Bayesian Regression

Bayesian Models for Ecologists

Becky Tang

June 07, 2024

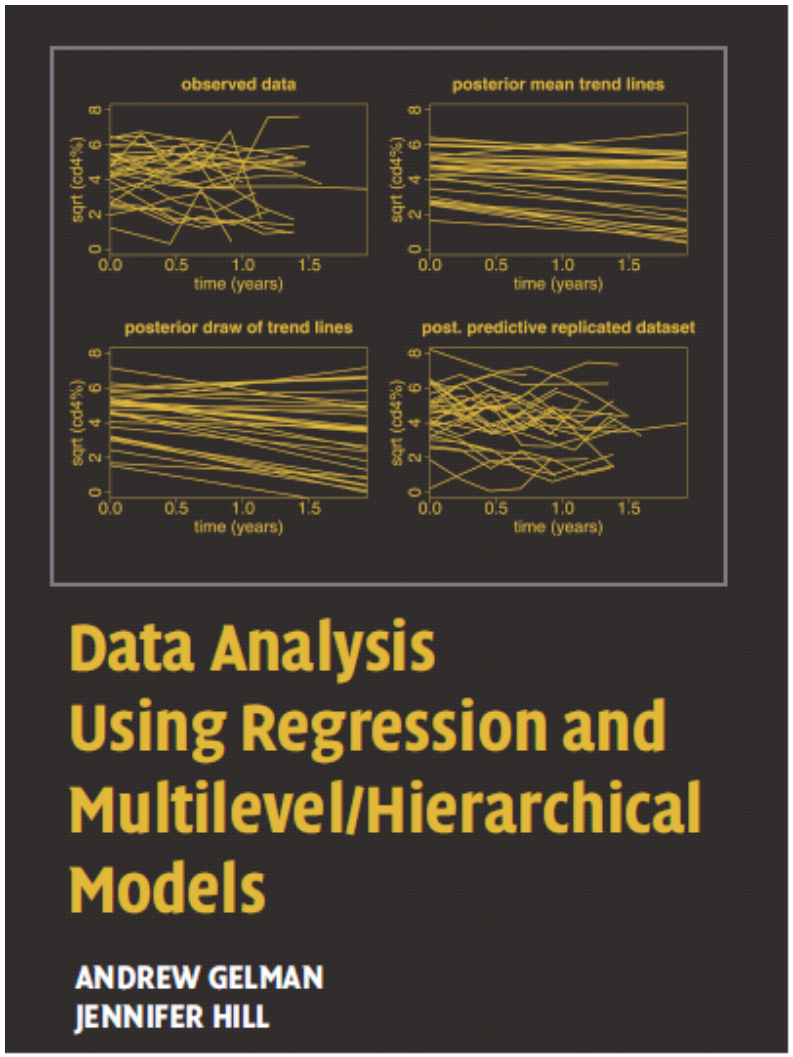
DEB 2042028

Outline

- Be able to write proper Bayesian regression models for different types of data.
- Appreciate one-to-one relationship between math and JAGS code.
- Be able to interpret coefficients of general linear models.

A great follow-up

This book should be in your library:



The general Bayesian set-up

Recall that the posterior distribution of the unobserved quantities conditional on the observed ones is proportional to their joint distribution:

$$[\theta, \sigma^2 | \mathbf{y}] \propto [\theta, \sigma^2, \mathbf{y}].$$

The joint distribution can be factored into a likelihood and priors for simple Bayesian models:

$$\left[\theta, \sigma^2, \mathbf{y}\right] = \left[\mathbf{y} \mid \theta, \sigma^2\right]\left[\theta, \sigma^2\right] \overset{ind.}{=} \left[\mathbf{y} \mid \theta, \sigma^2\right]\left[\theta\right]\left[\sigma^2\right]$$

A deterministic model of an ecological or socioenvironmental process is embedded in the likelihood like this:

$$\left[\boldsymbol{\theta}, \sigma^2, \mathbf{y}\right] \propto \left[\mathbf{y} \mid g\left(\boldsymbol{\theta}, x\right), \sigma^2\right]\left[\boldsymbol{\theta}\right]\left[\sigma^2\right]$$

Simple Bayesian regression models

We use likelihood to connect the underlying process to data:

$$\underbrace{[y_i \mid \mu_i, \sigma^2]}_{\text{stochastic model}}, \quad i = 1, \ldots, n$$

We formulate the deterministic model:

$$\mu_i = \underbrace{g(\beta, \mathbf{x}_i)}_{\text{deterministic model}}, \quad i = 1, \ldots, n$$

where β is a vector of regression coefficients and $\mathbf{x}_i$ is a vector of predictor variables.

