Introduction to Applied Bayesian Analysis in Wildlife Ecology

Jeffrey W. Doser

May 11, 2024

Bayesian linear models

General linear models

- Response variable (y) = deterministic part + stochastic part
  - Stochastic = random (unexplained information, error)
  - Deterministic = systematic
- Why are they called "linear" models?
  - Response is a function of explanatory variables whose effects are **additive**
  - "Linear in the parameters"
  - Can still model nonlinear relationships

Which of the following are linear models?

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$y_i = \beta_0 + \beta_1 x_i^2 + \epsilon_i$$

$$y_i = \beta_0 x_i^{\beta_1} + \epsilon_i$$

$$y_i = \beta_0 \exp(\beta_1 x_i) + \epsilon_i$$

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i + \epsilon_i$$

Which of the following are linear models?

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$y_i = \beta_0 + \beta_1 x_i^2 + \epsilon_i$$

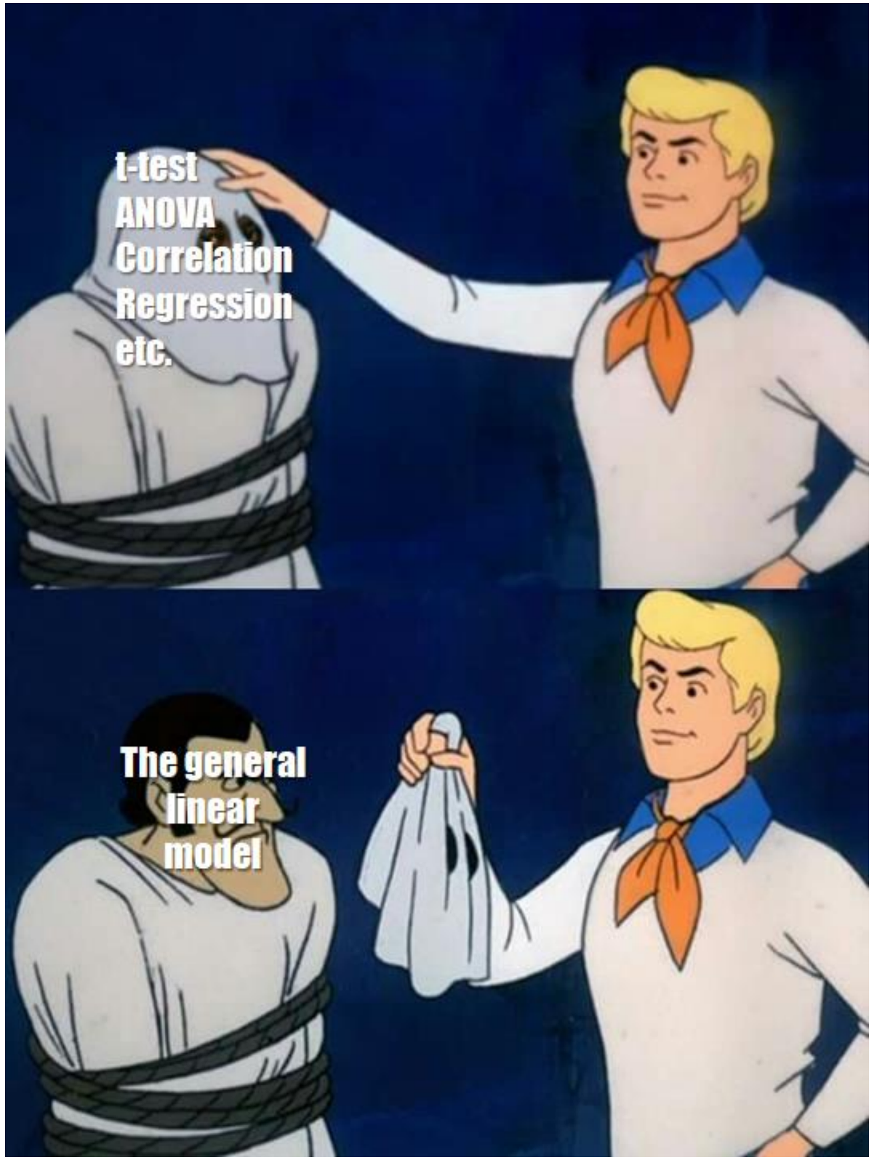
$$y_i = \beta_0 x_i^{\beta_1} + \epsilon_i$$

