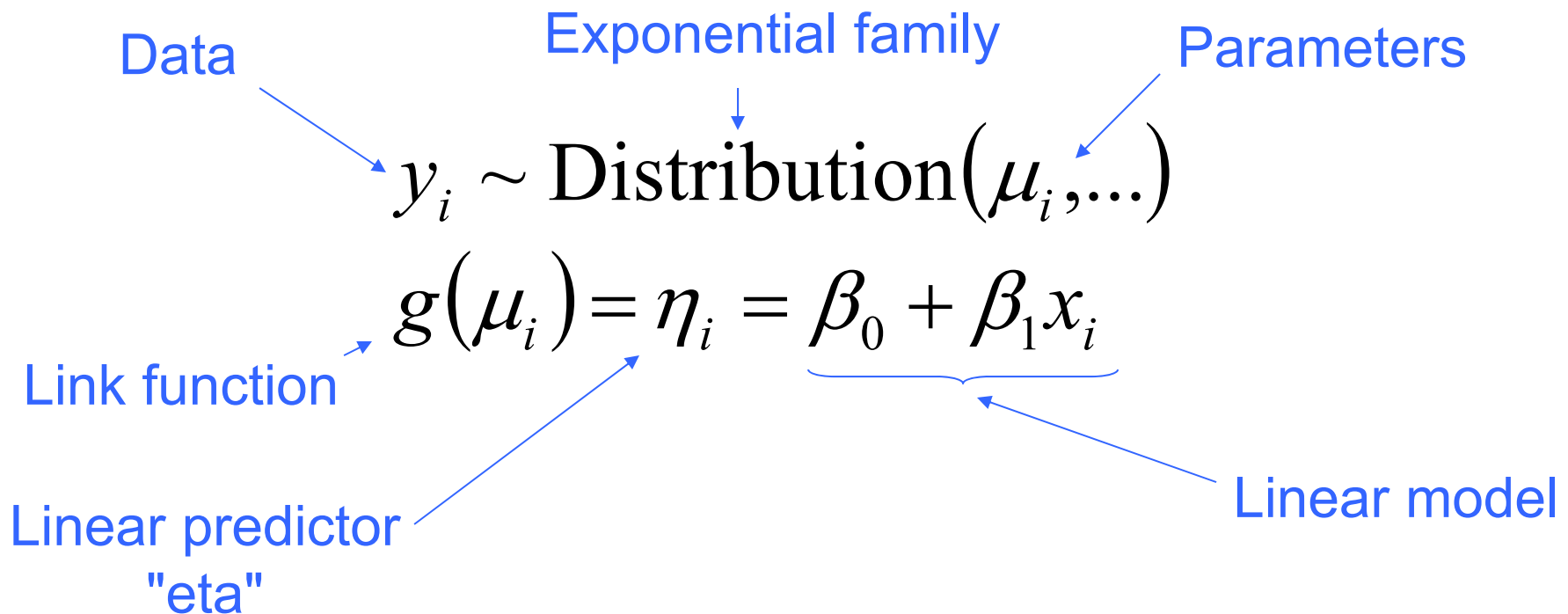# Diagnostics: model checking

- Systematic departures of the process (biological) model from the data
- Poor error distribution
- Mistakes in data
- Outliers
- Influential data points

Tools

- Plotting the fitted model with the data
- Residuals vs fitted
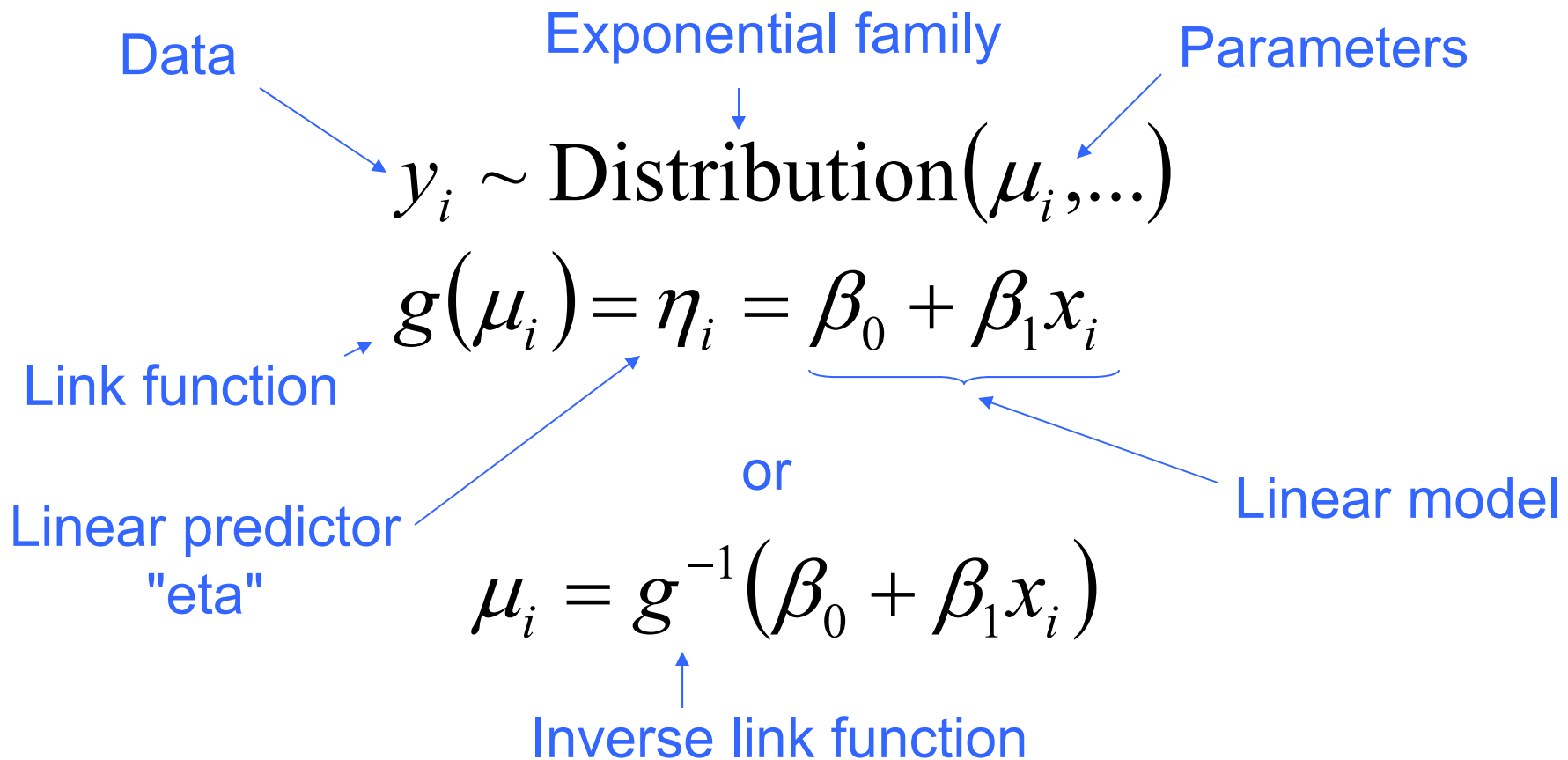- QQ plot, hist(residuals)
- Leave one out (LOO) influence algorithm