Assuming conditional independence of the data,

$$[\beta, \sigma^2 \mid \mathbf{y}] \propto \prod_{i=1}^{n} [y_i \mid g(\beta, x_i), \sigma^2] \times [\beta] [\sigma^2]$$

We choose appropriate deterministic functions (linear or non-linear) and appropriate probability distributions to compose specific models.

Identical notation

$$y_i = g(\beta, x_i) + \varepsilon_i$$
$$\varepsilon_i \sim \text{Normal}(0, \sigma^2)$$

is the same as:

$$y_i | \beta, \sigma^2 \sim \text{Normal}\big(g(\beta, x_i), \sigma^2\big),$$

but the second notation is much more flexible because it generalizes to distributions that do not have additive errors.

You don't have to be Normal!

Assuming you have one predictor x:

Data (y-values)	**Distribution**	**"Mean" function**	**Link**
continuous, real valued	Normal	$\mu = \beta_0 + \beta_1 x$	NA (i.e. identity)
discrete, strictly positive	Poisson	$\mu = e^{\beta_0+\beta_1 x}$	$\log(\mu) = \beta_0 + \beta_1 x$
0 or 1	Bernoulli	$\mu = \frac{\exp(\beta_0+\beta_1 x)}{\exp(\beta_0+\beta_1 x)+1}$	$\text{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 x$
$[0,1]$	Beta	$\mu = \frac{\exp(\beta_0+\beta_1 x)}{\exp(\beta_0+\beta_1 x)+1}$	$\text{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 x$
continuous, strictly positive, variance $\uparrow$ as a f(mean)	lognormal	$\mu = e^{\beta_0+\beta_1 x}$	$\log(\mu) = \beta_0 + \beta_1 x$
continuous, strictly positive, constant variance	Gamma	$\mu = e^{\beta_0+\beta_1 x}$	$\log(\mu) = \beta_0 + \beta_1 x$

Continuous and real valued data

Suppose you have collected some continuous data $\mathbf{y} = (-10.7, -4.3, \cdots, 49)$ at n sites, along with a predictor x_i measured at each site i which you believe is likely to affect these measurements. Write a model regressing y on x as follows:

1. Choose a specific stochastic and deterministic model.
2. Specify (vague) priors for your parameters.
3. Write out the DAG and express posterior distribution as proportional to joint distribution for your model.
4. Write the JAGS code for the model.
5. Interpret the coefficients of your model.

Normal data, continuous and real valued

Stochastic model:

$$y_i|\beta_0,\beta_1,\sigma^2 \overset{ind.}{\sim} \text{Normal}(g(\beta_0,\beta_1,x_i),\sigma^2) \qquad i=1,...,n$$

Deterministic model:

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

Priors:

$$\beta_0 \sim ? \qquad [\beta_0] = ?$$
$$\beta_1 \sim ? \qquad [\beta_1] = ?$$
$$\sigma^2 \sim ? \qquad [\sigma^2] = ?$$

Normal data, continuous and real valued

DAG:

Posterior distribution:

$$
\begin{aligned}
\left[\beta_0, \beta_1, \sigma \mid \mathbf{y}\right] &\propto [\beta_0, \beta_1, \sigma, \mathbf{y}] \\
&\propto [\mathbf{y}|\beta_0, \beta_1, \sigma][\beta_0][\beta_1][\sigma] \\
&\propto \prod_{i=1}^{n} \text{Normal}\left(y_i \mid g\left(\beta_0, \beta_1, x_i\right), \sigma^2\right) \\
&\quad \times \text{Normal}\left(\beta_0 \mid 0, 1000\right) \text{Normal}\left(\beta_1 \mid 0, 1000\right) \\
&\quad \times \text{uniform}\left(\sigma \mid 0, 100\right) \\
g\left(\beta_0, \beta_1, x_i\right) &= \beta_0 + \beta_1 x_i
\end{aligned}
$$

Normal data, continuous and real valued

JAGS code for the model:

```
b0 ~ dnorm(0, .001)
b1 ~ dnorm(0, .001)
sigma ~ dunif(0, 100)
tau <- 1/sigma^2
for (i in 1:length(y)){
  mu[i] <- b0 + b1 * x[i]
  y[i] ~ dnorm(mu[i], tau)
}
```

Interpretation:

- β_0 : expected outcome when $x = 0$
- β_1 : average change in the outcome for a one-unit change in x
- σ : std deviation of the outcomes about their respective means

Counts, discrete and non-negative

You have collected some count data ($y = 12, 17, 1, 0, 31, \ldots, 25$) at n sites, along with a covariate x_i at each location which you believe is likely to affect these counts. Write a model regressing y on x.