$$y_i = \beta_0 \exp(\beta_1 x_i) + \epsilon_i$$

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i + \epsilon_i$$

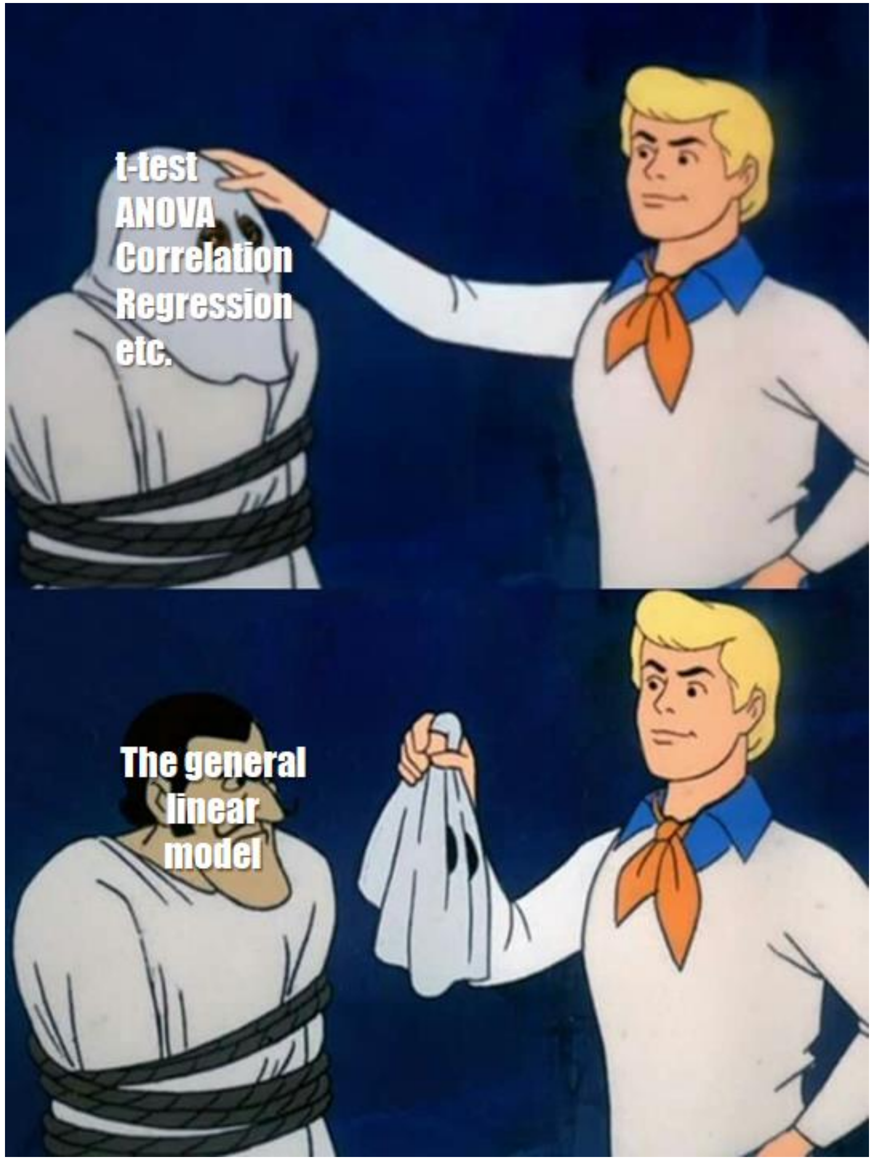
General linear models

- A broad class of models



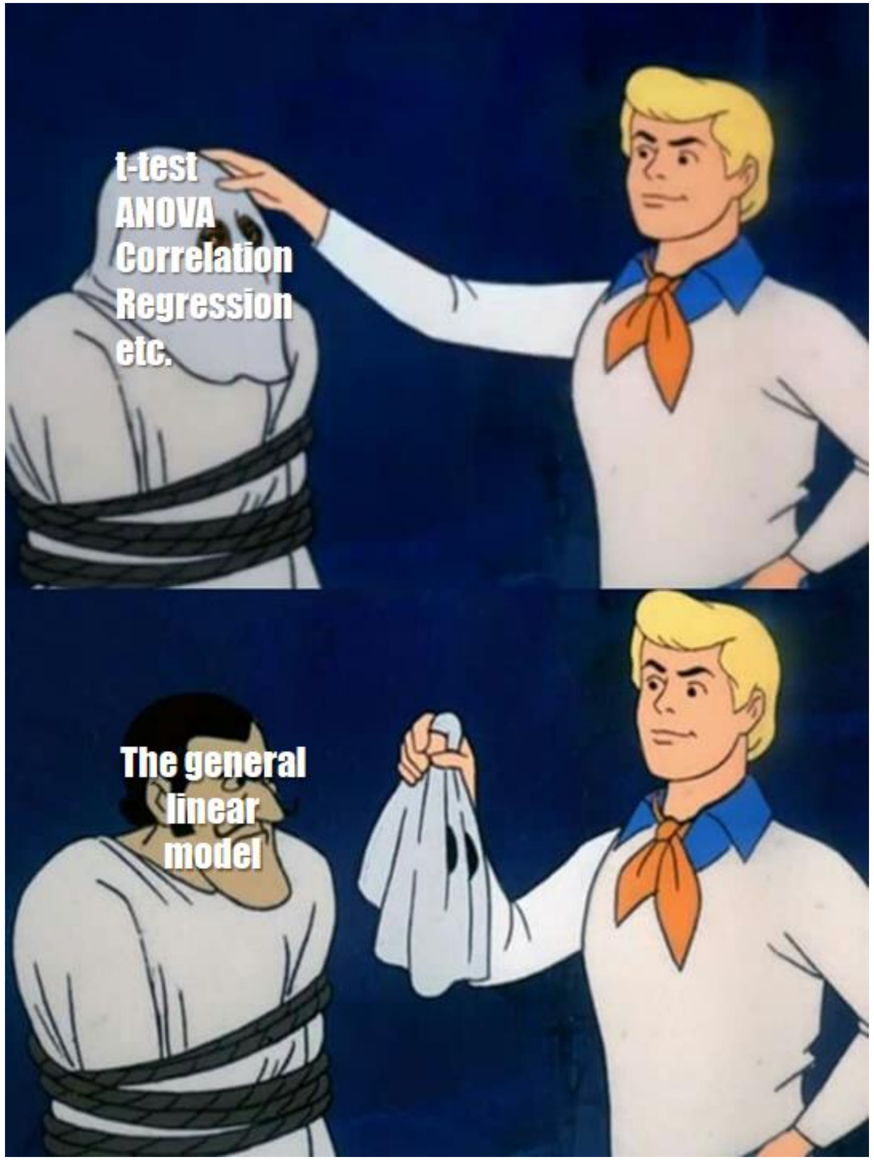
General linear models

- A broad class of models
- Lots of fancy design names for different linear models and designs
  - t test, ANOVA, ANCOVA, randomized complete block design, split plot design, etc.



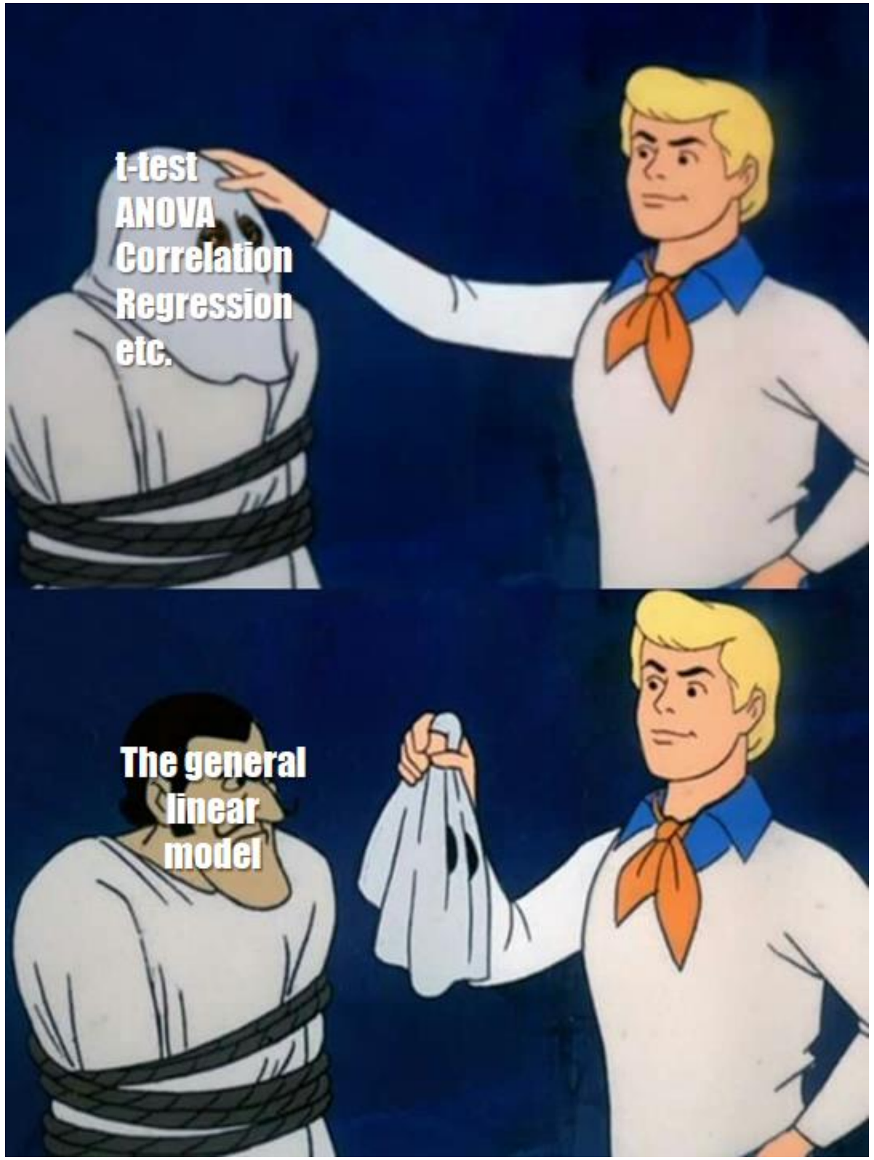
General linear models

- A broad class of models
- Lots of fancy design names for different linear models and designs
  - t test, ANOVA, ANCOVA, randomized complete block design, split plot design, etc.
- Don't focus on design names!