Main points McElreath Ch 9

- Generalized linear models

Data

Exponential family

Parameters

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

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

Link function

Linear predictor "eta"

Linear model

Main points McElreath Ch 9

- Generalized linear models

Data   Exponential family   Parameters

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

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

Link function

Linear predictor "eta"

Linear model

or

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

Inverse link function

Main points McElreath Ch 9

- Exponential family (some)
  - Exponential, Gamma, Normal, Poisson, Binomial
- Link functions (some)
  - identity, log, logit

Most common models

Normal	Poisson	Binomial
+	+	+
Identity link	Log link	Logit link

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

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

$$y_i \sim \text{Binomial}(\mu_i, n)$$

$$\mu_i = \beta_0 + \beta_1 x_i$$

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

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = \beta_0 + \beta_1 x_i$$

Inverse link functions:

$$\mu_i = \eta_i$$

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

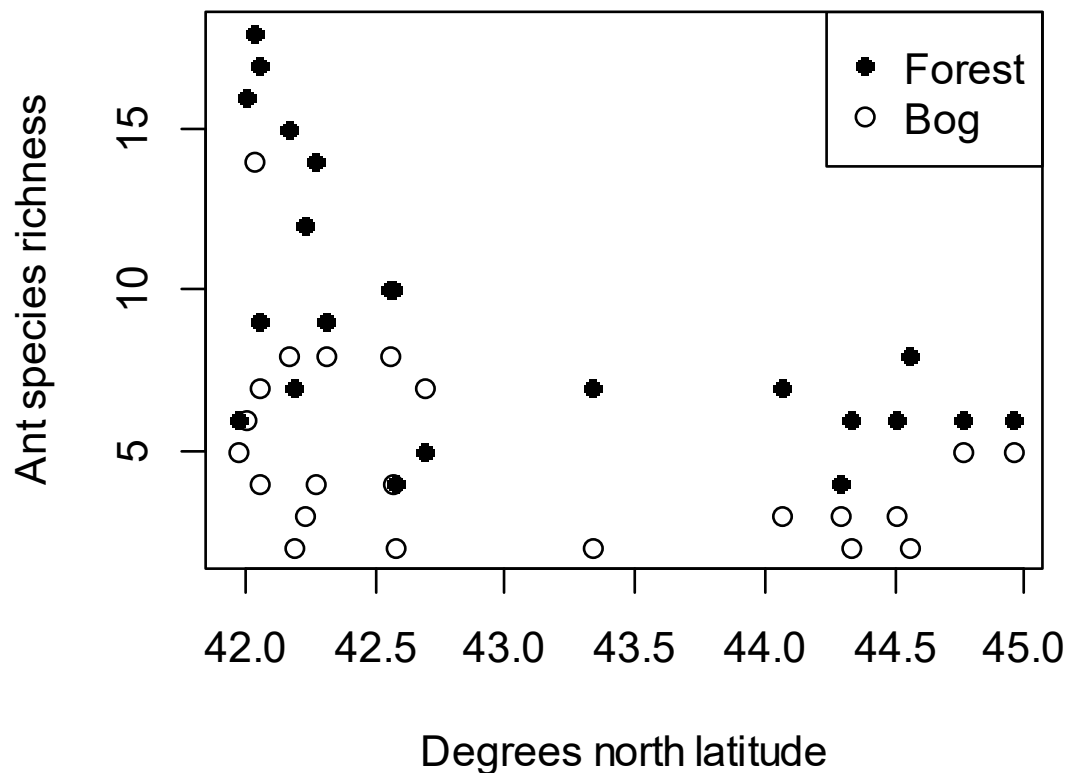$$\mu_i = \frac{e^{\eta_i}}{1 + e^{\eta_i}}$$

Dataset to analyze

How different is species richness between habitats?

How does species richness vary with latitude?

Is this relationship different between habitats?

What will the data-generating model be?
Ignore pairs for now