Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	25.56	0.97	23.63	27.51
landscapeBauxite	-6.93	1.27	-9.43	-4.47
landscapeForest	-1.06	1.24	-3.50	1.37
landscapeUrban	-6.78	1.35	-9.44	-4.08
arealarge	9.89	1.81	6.32	13.47
landscapeBauxite:arealarge	0.87	2.87	-4.56	6.46
landscapeForest:arealarge	-5.76	2.91	-11.69	-0.29
landscapeUrban:arealarge	-4.81	2.91	-10.52	0.97

Q: Does area effect change between landscape levels?

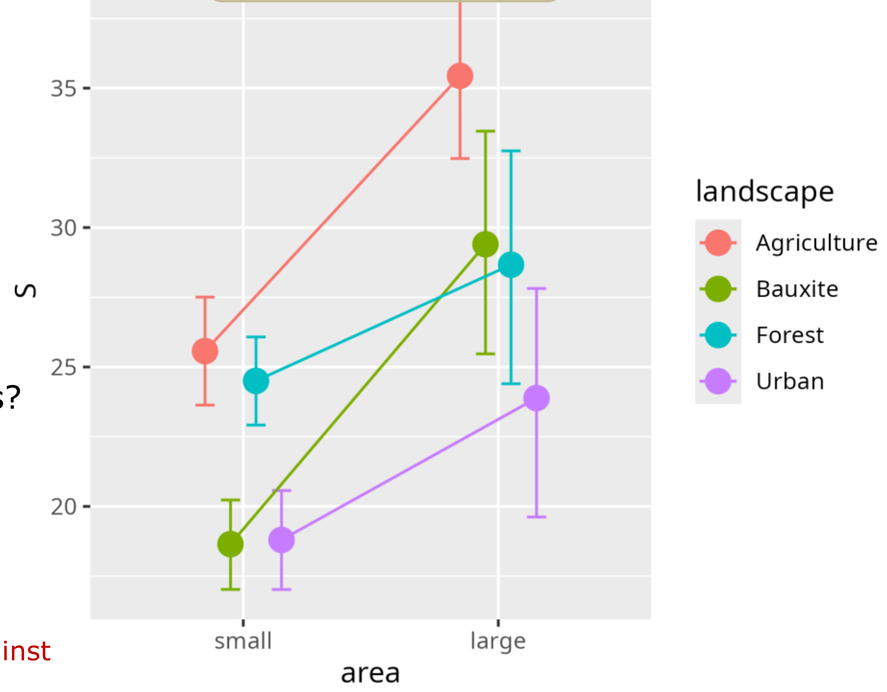
```
> LOO(fit_factorial, fit_additive)
```

Model comparisons:

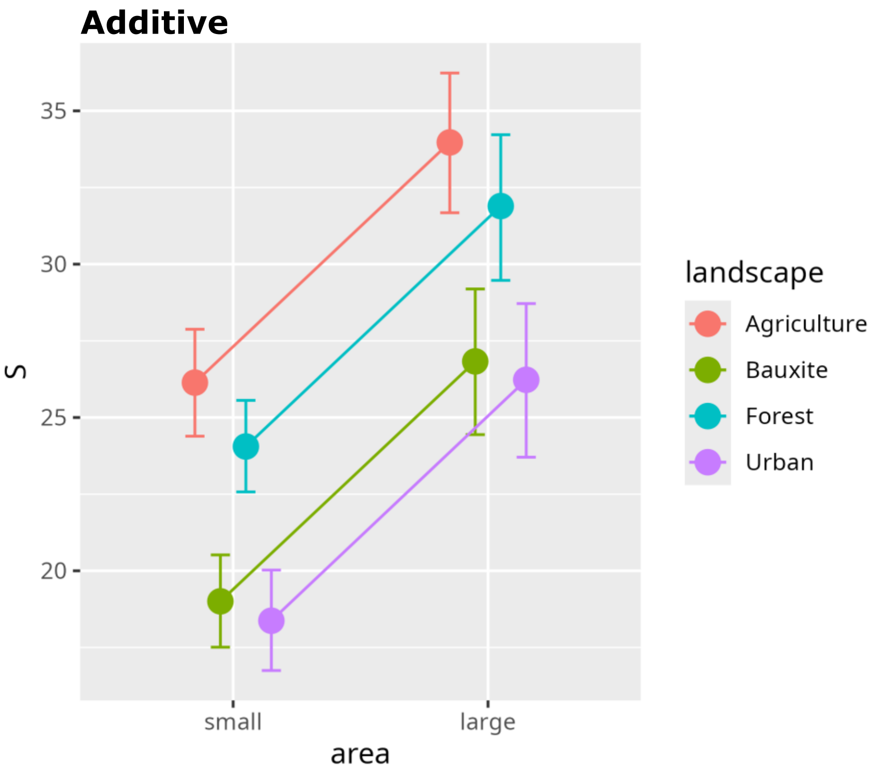
	elpd_diff	se_diff
fit_factorial	0.0	0.0
fit_additive	-0.7	2.5

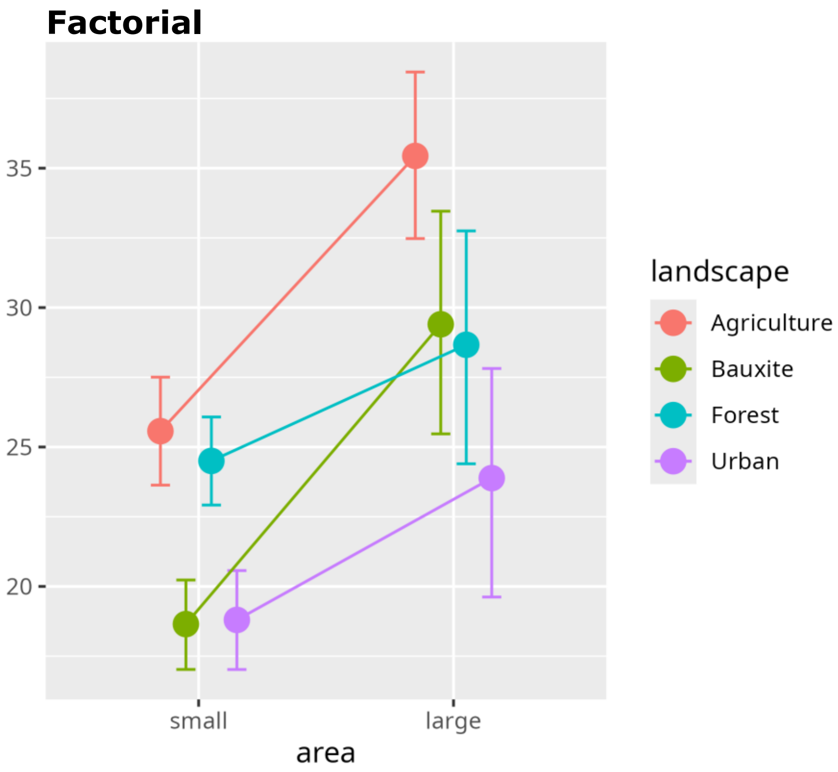
No strong evidence for **S~landscape*area** against **S~landscape+area**

Lines not parallel = interaction effect



Factorial vs additive





No strong evidence for interaction found → select additive as the best model

Post-hoc analysis

What is post-hoc analysis?