General linear models

- A broad class of models
- Lots of fancy design names for different linear models and designs
  - t test, ANOVA, ANCOVA, randomized complete block design, split plot design, etc.
- Don't focus on design names!
- The Bayesian approach de-emphasizes these classical design names because the estimation approach is always the same!!



Linear models

Deterministic Component

$$\mu_i = \beta_0 + \beta_1 \cdot x_{1,i} + \beta_2 \cdot x_{2,i} + \cdots + \beta_p \cdot x_{p,i}$$

Linear models

Can include both continuous and categorical predictor variables

Deterministic Component

$$\mu_i = \beta_0 + \beta_1 \cdot x_{1,i} + \beta_2 \cdot x_{2,i} + \cdots + \beta_p \cdot x_{p,i}$$

Linear models

Can include both continuous and categorical predictor variables

Deterministic Component

$$\mu_i = \beta_0 + \beta_1 \cdot x_{1,i} + \beta_2 \cdot x_{2,i} + \cdots + \beta_p \cdot x_{p,i}$$

Stochastic Component (the likelihood)

$$y_i \sim \mathrm{Normal}(\mathrm{mean} = \mu_i, \mathrm{variance} = \sigma^2)$$

Linear models

Can include both continuous and categorical predictor variables

Deterministic Component

$$\mu_i = \beta_0 + \beta_1 \cdot x_{1,i} + \beta_2 \cdot x_{2,i} + \cdots + \beta_p \cdot x_{p,i}$$

Stochastic Component (the likelihood)

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

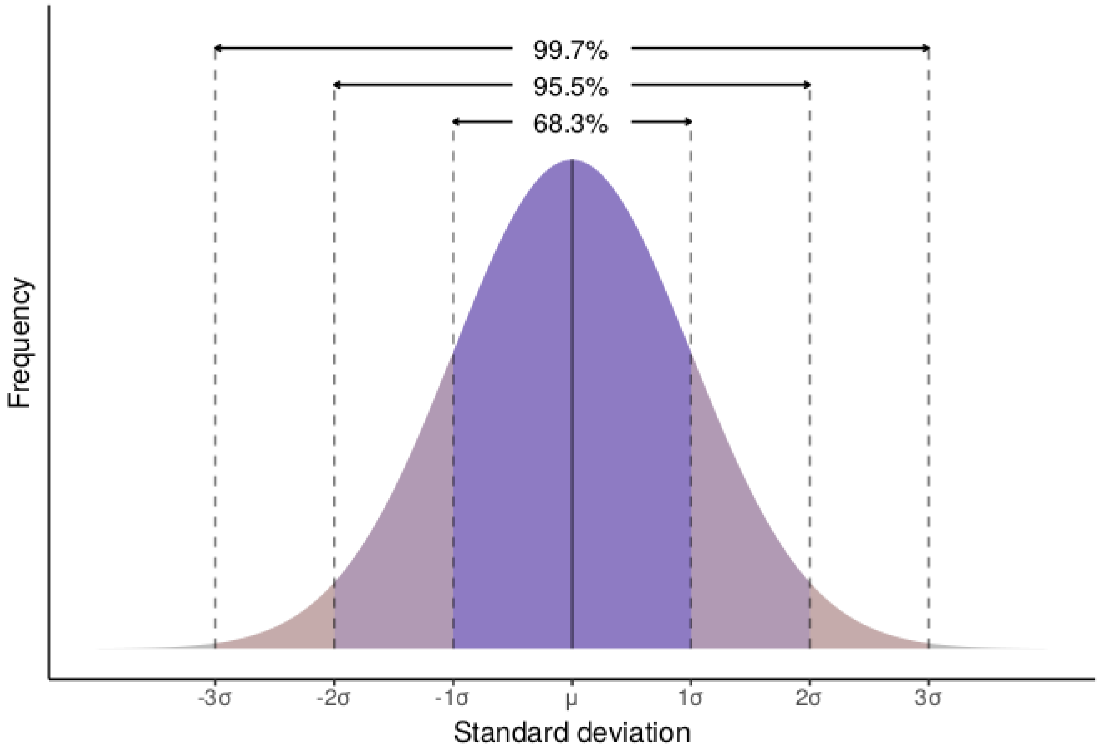
$$y_i \sim \text{Normal}(\text{mean} = \mu_i, \text{sd} = \sigma)$$

Priors in a Bayesian linear model

- We need priors for all regression coefficients (intercept and slopes) and *either* the residual variance or standard deviation
- Some software puts prior on variance, some software puts prior on standard deviation
- `brms` puts the prior on the residual standard deviation
- If you don't care about the priors much and just want them to be vague and weakly informative, then the default brms priors are a good option

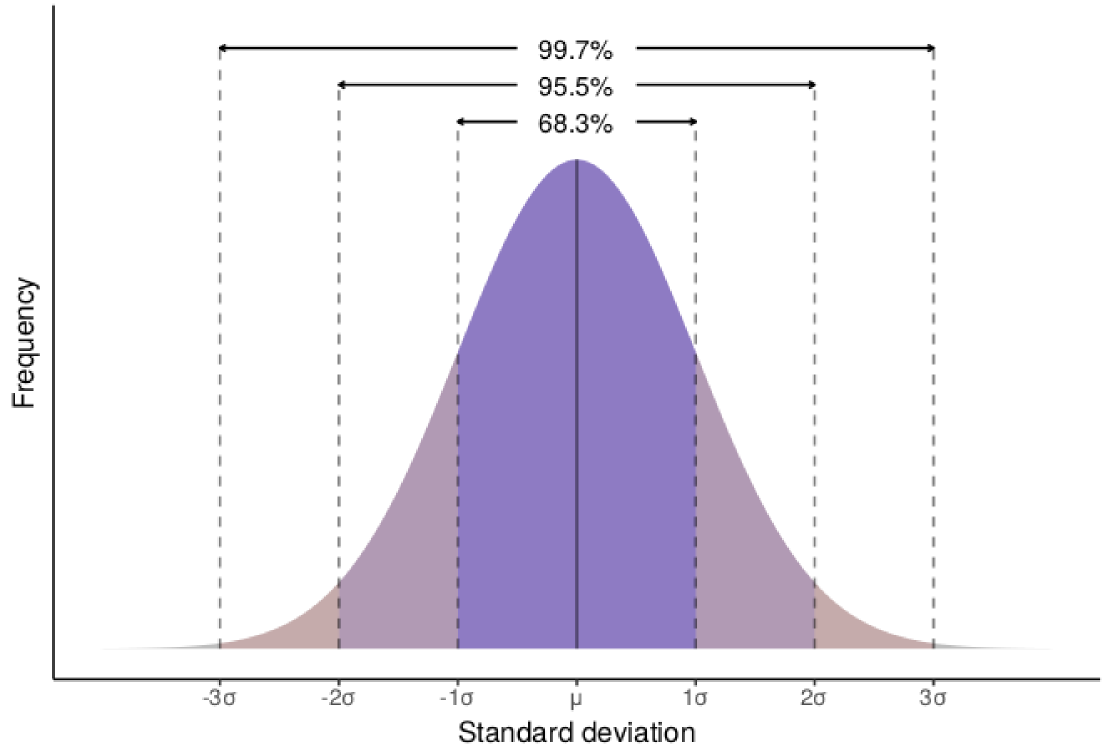
Priors for regression coefficients

- Most common choice of prior distribution is a normal distribution with a mean of 0 and a very large variance