1. Choose a specific stochastic and deterministic model.
2. Specify (vague) priors for your parameters.
3. Write out the DAG and express posterior distribution as proportional to joint distribution for your model.
4. Write the JAGS code for the model.
5. Interpret the coefficients of your model.

Poisson, discrete and non-negative

Stochastic model:

$$y_i|\beta_0, \beta_1 \overset{ind.}{\sim} \text{Poisson}(g(\beta_0, \beta_1, x_i)) \qquad i = 1, ..., n$$

Deterministic model:

$$\mu_i = g(\beta_0, \beta_1, x_i) = e^{\beta_0 + \beta_1 x_i}$$

Priors:

$$\beta_0 \sim ? \qquad [\beta_0] = ?$$
$$\beta_1 \sim ? \qquad [\beta_1] = ?$$

Poisson, discrete and non-negative

DAG:

Posterior distribution:

$$
\begin{aligned}
\left[\beta_0, \beta_1 \mid \mathbf{y}\right] &\propto \prod_{i=1}^{n} \text{Poisson}\left(y_i \mid g\left(\beta_0, \beta_1, x_i\right)\right) \\
&\quad \times \text{Normal}\left(\beta_0 \mid 0, 1000\right) \text{Normal}\left(\beta_1 \mid 0, 1000\right) \\
g\left(\beta_0, \beta_1, x_i\right) &= e^{\beta_0 + \beta_1 x_i}
\end{aligned}
$$

Poisson, discrete and non-negative

JAGS code:

```
b0 ~ dnorm(0, .001)
b1 ~ dnorm(0, .001)
for(i in 1:length(y)){
  log(mu[i]) <- b0 + b1 * x[i]
  y[i] ~ dpois(mu[i])
}
```

or

```
  mu[i] <- exp(b0 + b1 * x[i])
  y[i] ~ dpois(mu[i])
```

Poisson, discrete and non-negative

$$
\begin{aligned}
[\beta_0, \beta_1 \mid \mathbf{y}] &\propto \prod_{i=1}^{n} \text{Poisson}\left(y_i \mid g\left(\beta_0, \beta_1, x_i\right)\right) \\
&\quad \times \text{Normal}\left(\beta_0 \mid 0, 1000\right) \text{Normal}\left(\beta_1 \mid 0, 1000\right) \\
\mu_i = g\left(\beta_0, \beta_1, x_i\right) &= e^{\beta_0 + \beta_1 x_i} = e^{\beta_0} e^{\beta_1 x_i}
\end{aligned}
$$

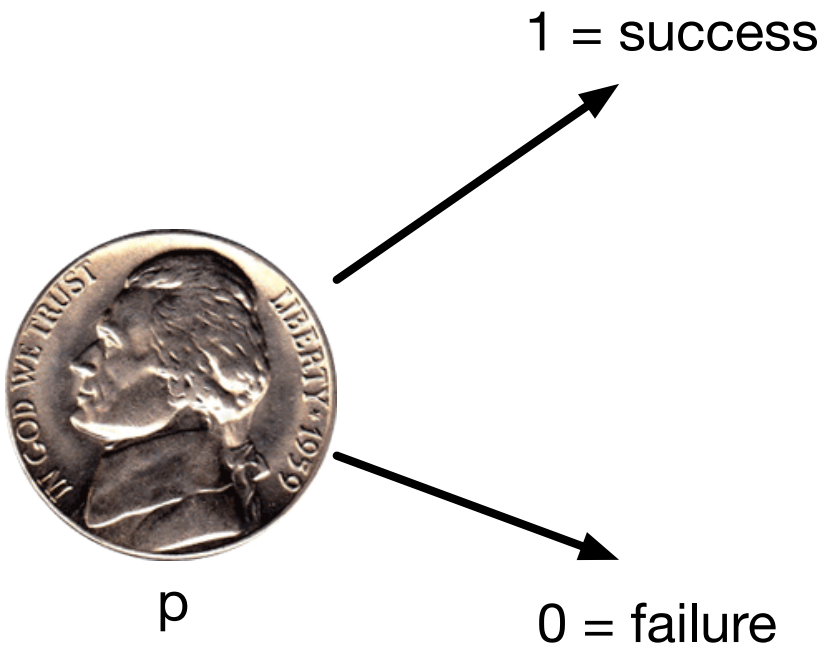
Interpretation: (Exponentiate coeff. and report multiplicative change in mean counts.)