Model comparison (LOO)

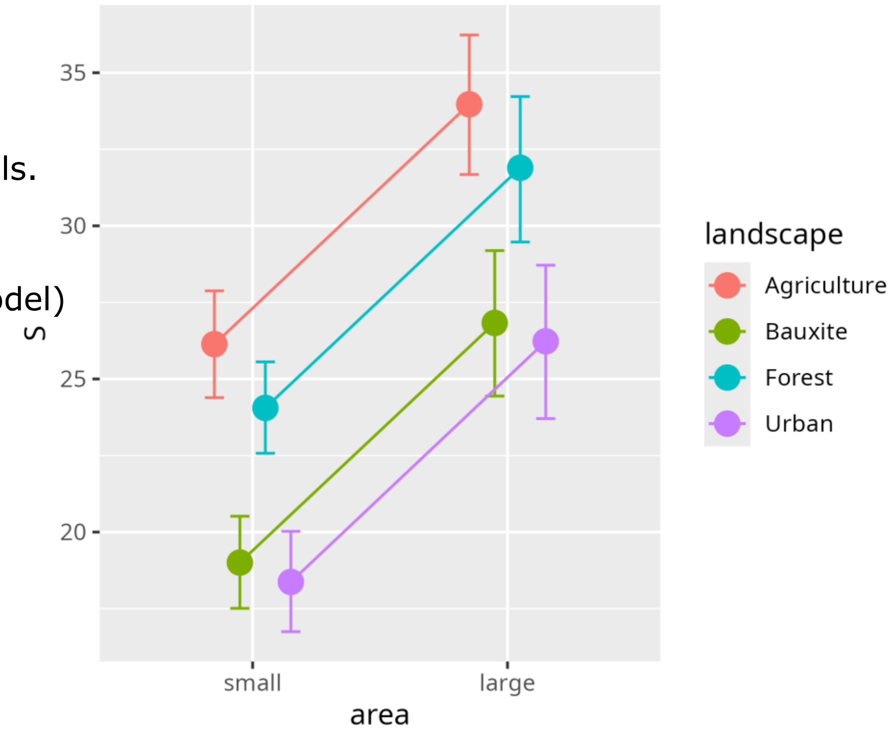
tells you **IF** there is a difference between group-levels.

Post-hoc analysis (**after** selecting an appropriate model)

tells you **WHAT** the difference is.

Analysis is **model-based**.

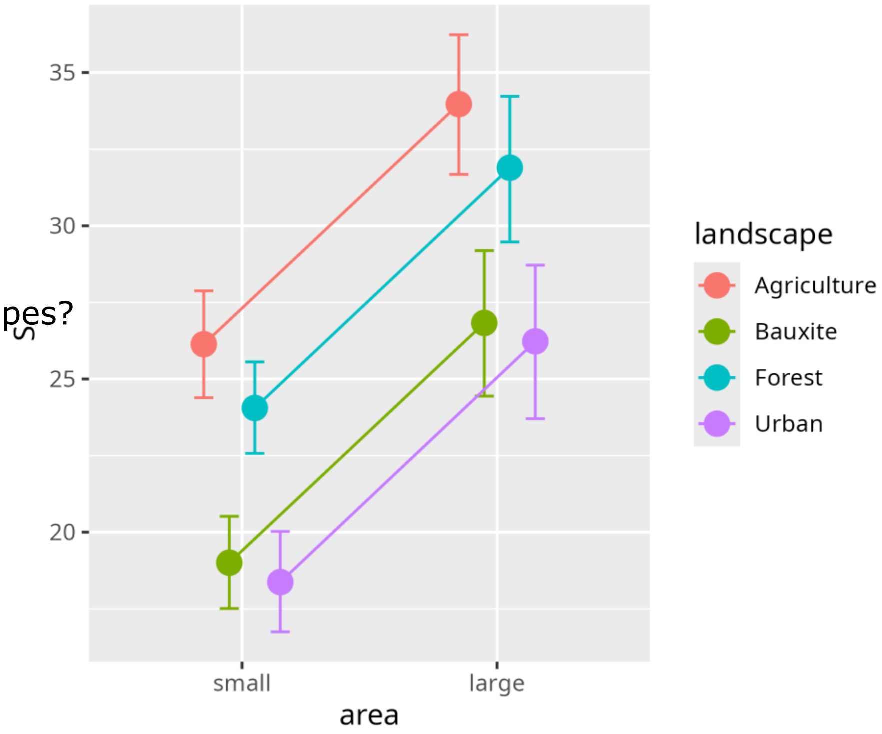
Do not just compute empirical means from the data.



What is post-hoc analysis?

Answer questions like:

- What is the mean species richness in small areas?
  → Average over landscapes.
- What is the mean species richness in urban landscapes?
  → Average over area sizes.
- What is the mean difference between urban and agricultural landscapes?
- And what are all their associated uncertainties?



Bayesian post-hoc analysis

Make predictions & compute their average or difference etc depending on the question.

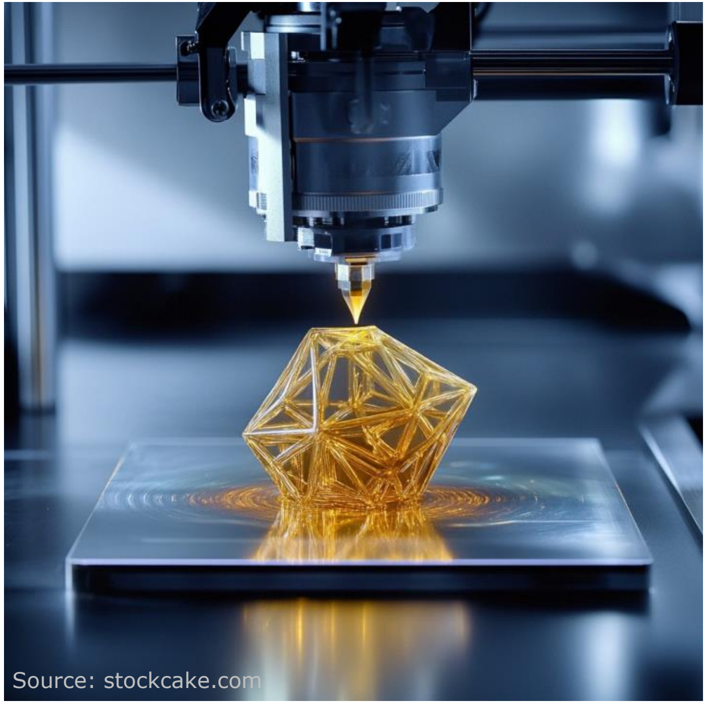
Remember: everything is a distribution!

- For each sample $\boldsymbol{k}$ of the posterior ($k = 1 \ldots 1000$) use it's predictions, compute what's required, e.g. $\boldsymbol{a_k} - \boldsymbol{b_k}$
- That is a sample of posterior distribution for $\boldsymbol{a} - \boldsymbol{b}$
- Compute mean, standard deviation, quantiles, etc

→ The **`emmeans`** package can automate these steps!

→ Alternative: **`marginaleffects`** package.
Powerful but a bit more complex

The Bayesian 3D printer



Post-hoc analysis: 1 predictor

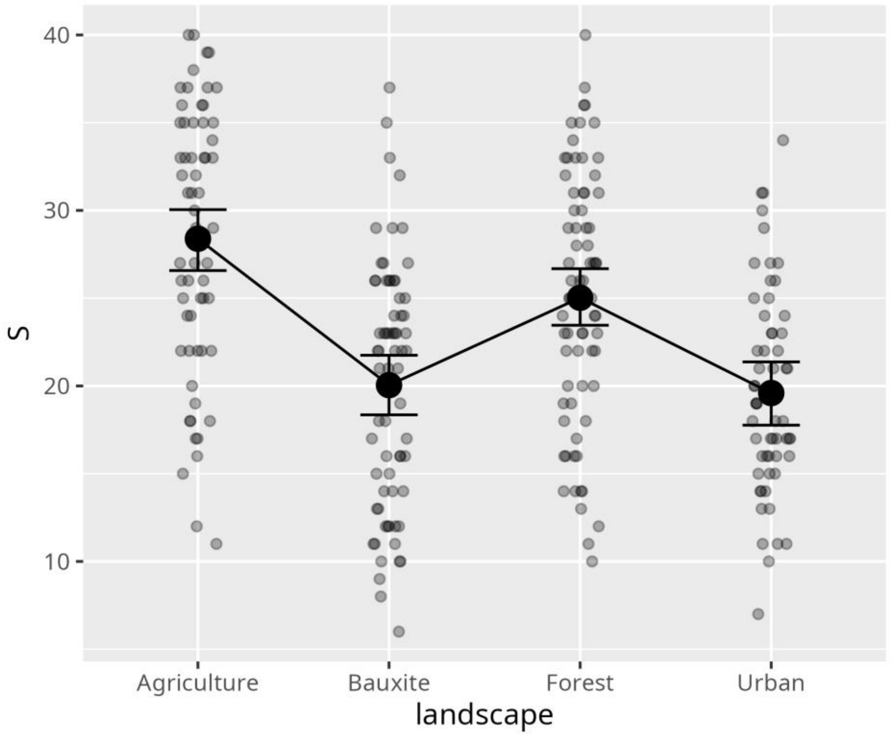
```
> brm( S ~ landscape )
```

```
Regression Coefficients:
                  Estimate Est.Error l-95% CI u-95% CI
Intercept            28.38      0.88    26.58    30.05
landscapeBauxite     -8.33      1.22   -10.63    -5.84
landscapeForest      -3.35      1.21    -5.67    -0.97
landscapeUrban       -8.78      1.29   -11.31    -6.25
```