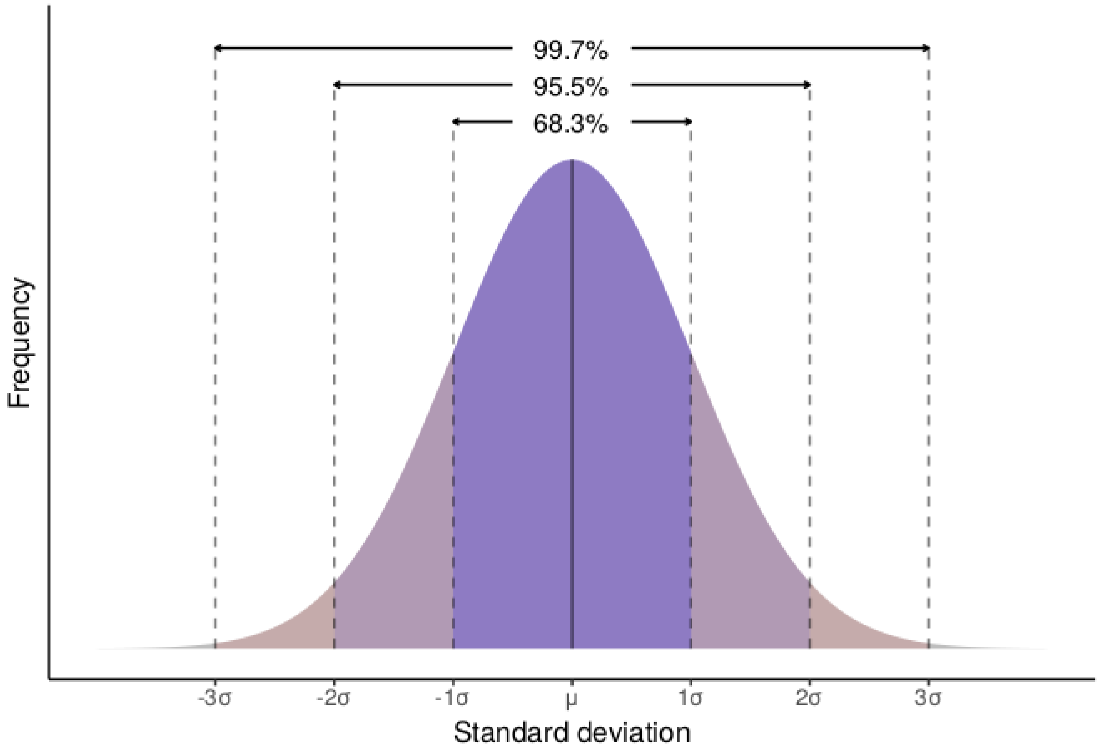
Priors for regression coefficients

- Most common choice of prior distribution is a normal distribution with a mean of 0 and a very large variance
- "Flat prior": very common, gives equal probability to all possible values of the parameter



Priors for regression coefficients

- Most common choice of prior distribution is a normal distribution with a mean of 0 and a very large variance
- "Flat prior": very common, gives equal probability to all possible values of the parameter
- Other options: Student's t distribution
- `brms` defaults:
  - Intercept: Student's t
  - Slope terms: flat prior



Priors for the standard deviation

- Remember: the standard deviation only takes positive values
- Common choices:
  - Gamma
  - Inverse-gamma
  - Half Student's t
  - Half Cauchy
- `brms` default: half Student's t

Priors for the standard deviation

- Remember: the standard deviation only takes positive values
- Common choices:
  - Gamma
  - Inverse-gamma
  - Half Student's t
  - Half Cauchy
- `brms` default: half Student's t

A key benefit to modern Bayesian software like brms is the default priors are suitable for many applications

Example: Bird richness across an elevation gradient

How does bird richness vary across an elevation gradient and in relation to landowner type (private, state, federal)?

`3a-linear-model-brms.R`

What do you do after you fit a Bayesian model?

1. Convergence assessment
2. Model checking (Goodness of Fit assessment)
3. Model selection (model comparison)

What do you do after you fit a Bayesian model?

1. Convergence assessment
2. **Model checking (Goodness of Fit assessment)**
3. Model selection (model comparison)

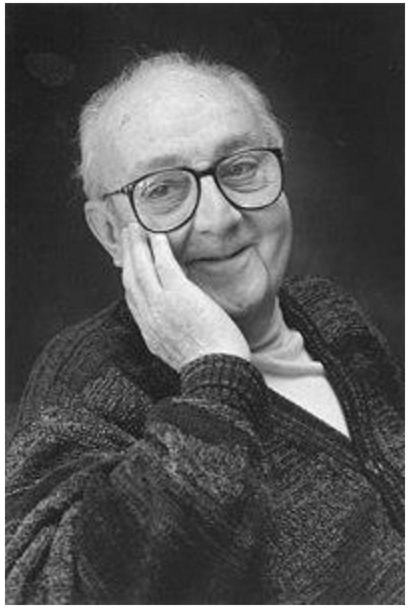
Model checking (GoF assessment)

- Does your model adequately reflect the data that you have?
- Does your data meet the assumptions of the given model?
- A good model should generate "replicate" data that closely resemble the true data you have collected
- If our model cannot generate new data similar to the data we collected, it may not adequately represent the true process we are interested in

Model checking (GoF assessment)

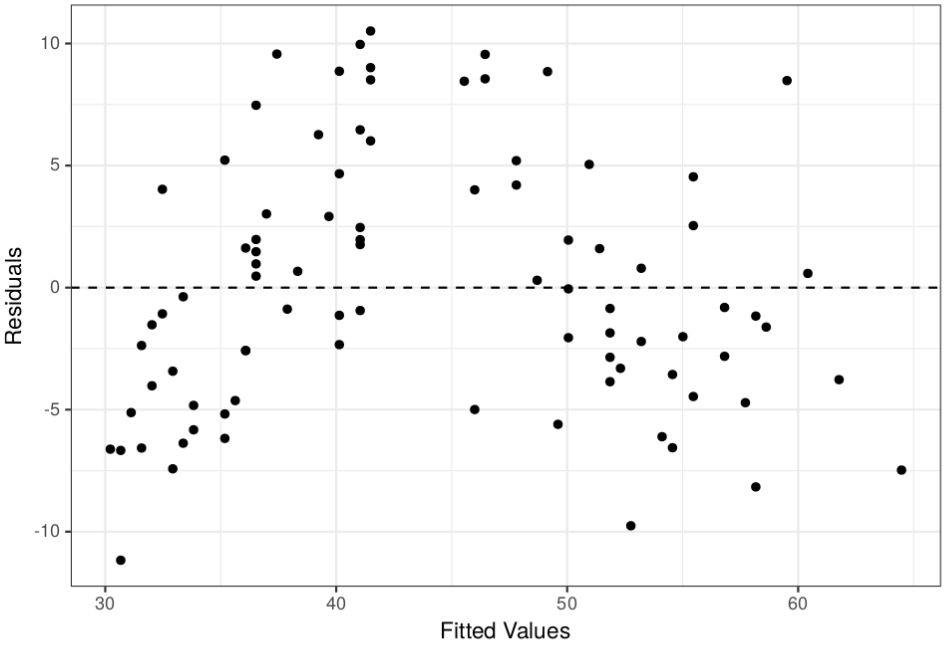
- Does your model adequately reflect the data that you have?
- Does your data meet the assumptions of the given model?
- A good model should generate "replicate" data that closely resemble the true data you have collected
- If our model cannot generate new data similar to the data we collected, it may not adequately represent the true process we are interested in

