Today

- Finish off model checking
- Generalized linear models (GLM)
  - McElreath Ch 9

Independent project

- Complete analysis (EDA through inference & conclusions)
- ggplot, dplyr
- Preferably hierarchical model:
  - rstanarm: stan_glmer or stan_lmer
- Submit .md from .R or .Rmd
- Due end of semester

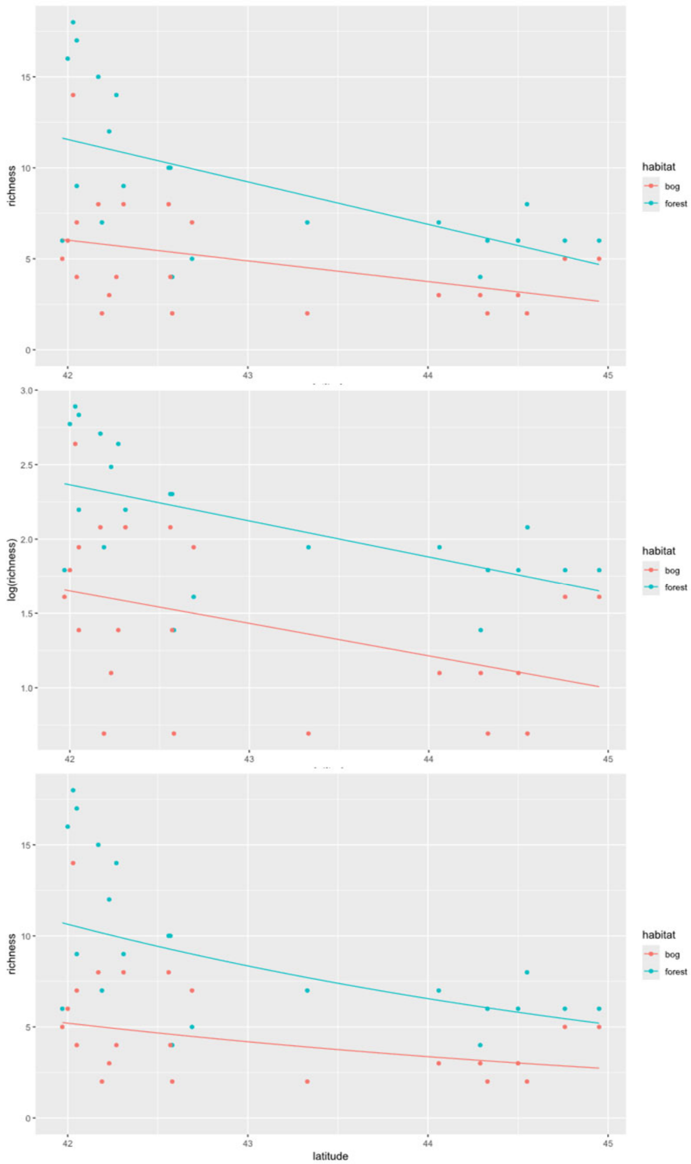
Model checking

- When is normality of errors important?
  - Not critical for inference about means
    - Frequentist: sampling distribution will still be approximately normal
    - Bayesian: posterior distribution will still be approximately normal or insensitive
  - Can be important for inference about prediction
    - Because: data generating process
- Good information for improving model

Model checking

- How can I know what to look for in a diagnostic plot?
  - simulate it!
- Continue coding demo

How to proceed?



Linear

Log-linear

Nonlinear

Scientific questions:

How does species richness vary with latitude?

Is this relationship different between habitats?

How different is species richness between habitats?

Main points McElreath Ch 9

- Generalized linear models

Data

Parameters

$$y_i \sim \mathrm{Distribution}(\mu_i, \ldots)$$

$$g(\mu_i) = \eta_i = \underbrace{\beta_0 + \beta_1 x_i}_{\text{Linear model}}$$

Link function

Linear predictor
"eta"

Main points McElreath Ch 9

- Generalized linear models

Data

Parameters

$$y_i \sim \text{Distribution}(\mu_i, \ldots)$$

$$g(\mu_i) = \eta_i = \underbrace{\beta_0 + \beta_1 x_i}$$

Link function

or

Linear model

Linear predictor
"eta"

$$\mu_i = g^{-1}(\beta_0 + \beta_1 x_i)$$

Inverse link function

Main points McElreath Ch 9

- Exponential family (some)
  - Exponential, Gamma, Normal, Poisson, Binomial
- Other distributions
  - with write-your-own or Bayesian, this doesn’t have to be from the exponential family
  - even more generalized!
- Link functions (some)
  - identity, log, logit

Most common models

Normal + Identity link	Poisson + Log link	Binomial + Logit link
$y_i \sim \mathrm{Normal}(\mu_i, \sigma)$	$y_i \sim \mathrm{Poisson}(\mu_i)$	$y_i \sim \mathrm{Binomial}(\mu_i, n)$
$\mu_i = \beta_0 + \beta_1 x_i$	$\log(\mu_i) = \beta_0 + \beta_1 x_i$	$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 x_i$

Key properties:

Normal	Poisson	Binomial
y: -∞ to ∞, continuous	y: 0 to ∞, discrete, integer	y: 0, 1, discrete, binary
μ: -∞ to ∞, continuous	μ: 0 to ∞, continuous	μ: 0 to 1, probability

Most common models

Normal + Identity link	Poisson + Log link	Binomial + Logit link
$y_i \sim \mathrm{Normal}(\mu_i, \sigma)$	$y_i \sim \mathrm{Poisson}(\mu_i)$	$y_i \sim \mathrm{Binomial}(\mu_i, n)$
$\mu_i = \beta_0 + \beta_1 x_i$	$\log(\mu_i) = \beta_0 + \beta_1 x_i$	$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 x_i$