Q: What are the predicted means (and their uncertainties)?

```
> emmeans(fit_landscape, "landscape")
 landscape   emmean lower.HPD upper.HPD
 Agriculture   28.4      26.6      30.0
 Bauxite       20.0      18.5      21.8
 Forest        25.0      23.4      26.6
 Urban         19.6      17.8      21.4

Point estimate displayed: median
HPD interval probability: 0.95
```



Post-hoc analysis: 1 predictor

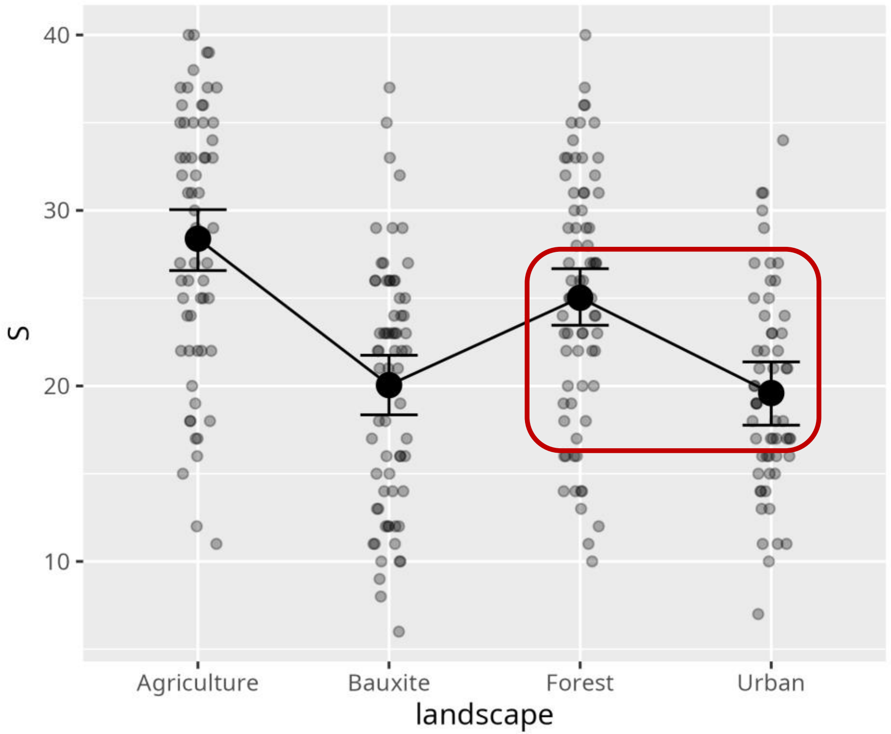
```
> brm( S ~ landscape )

Regression Coefficients:
                  Estimate Est.Error l-95% CI u-95% CI
Intercept            28.38      0.88    26.58    30.05
landscapeBauxite     -8.33      1.22   -10.63    -5.84
landscapeForest      -3.35      1.21    -5.67    -0.97
landscapeUrban       -8.78      1.29   -11.31    -6.25
```

Q: What is the difference between forest & urban?

```
> emmeans(fit_landscape, "landscape") |> pairs()
 contrast               estimate lower.HPD upper.HPD
 Agriculture - Bauxite      8.35     5.857     10.56
 Agriculture - Forest       3.30     0.966      5.63
 Agriculture - Urban        8.73     6.290     11.24
 Bauxite - Forest          -5.00    -7.217     -2.60
 Bauxite - Urban            0.45    -2.232      2.86
 Forest - Urban             5.45     2.940      7.83

Point estimate displayed: median
HPD interval probability: 0.95
```



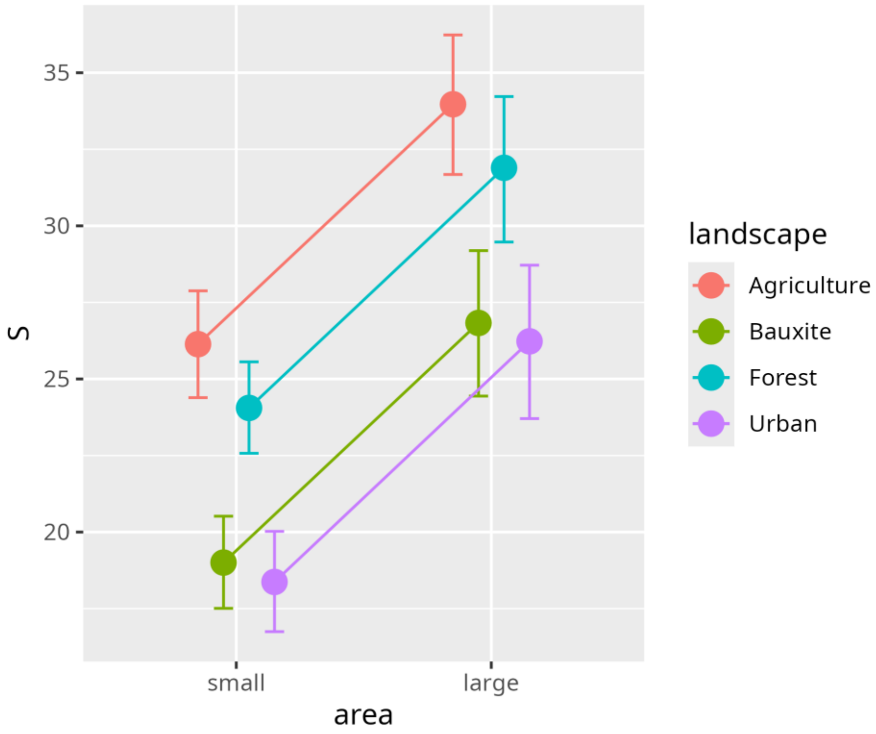
Post-hoc analysis: multiple predictors

```
> brm( S ~ landscape + area )
```

Q: What are group-level means?

```
> emmeans(fit_additive, ~area*landscape)
 area  landscape   emmean lower.HPD upper.HPD
 small Agriculture   26.2      24.4      27.9
 large Agriculture   33.9      31.7      36.1
 small Bauxite       19.0      17.6      20.6
 large Bauxite       26.8      24.6      29.2
 small Forest        24.1      22.6      25.6
 large Forest        31.9      29.5      34.3
 small Urban         18.4      16.8      20.2
 large Urban         26.2      23.8      28.5

Point estimate displayed: median
HPD interval probability: 0.95
```