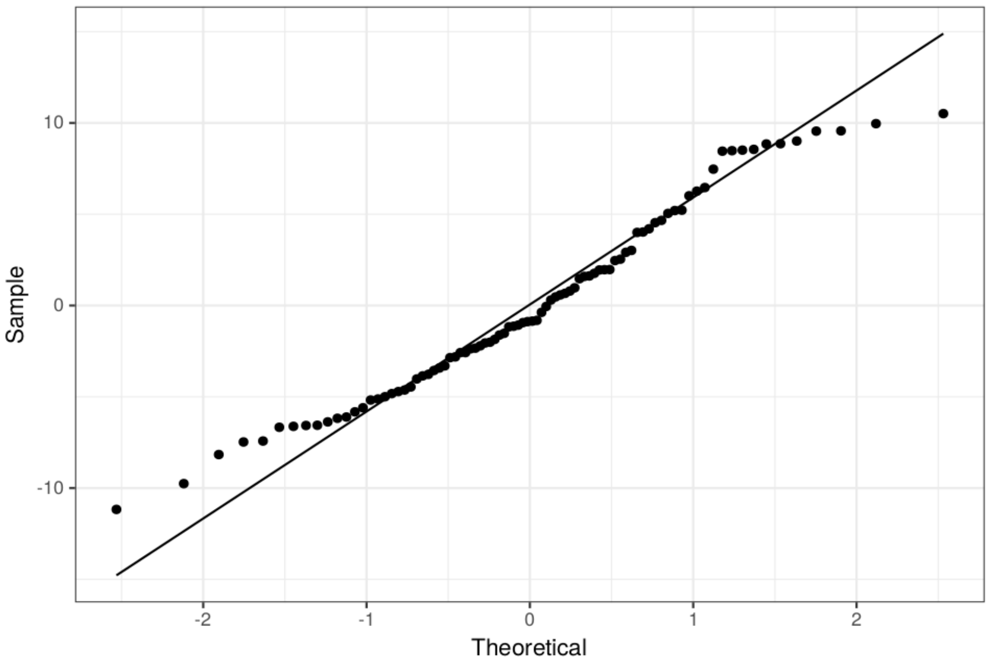
"All models are wrong, but some are useful"
- George Box



Residual diagnostics are a form of model checking

- Standard residual diagnostics can be used for Bayesian linear models
- Plot residuals vs. Model fitted (predicted) values: should show no clear and obvious patterns
- Normal probability plot (qq plot): should fall along a diagonal line





Model checking: posterior predictive checks

Model checking: posterior predictive checks

- **Posterior predictive checks**: a broad Bayesian approach for model checking

Model checking: posterior predictive checks

- **Posterior predictive checks**: a broad Bayesian approach for model checking
- Basic algorithm:
  1. Fit the model

Model checking: posterior predictive checks

- **Posterior predictive checks**: a broad Bayesian approach for model checking
- Basic algorithm:
  1. Fit the model
  2. Use the model results to predict a "new" data set using each posterior MCMC sample

Model checking: posterior predictive checks

- **Posterior predictive checks**: a broad Bayesian approach for model checking
- Basic algorithm:
  1. Fit the model
  2. Use the model results to predict a "new" data set using each posterior MCMC sample
  3. Compare the "new" data sets to the true data set

I strongly suggest using visual approaches for posterior predictive checks

Model checking: posterior predictive checks

- **Posterior predictive checks**: a broad Bayesian approach for model checking
- Basic algorithm:
  1. Fit the model
  2. Use the model results to predict a "new" data set using each posterior MCMC sample
  3. Compare the "new" data sets to the true data set
  4. Assess visually (and/or using numeric values) how well the "new" data sets line up with the true data set.

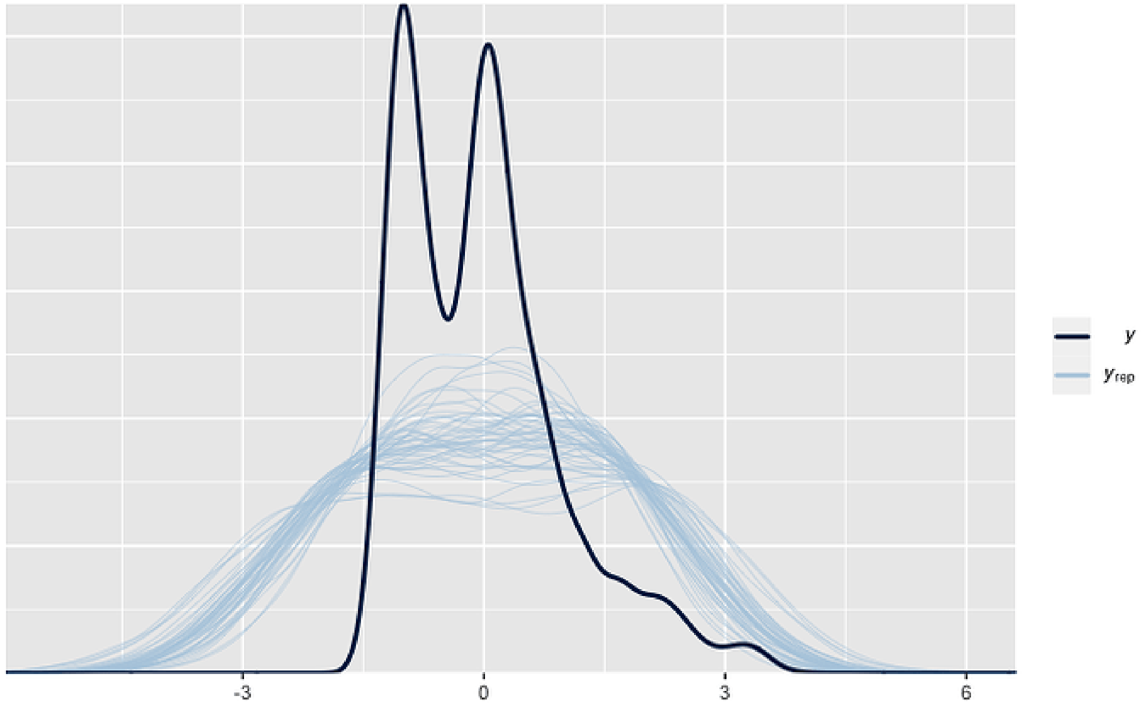
Model checking: posterior predictive checks

- **Posterior predictive checks**: a broad Bayesian approach for model checking
- Basic algorithm:
  1. Fit the model
  2. Use the model results to predict a "new" data set using each posterior MCMC sample
  3. Compare the "new" data sets to the true data set
  4. Assess visually (and/or using numeric values) how well the "new" data sets line up with the true data set.