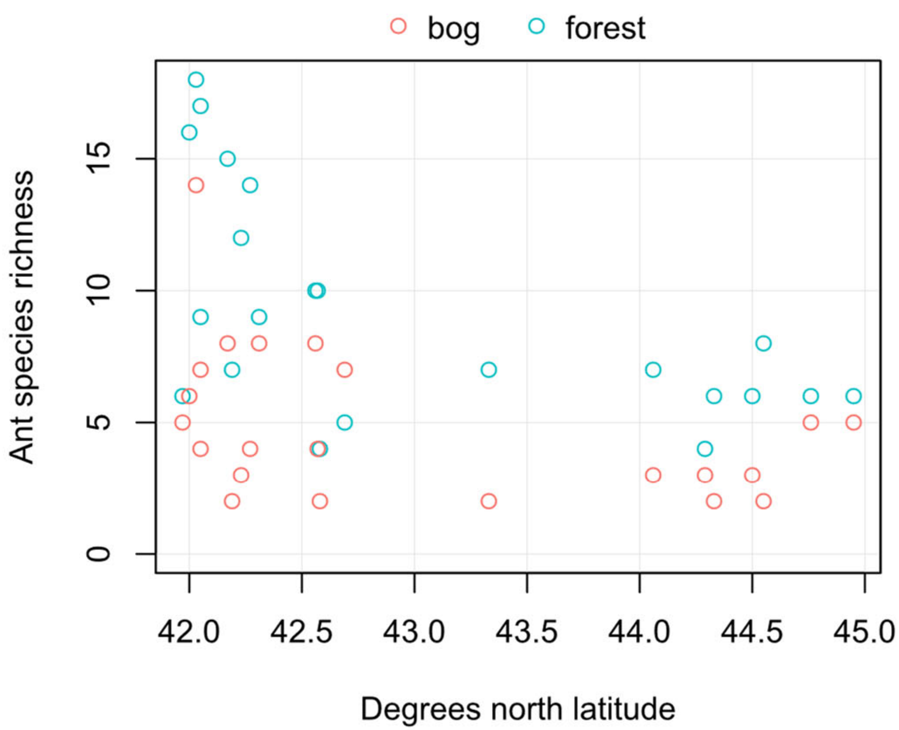
Linear predictor (always the same):

$\eta_i = \beta_0 + \beta_1 x_i$	$\eta_i = \beta_0 + \beta_1 x_i$	$\eta_i = \beta_0 + \beta_1 x_i$

Inverse link function:

$\mu_i = \eta_i$	$\mu_i = e^{\eta_i}$	$\mu_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}$

Which makes the most sense?



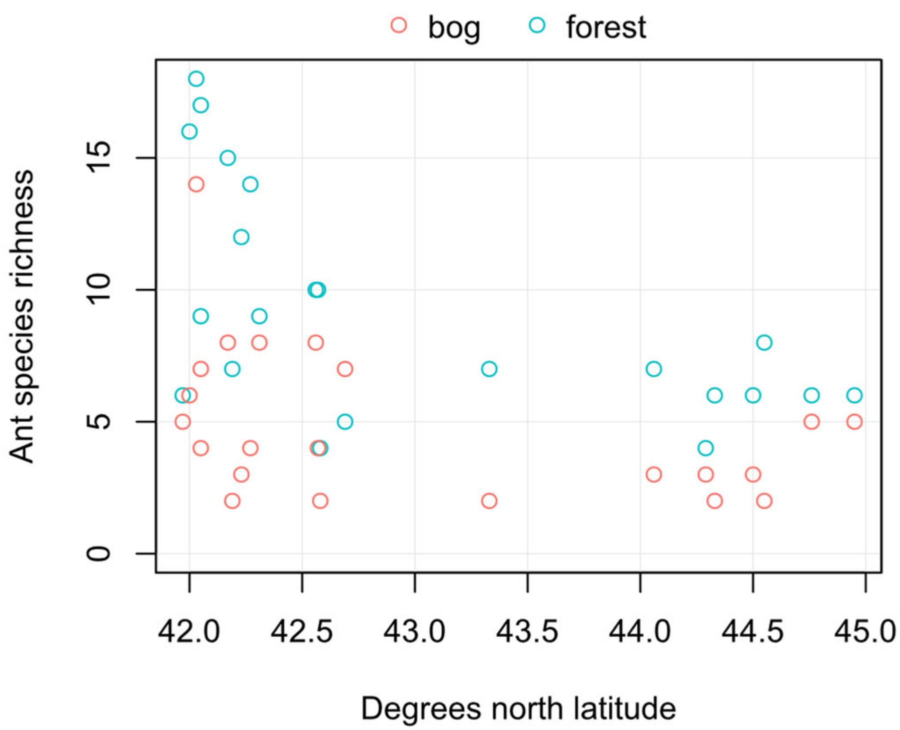
Most common models

Normal + Identity link	Poisson + Log link	Binomial + Logit link
$y_i \sim \mathrm{Normal}(\mu_i, \sigma)$	$y_i \sim \mathrm{Poisson}(\mu_i)$	$y_i \sim \mathrm{Binomial}(\mu_i, n)$
$\mu_i = \beta_0 + \beta_1 x_i$	$\log(\mu_i) = \beta_0 + \beta_1 x_i$	$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 x_i$

Inverse link functions:

$\mu_i = \eta_i$	$\mu_i = e^{\eta_i}$	$\mu_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}$

Which makes the most sense?



Poisson

+

Log link

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

$$\log(\mu_i) = \beta_0 + \beta_1 x_i + \ldots$$

$$\eta_i = \beta_0 + \beta_1 x_i + \ldots$$

$$\mu_i = e^{\eta_i}$$