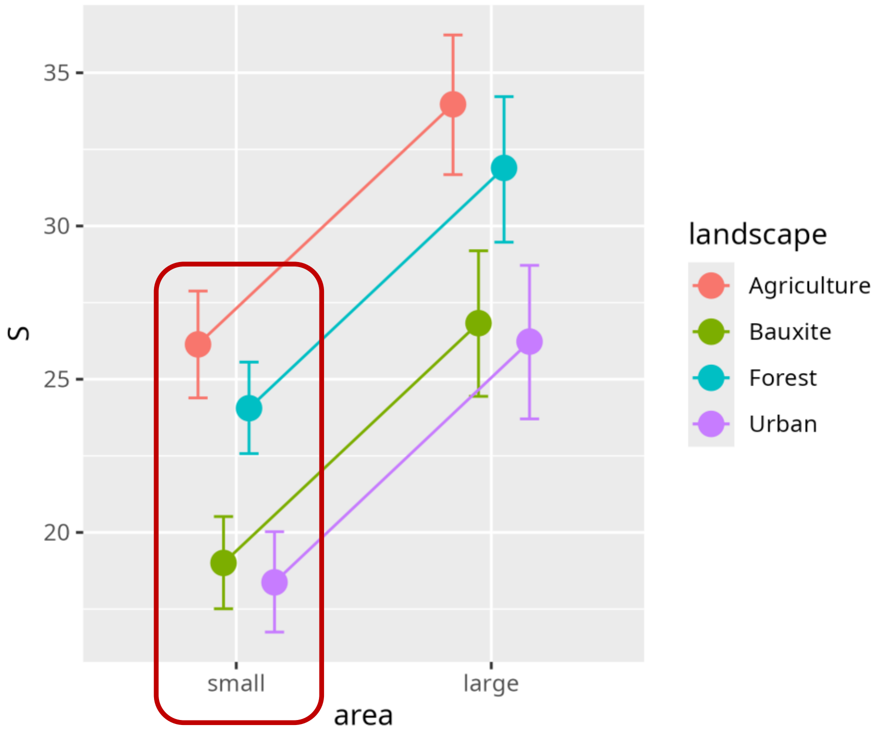
Post-hoc analysis: multiple predictors

```
> brm( S ~ landscape + area )
```

Q: What is the mean species richness in small areas?

```
> emmeans(fit_additive, ~area)
 area   emmean lower.HPD upper.HPD
 small    21.9      21.0      22.7
 large    29.7      27.8      31.6

Results are averaged over the levels of: landscape
Point estimate displayed: median
HPD interval probability: 0.95
```



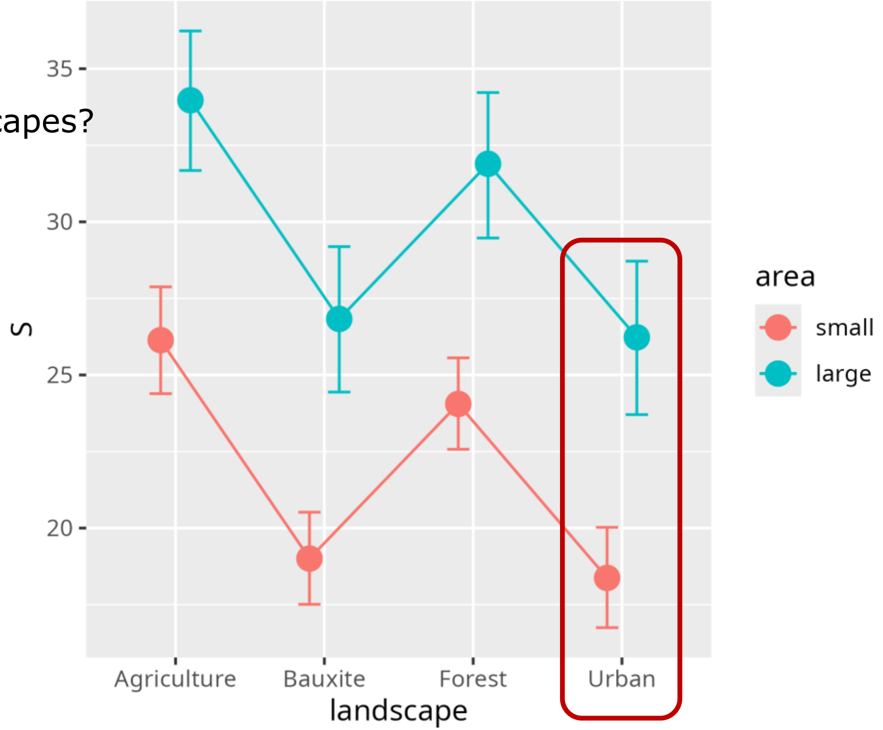
Post-hoc analysis: multiple predictors

```
> brm( S ~ landscape + area )
```

Q: What is the mean species richness in urban landscapes?

```
> emmeans(fit_additive, ~landscape)
 landscape   emmean lower.HPD upper.HPD
 Agriculture   30.0      28.4      31.7
 Bauxite       22.9      21.3      24.6
 Forest        28.0      26.3      29.7
 Urban         22.3      20.5      24.2

Results are averaged over the levels of: area
Point estimate displayed: median
HPD interval probability: 0.95
```



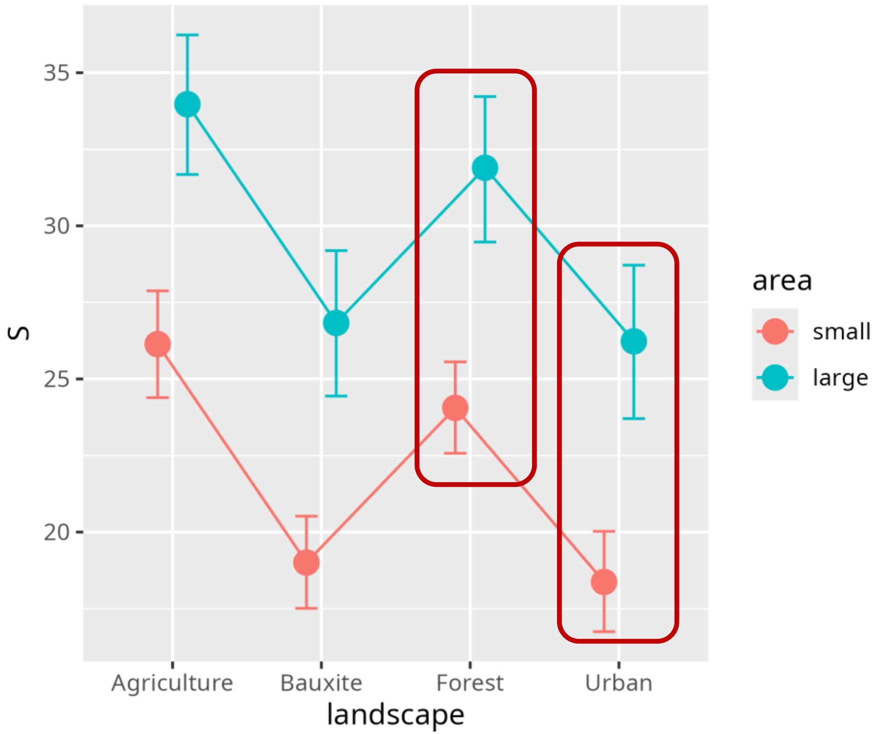
Post-hoc analysis: multiple predictors

```
> brm( S ~ landscape + area )
```

Q: What is the mean difference between urban and forest landscapes?

```
> emmeans(fit_additive, ~landscape) |> pairs()
 contrast                estimate lower.HPD upper.HPD
 Agriculture - Bauxite      7.128    4.9378      9.35
 Agriculture - Forest       2.073   -0.0856      4.36
 Agriculture - Urban        7.779    5.4187     10.13
 Bauxite - Forest          -5.064   -7.1364     -2.98
 Bauxite - Urban            0.653   -1.6066      2.70
 Forest - Urban             5.721    3.4428      7.91

Results are averaged over the levels of: area
Point estimate displayed: median
HPD interval probability: 0.95
```



Categorical & continuous predictors (ANCOVA)