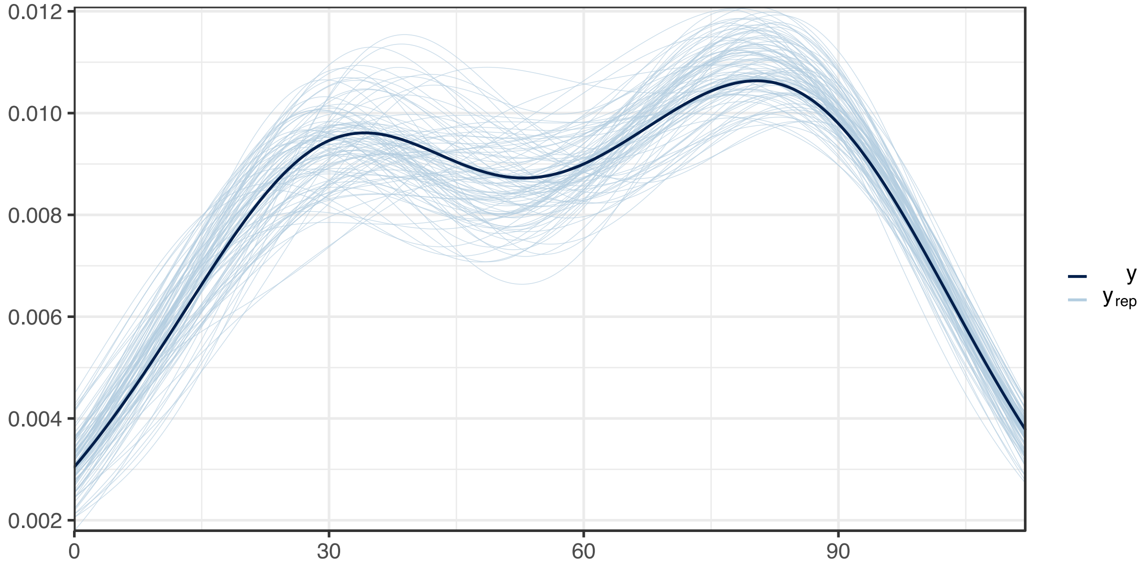
I strongly suggest using visual approaches for posterior predictive checks