Categorical & continuous predictor

Example: bird species richness

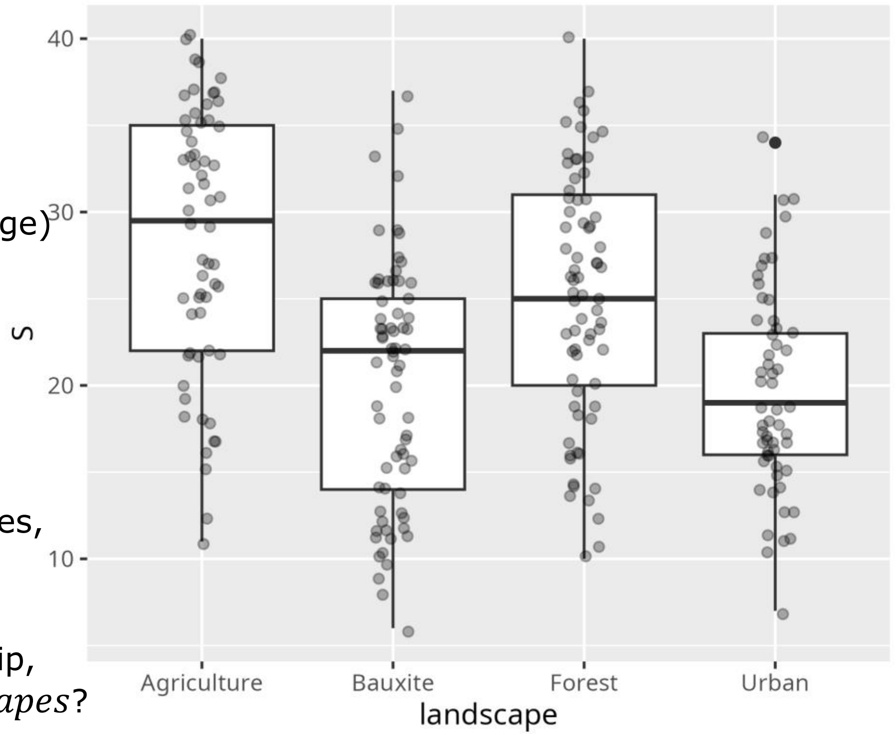
A lot of unexplained variation in `S ~ landscape`

Last section: added predictor $area$ (levels: small, large)
`S ~ landscape + area`

If we have better resolved data for area (in km²)
→ Use **continuous predictor $log.area$**

Q: Is species richness S different over $landscape$ types, while controlling for $log.area$?

Q: How strong is the average species-area relationship, while controlling (acknowledging) different $landscapes$?



Categorical & continuous predictor

Example: S vs. landscape type & log.area

Fit a regression line to each landscape level

Additive model `S ~ landscape + log.area`

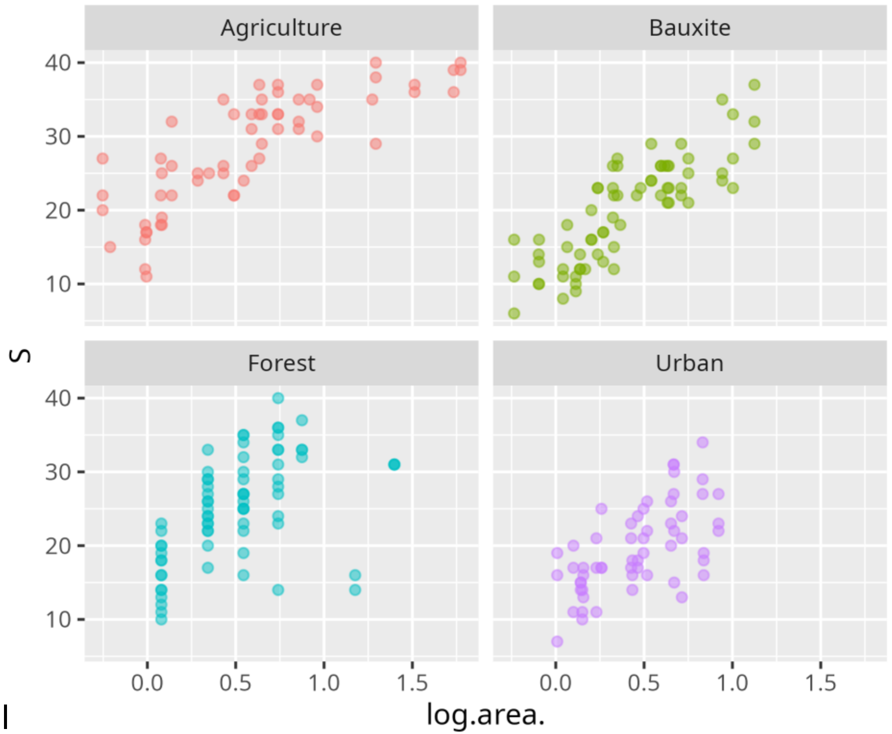
Slope (log.area) independent of landscape
→ identical slope

Individual intercepts for each landscape level

Factorial model `S ~ landscape * log.area`

Slope (log.area) depends on landscape

Individual intercepts & slopes for each landscape level



Additive model

`S ~ landscape + log.area`

$$\mu = \alpha(landscape) + \beta \cdot log.area$$

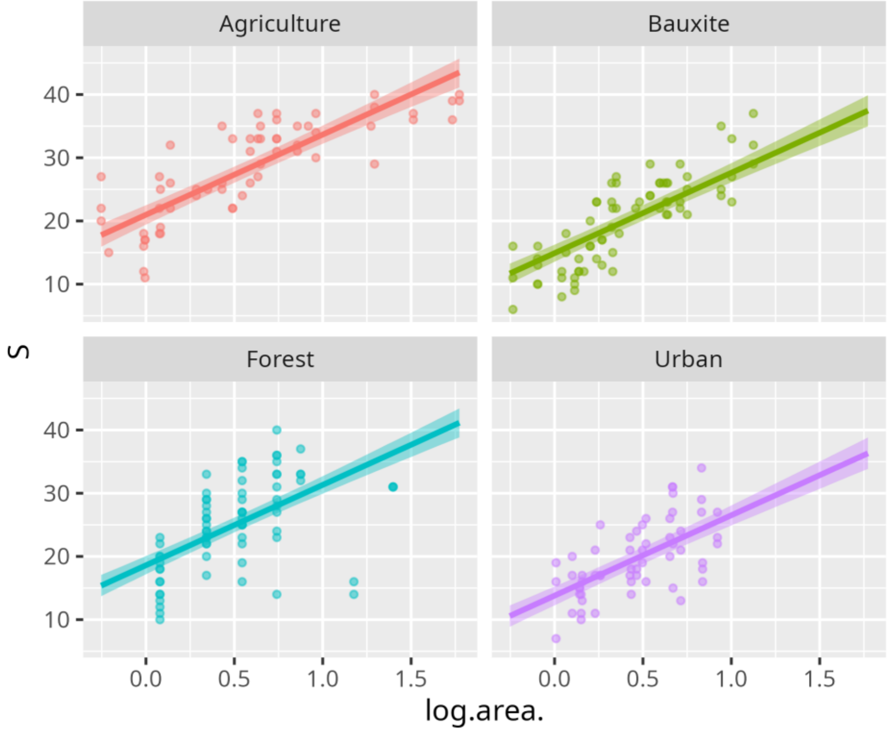
4 intercepts: $\alpha_{Agriculture}, \alpha_{Bauxite}, \alpha_{Forest}, \alpha_{Urban}$

1 slope: β

1 sdev: σ

Dummy-coding of intercepts:

$$\mu = a_0 + a_1 \cdot x_{Bauxite} + a_2 \cdot x_{Forest} + a_3 \cdot x_{Urban} + \beta \cdot log.area$$



Additive model

```
> brm(S ~ landscape + log.area)
```

Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	20.96	0.79	19.42	22.48
landscapeBauxite	-6.06	0.90	-7.77	-4.23
landscapeForest	-2.34	0.90	-4.08	-0.56
landscapeUrban	-7.16	0.93	-9.00	-5.41
log.area.	12.71	0.81	11.09	14.29