Model checking: posterior predictive checks

Bad model fit

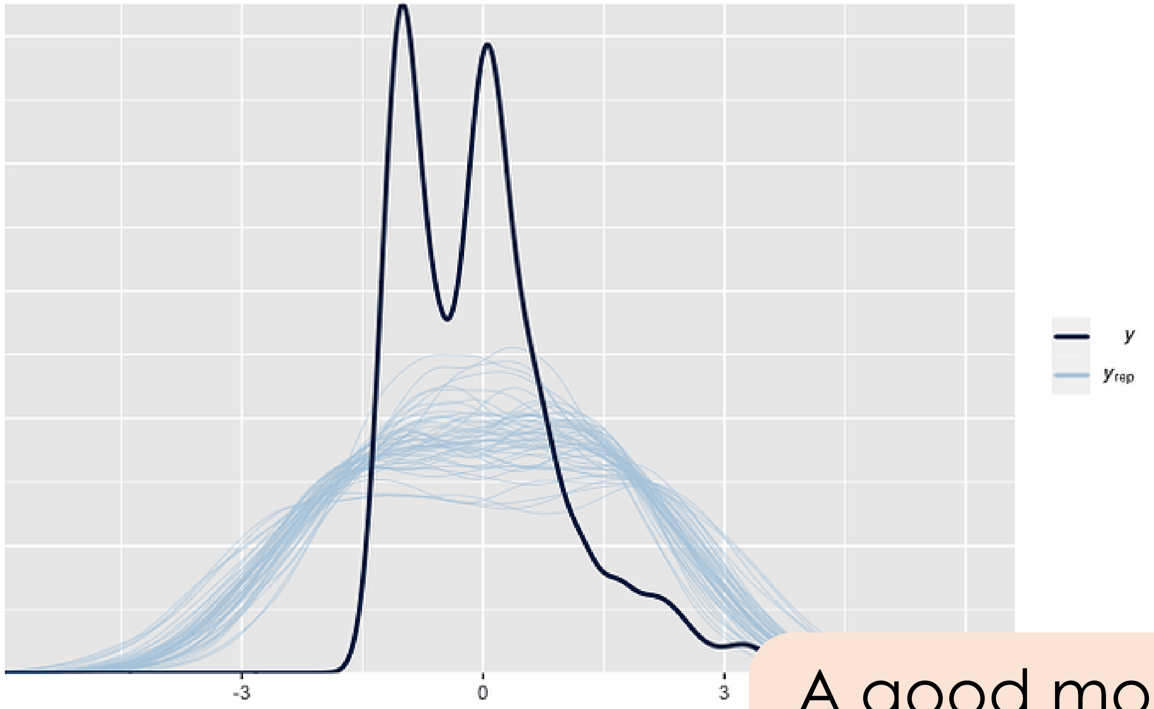


Good model fit

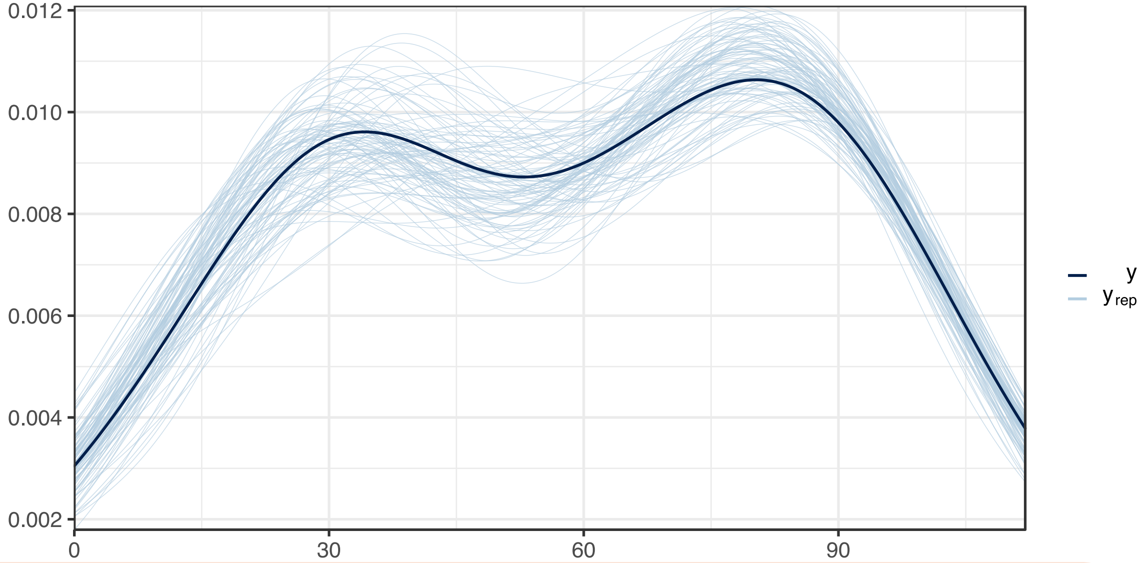


Model checking: posterior predictive checks

Bad model fit



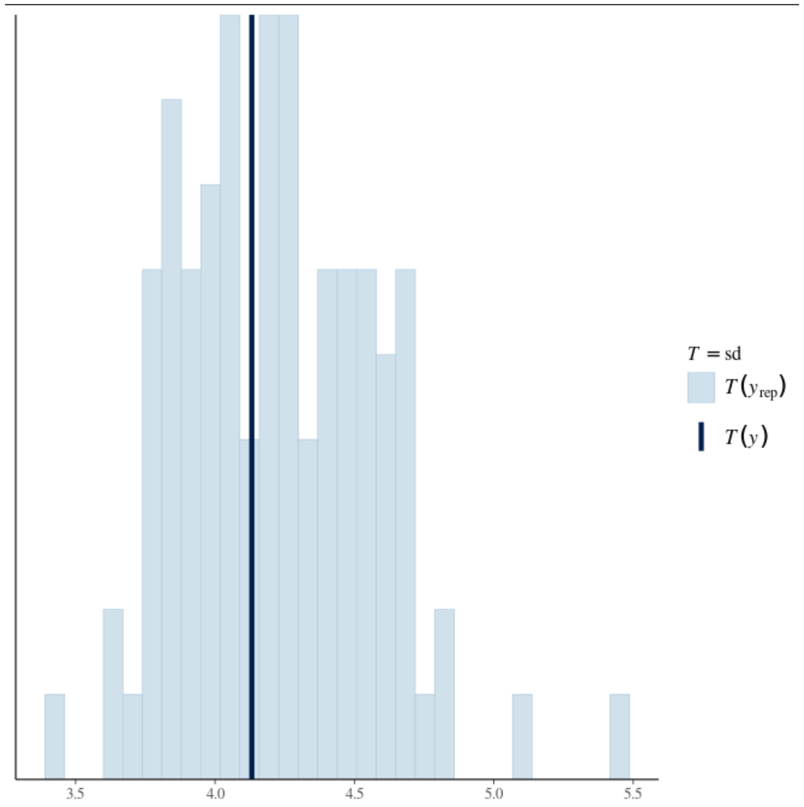
Good model fit



A good model fit is one where the curve for the true data could easily be one of the curves from the model fit

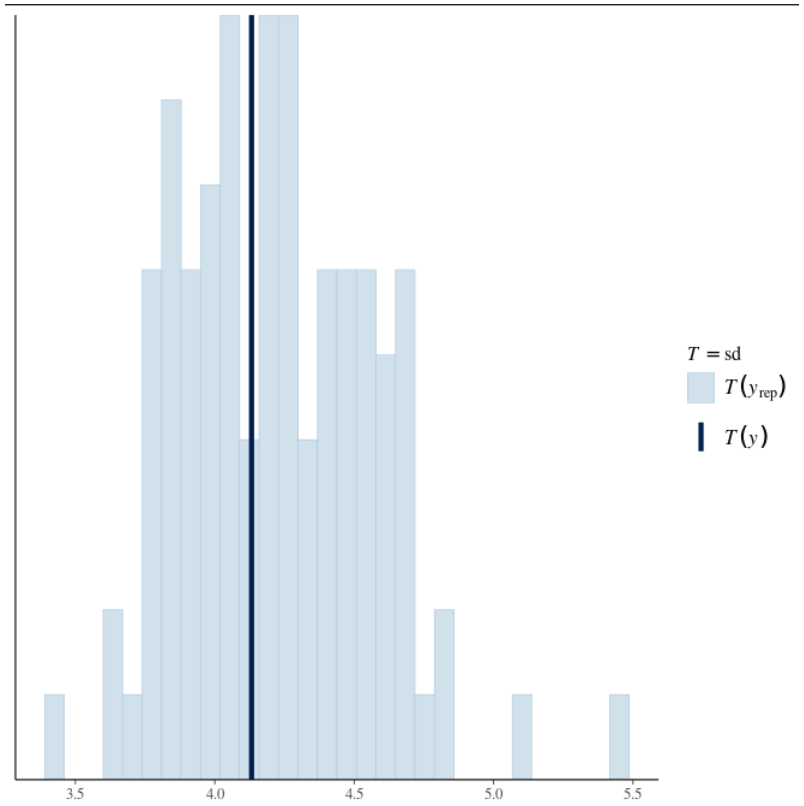
Model checking: posterior predictive checks

- Posterior predictive checks can be summarized with what is called a Bayesian p-value



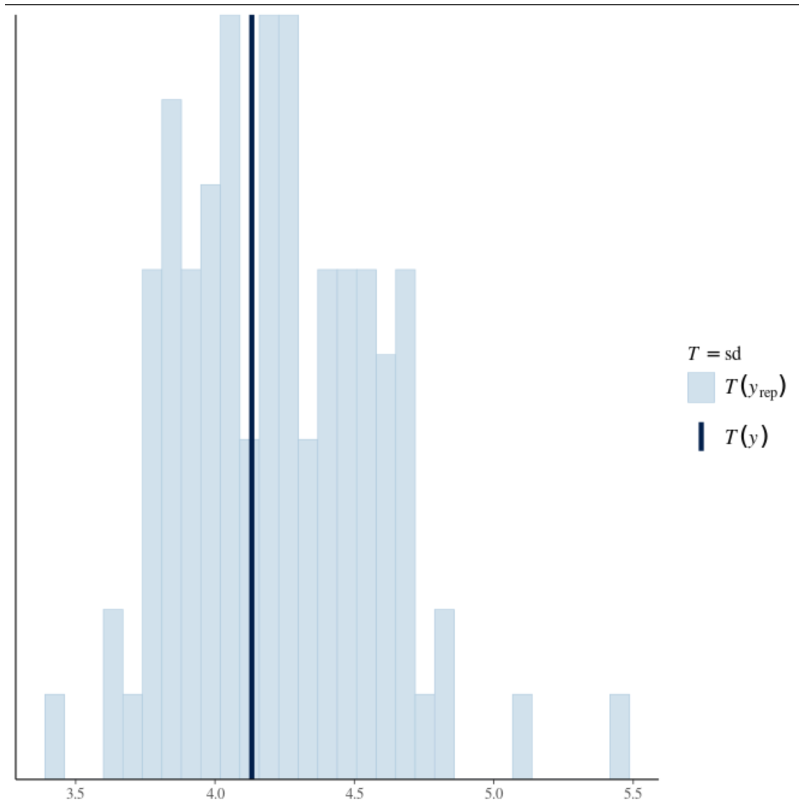
Model checking: posterior predictive checks

- Posterior predictive checks can be summarized with what is called a Bayesian p-value
- **Bayesian p-value**: the proportion of MCMC samples where a summary metric of the replicate data set is greater than the same metric calculated on the true data set.