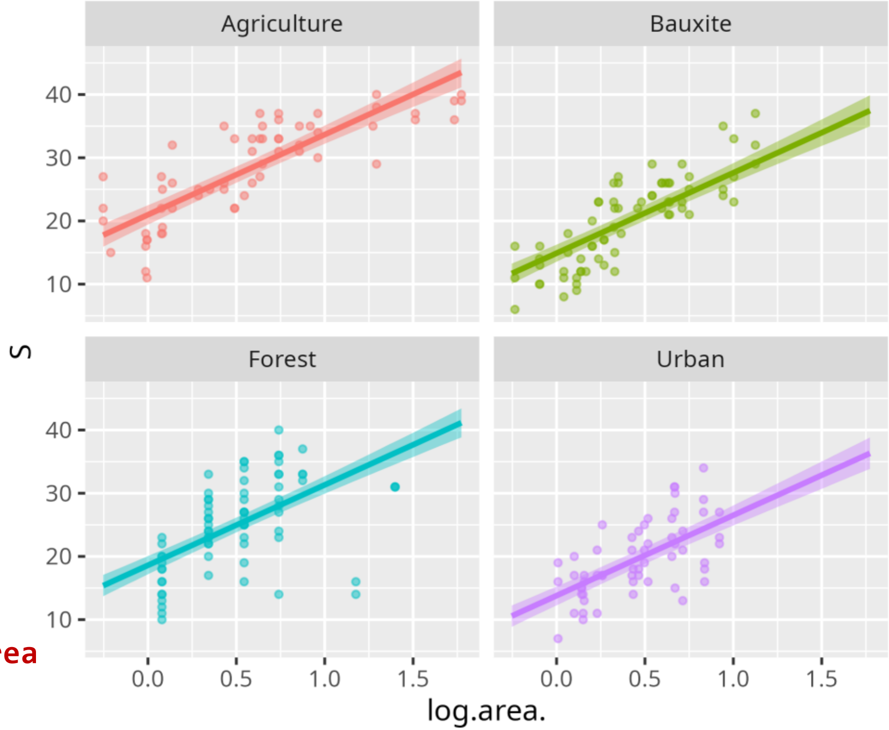
Q: Is there a difference between landscape types, while accounting for area size?

```
LOO(fit_additive, fit_logarea)
```

Model comparisons:

	elpd_diff	se_diff
fit_additive	0.0	0.0
fit_logarea	-31.8	7.7

Yes, **S~landscape+logarea** is a better model than **S~logarea**



Additive model

```
> brm(S ~ landscape + log.area)
```

Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	20.96	0.79	19.42	22.48
landscapeBauxite	-6.06	0.90	-7.77	-4.23
landscapeForest	-2.34	0.90	-4.08	-0.56
landscapeUrban	-7.16	0.93	-9.00	-5.41
log.area.	12.71	0.81	11.09	14.29

Q: Is there a difference between landscape types, while accounting for area size?

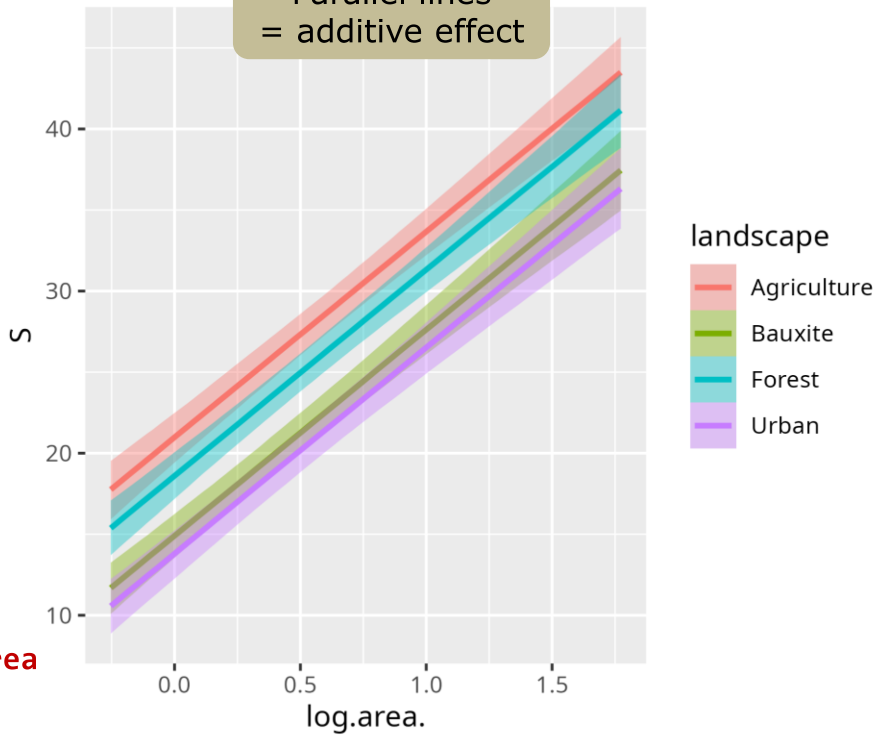
```
LOO(fit_additive, fit_logarea)
```

Model comparisons:

	elpd_diff	se_diff
fit_additive	0.0	0.0
fit_logarea	-31.8	7.7

Yes, **S~landscape+logarea** is a better model than **S~logarea**

Parallel lines
= additive effect



Additive model

```
> brm(S ~ landscape + log.area)
```

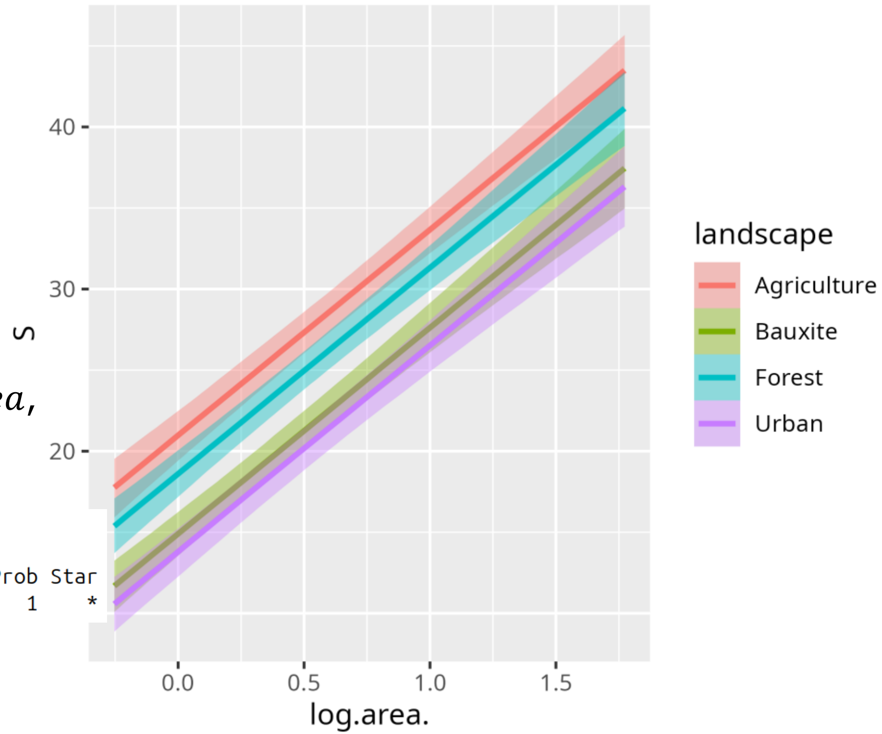
```
Regression Coefficients:
                 Estimate Est.Error l-95% CI u-95% CI
Intercept           20.96      0.79    19.42    22.48
landscapeBauxite    -6.06      0.90    -7.77    -4.23
landscapeForest     -2.34      0.90    -4.08    -0.56
landscapeUrban      -7.16      0.93    -9.00    -5.41
log.area.           12.71      0.81    11.09    14.29
```

Q: Is there a positive relation between S and $log.area$, while acknowledging different $landscapes$?

```
> hypothesis(fit_additive, "log.area.>0")
Hypothesis Tests for class b:
      Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Star
1 (log.area.) > 0    12.71      0.81    11.41    14.03        Inf         1    *
```

Yes, posterior distribution of slope positive

Alternatively, you could also do a model comparison (LOO)