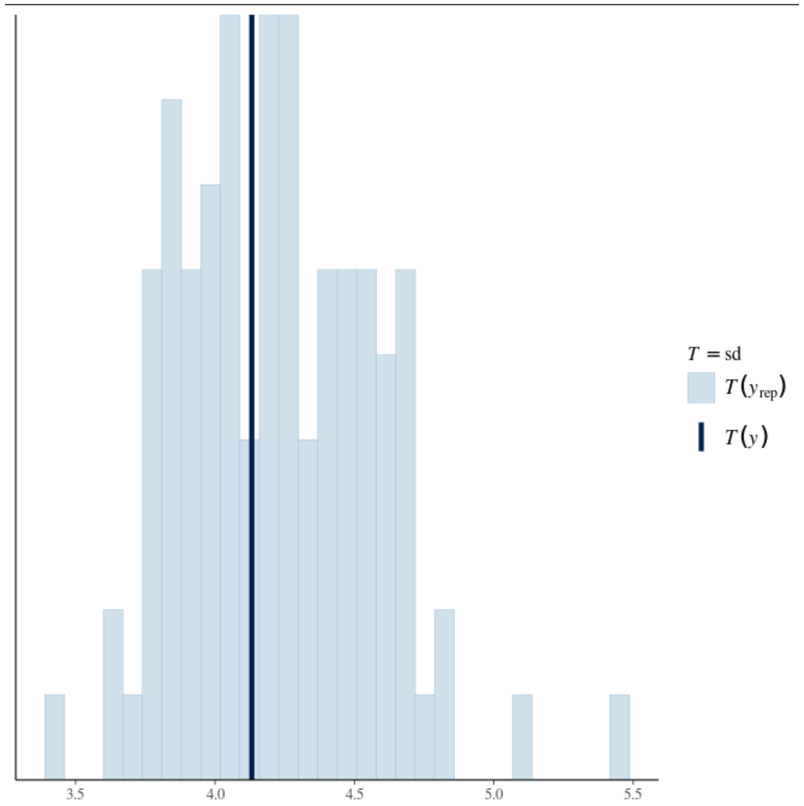
Model checking: posterior predictive checks

- Posterior predictive checks can be summarized with what is called a Bayesian p-value
- **Bayesian p-value**: the proportion of MCMC samples where a summary metric of the replicate data set is greater than the same metric calculated on the true data set.
- Ideally, Bayesian p-value = 0.5



Model checking: posterior predictive checks

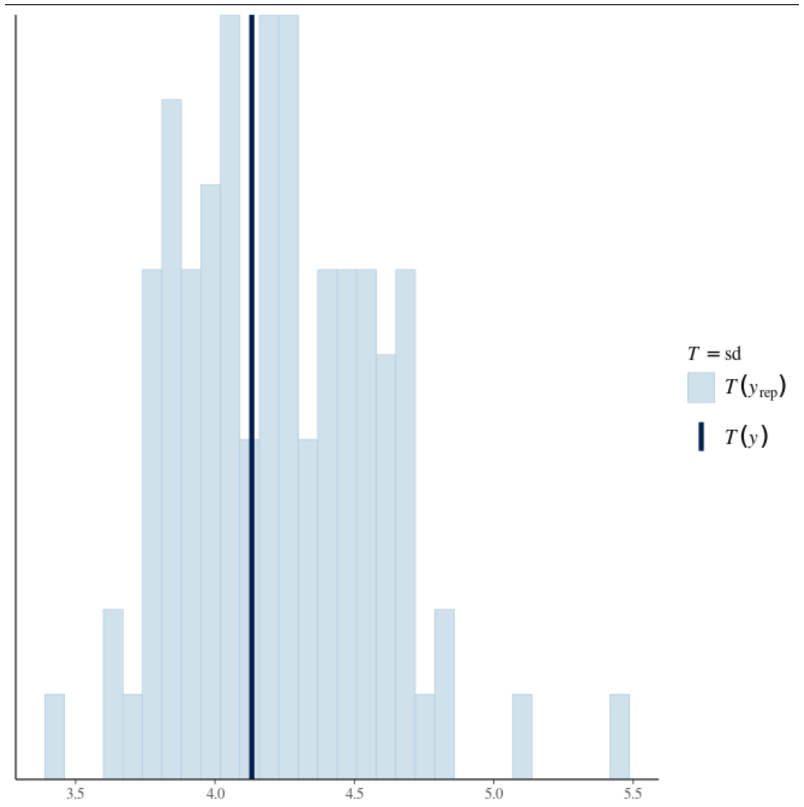
- Posterior predictive checks can be summarized with what is called a Bayesian p-value
- **Bayesian p-value**: the proportion of MCMC samples where a summary metric of the replicate data set is greater than the same metric calculated on the true data set.
- Ideally, Bayesian p-value = 0.5
- Bayesian p-value > 0.9 or < 0.1 suggest potentially bad model fit



Model checking: posterior predictive checks

- Posterior predictive checks can be summarized with what is called a Bayesian p-value
- **Bayesian p-value**: the proportion of MCMC samples where a summary metric of the replicate data set is greater than the same metric calculated on the true data set.
- Ideally, Bayesian p-value = 0.5
- Bayesian p-value > 0.9 or < 0.1 suggest potentially bad model fit

The Bayesian p-value is the proportion of bars to the right of the black line



Use caution with Bayesian p-values

- Can be extremely sensitive to outliers
- A Bayesian p-value of 0.5 does not necessarily indicate you have a good model

Bayesian p-values are useful, but you should always use graphical methods to assess PPCs as well

More on Bayesian model checking

ECOLOGICAL MONOGRAPHS
ECOLOGICAL SOCIETY OF AMERICA

Concepts & Synthesis | Full Access

A guide to Bayesian model checking for ecologists

Paul B. Conn, Devin S. Johnson, Perry J. Williams, Sharon R. Melin, Mevin B. Hooten

What do you do after you fit a Bayesian model?

1. Convergence assessment
2. Model checking (Goodness of Fit assessment)
3. **Model selection (model comparison)**

Model selection/comparison

- How can we compare different models?
- Which model among a set of candidate models is the best?
- In frequentist methods, we can use AIC (the most common type of information criterion).
- Bayesian methods have a variety of information criteria as well.
- Two common approaches: WAIC and LOO

WAIC and LOO

- Widely Applicable Information Criterion or Watanabe-Akaike Information Criterion
- Leave-one-out cross validation
- Can be used in the same manner as AIC
- Very applicable for a variety of complex models, unlike other information criteria (DIC) for Bayesian models
- Can easily calculate using `brms` and `Stan`, and only slightly more difficult with `NIMBLE`

WAIC and LOO

- Widely Applicable Information Criterion or Watanabe-Akaike Information Criterion
- Leave-one-out cross validation
- Can be used in the same manner as AIC
- Very applicable for a variety of complex models, unlike other information criteria (DIC) for Bayesian models
- Can easily calculate using `brms` and `Stan`, and only slightly more difficult with `NIMBLE`

WAIC can be calculated in different ways, which can lead to differences across software

WAIC and LOO

- Widely Applicable Information Criterion or Watanabe-Akaike Information Criterion
- Leave-one-out cross validation
- Can be used in the same manner as AIC
- Very applicable for a variety of complex models, unlike other information criteria (DIC) for Bayesian models
- Can easily calculate using `brms` and `Stan`, and only slightly more difficult with `NIMBLE`

WAIC can be calculated in different ways, which can lead to differences across software

Home > Statistics and Computing > Article

Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC

Published: 30 August 2016

Volume 27, pages 1413–1432, (2017) Cite this article

Aki Vehtari, Andrew Gelman & Jonah Gabry

Example: Bird richness across an elevation gradient

How does bird richness vary across an elevation gradient and in relation to landowner type (private, state, federal)?

`3a-linear-model-brms.R`

Questions?