Additive model

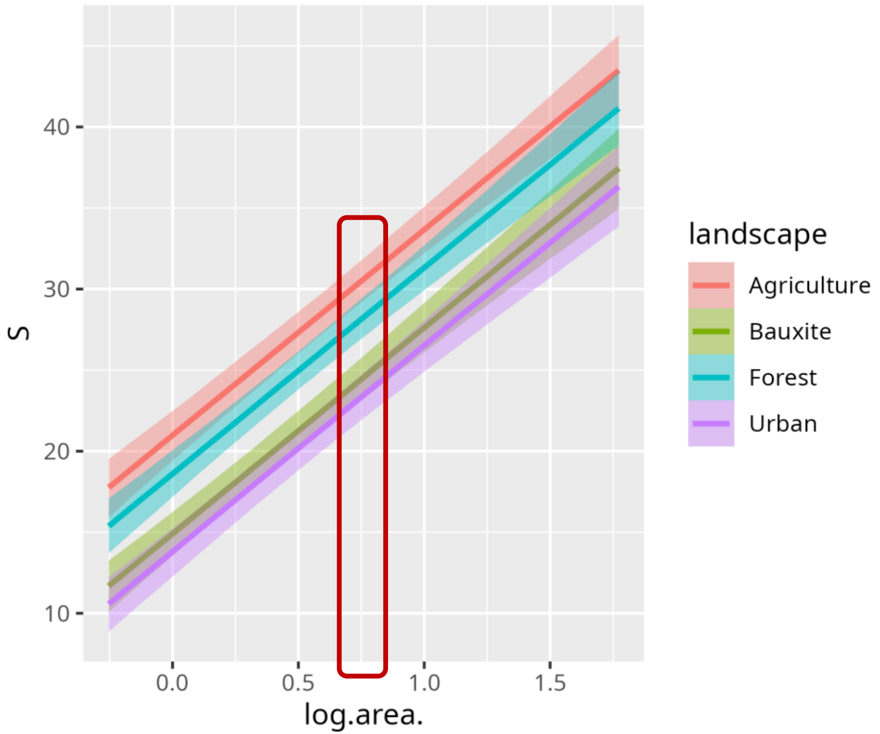
Post-hoc analysis

```
> brm(S ~ landscape + log.area)
```

Q: What are mean intercepts and
pairwise differences between landscape types?

```
> emmeans(fit_additive, ~landscape)
 landscape   emmean lower.HPD upper.HPD
 Agriculture   27.1      25.8      28.4
 Bauxite       21.1      19.9      22.4
 Forest        24.8      23.6      25.9
 Urban         20.0      18.7      21.3

Point estimate displayed: median
HPD interval probability: 0.95
> emmeans(fit_additive, ~landscape) |> pairs()
 contrast              estimate lower.HPD upper.HPD
 Agriculture - Bauxite     6.05     4.301      7.80
 Agriculture - Forest      2.35     0.648      4.13
 Agriculture - Urban       7.14     5.426      9.02
 Bauxite - Forest         -3.71    -5.312     -1.99
 Bauxite - Urban           1.10    -0.557      3.01
 Forest - Urban            4.78     3.104      6.62
```



„Intercepts“ at mean log.area

Interaction model

```
S ~ landscape * log.area
```

$$\mu = \alpha(landscape) + \beta(landscape) \cdot log.area$$

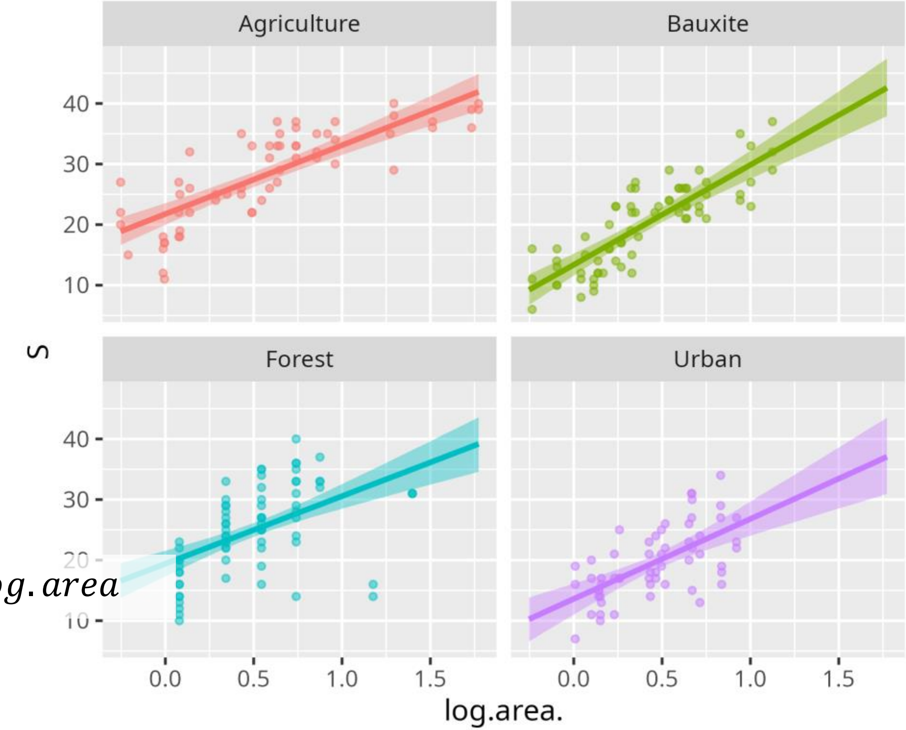
4 intercepts: $\alpha_{Agriculture}, \alpha_{Bauxite}, \alpha_{Forest}, \alpha_{Urban}$

4 slopes: $\beta_{Agriculture}, \beta_{Bauxite}, \beta_{Forest}, \beta_{Urban}$

1 sdev: σ

Dummy-coding of intercepts & slopes:

$$\mu = a_0 + a_1 \cdot x_{Bauxite} + a_2 \cdot x_{Forest} + a_3 \cdot x_{Urban} + (b_0 + b_1 \cdot x_{Bauxite} + b_2 \cdot x_{Forest} + b_3 \cdot x_{Urban}) \cdot log.area$$



Interaction model

```
> brm(S ~ landscape * log.area)
```

Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	21.74	0.92	19.96	23.54
landscapeBauxite				-5.74
landscapeForest				0.47
landscapeUrban				-5.11
log.area.				13.64
landscapeBauxite:log.area.	5.10	2.08	1.01	9.21
landscapeForest:log.area.	-0.23	2.10	-4.42	3.93
landscapeUrban:log.area.	1.86	2.60	-3.27	7.08

Summary table not helpful for predictors with >2 levels

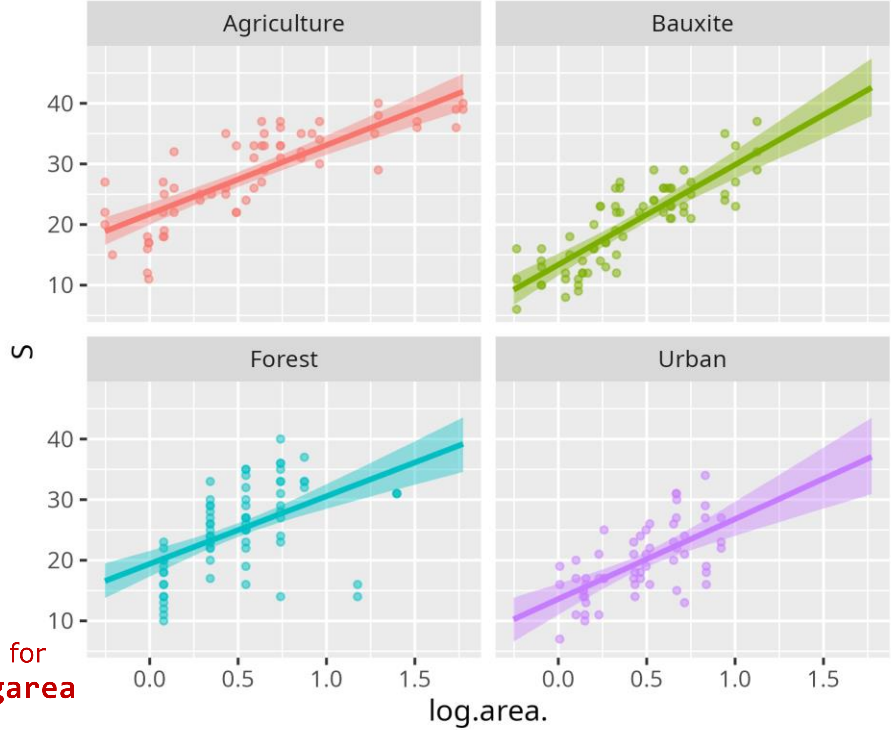
Q: Is the species-area relationship different between landscape types?

```
LOO(fit_factorial, fit_additive)
```

Model comparisons:

	elpd_diff	se_diff
fit_factorial	0.0	0.0
fit_additive	-0.4	2.4

No strong evidence for **S~landscape*logarea** against **S~landscape+logarea**



Interaction model

```
> brm(S ~ landscape * log.area)
```

Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	21.74	0.92	19.96	23.54
landscapeBauxite				-5.74
landscapeForest				0.47
landscapeUrban				-5.11
log.area.				13.64
landscapeBauxite:log.area.	5.10	2.08	1.01	9.21
landscapeForest:log.area.	-0.23	2.10	-4.42	3.93
landscapeUrban:log.area.	1.86	2.60	-3.27	7.08

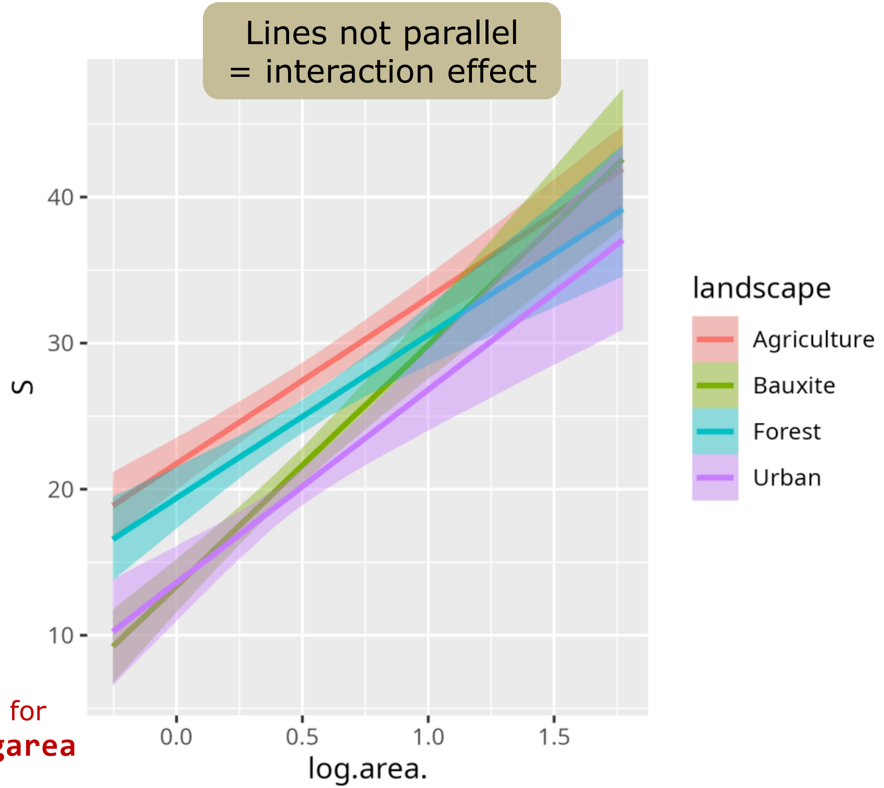
Summary table not helpful for predictors with >2 levels

Q: Is the species-area relationship different between landscape types?

```
LOO(fit_factorial, fit_additive)
```

```
Model comparisons:
              elpd_diff se_diff
fit_factorial  0.0       0.0
fit_additive  -0.4       2.4
```

No strong evidence for **S~landscape*logarea** against **S~landscape+logarea**



Interaction model

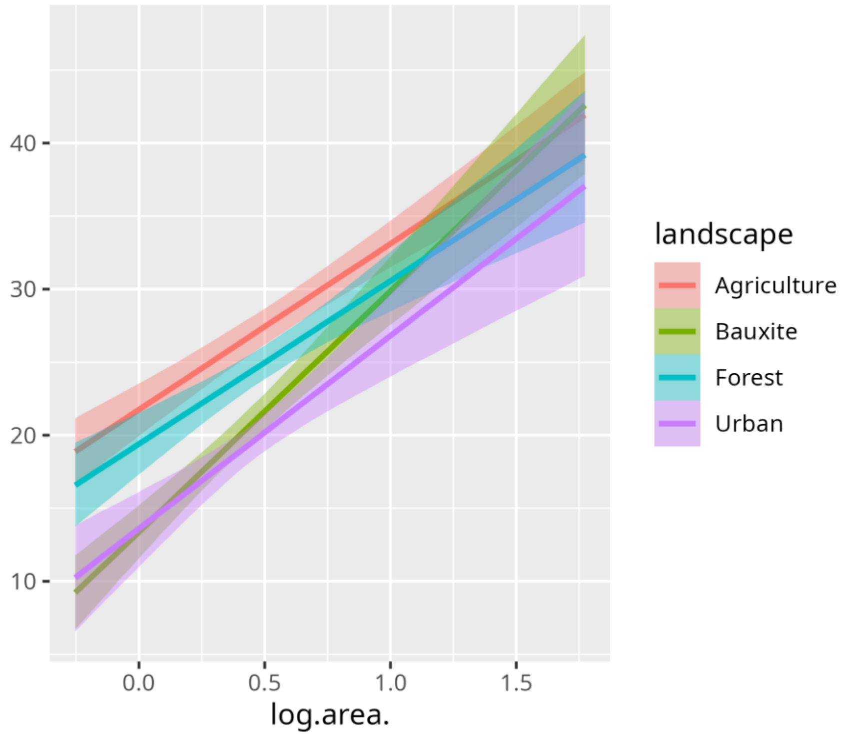
Post-hoc analysis
(If there was support for this model)

```
> brm(S ~ landscape * log.area)
```

Q: What are the predicted slopes?

```
> emtrends(fit_factorial, ~landscape, var="log.area.")
 landscape   log.area..trend lower.HPD upper.HPD
 Agriculture            11.4      9.12      13.8
 Bauxite                16.5     13.26      19.9
 Forest                 11.1      7.81      14.7
 Urban                  13.2      8.77      18.1

Point estimate displayed: median
HPD interval probability: 0.95
> emtrends(fit_factorial, ~landscape, var="log.area.") |> pairs()
 contrast              estimate lower.HPD upper.HPD
 Agriculture - Bauxite   -5.066    -9.224     -1.04
 Agriculture - Forest     0.224    -3.914      4.42
 Agriculture - Urban     -1.867    -7.125      3.19
 Bauxite - Forest         5.325     0.292     10.10
 Bauxite - Urban          3.206    -2.198      9.12
 Forest - Urban          -2.070    -8.073      3.42
```



Summary

Summary

- Regression, ANOVA, ANCOVA are just linear models
- Categorical variables can often be expressed by „dummy-coding" or by „effects-coding",
  brms uses dummy-coding as default
- In Bayesian stats, linearity is not that important
- But always check your model assumptions (e.g. PPC, check_model)
- Research question should guide you which model to fit and which „tests" to perform
- „Test" just means a statement about a research question,
  quantified through posterior distribution of effect sizes, model comparisons, or post-hoc analysis
- brms flexible and „all-in-one" package

Further reading

Bürkner, P. (2024). The brms Book [in progress]. https://paulbuerkner.com/software/brms-book/

Cinelli, C., Forney, A., & Pearl, J. (2024). A crash course in good and bad controls. *Sociological Methods & Research*, 53(3), 1071–1104. https://doi.org/10.1177/00491241221099552

Conn, P. B., Johnson, D. S., Williams, P. J., Melin, S. R., & Hooten, M. B. (2018). A guide to Bayesian model checking for ecologists. *Ecological Monographs*, 88(4), 526–542. https://doi.org/10.1002/ecm.1314

Fieberg, J. (2024). Statistics 4 Ecologists. https://statistics4ecologists-v2.netlify.app/ [Chapters 1,3]

Gelman, A., Hill, J., & Vehtari, A. (2020). Regression and Other Stories. *Cambridge University Press*. https://doi.org/10.1017/9781139161879 [Chapters 6-12]

Inchausti, P. (2023). Statistical Modeling With R: a dual frequentist and Bayesian approach for life scientists. *Oxford University Press*. [Chapters 4-7]

Kery, M. & Kellner, F. (2024): Applied Statistical Modelling for Ecologists. *Elsevier.* [Chapters 5-9]