- e^{β_0}: average count when $x = 0$
- e^{β_1}: multiplicative change in the mean count per one unit change in x

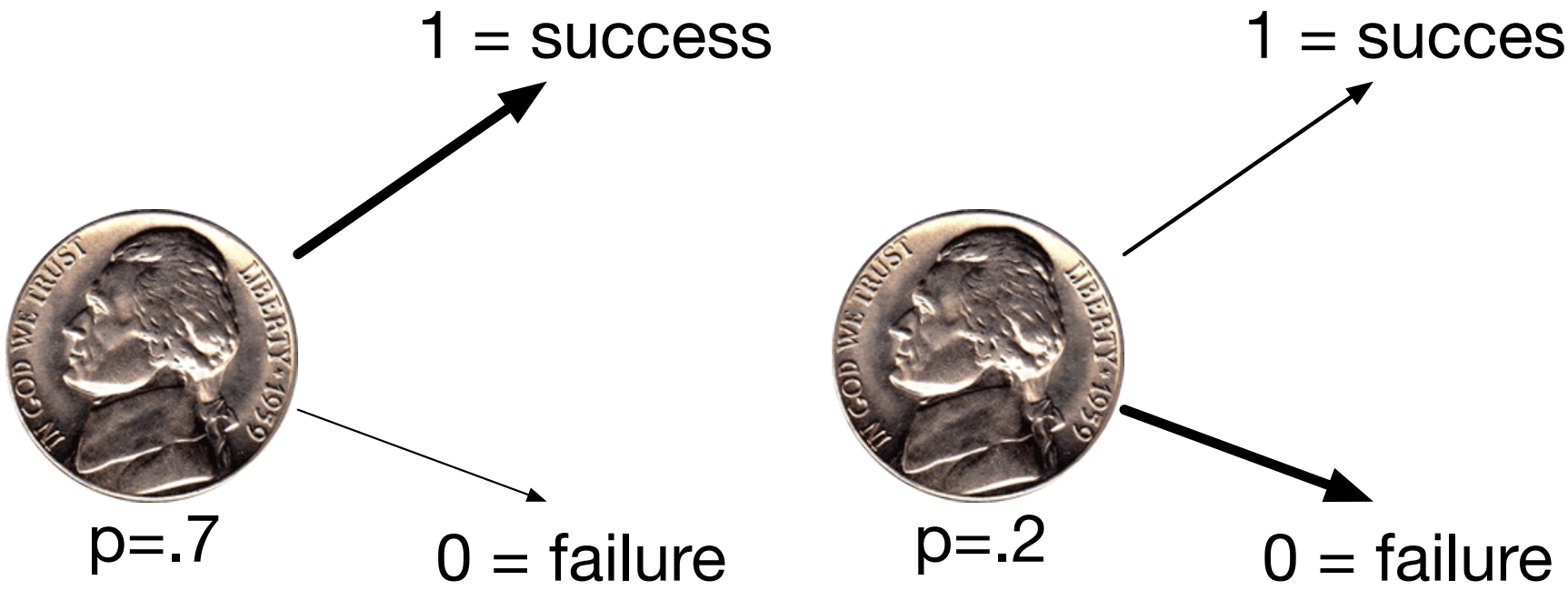
For example: "Mean western toad juvenile abundance is reduced by a factor of 5.1 (95% CI: 3.4, 10.8) per unit change in UV-B radiation."

What was the estimate of β_1?

Binary data, 0 or 1 (aka logistic)



Binary data, 0 or 1 (aka logistic)



Binary data, 0 or 1 (aka logistic)

What happens when we want to relate p to a predictor x, where x can be any value on the real line? How do we connect x to $p \in [0,1]$?

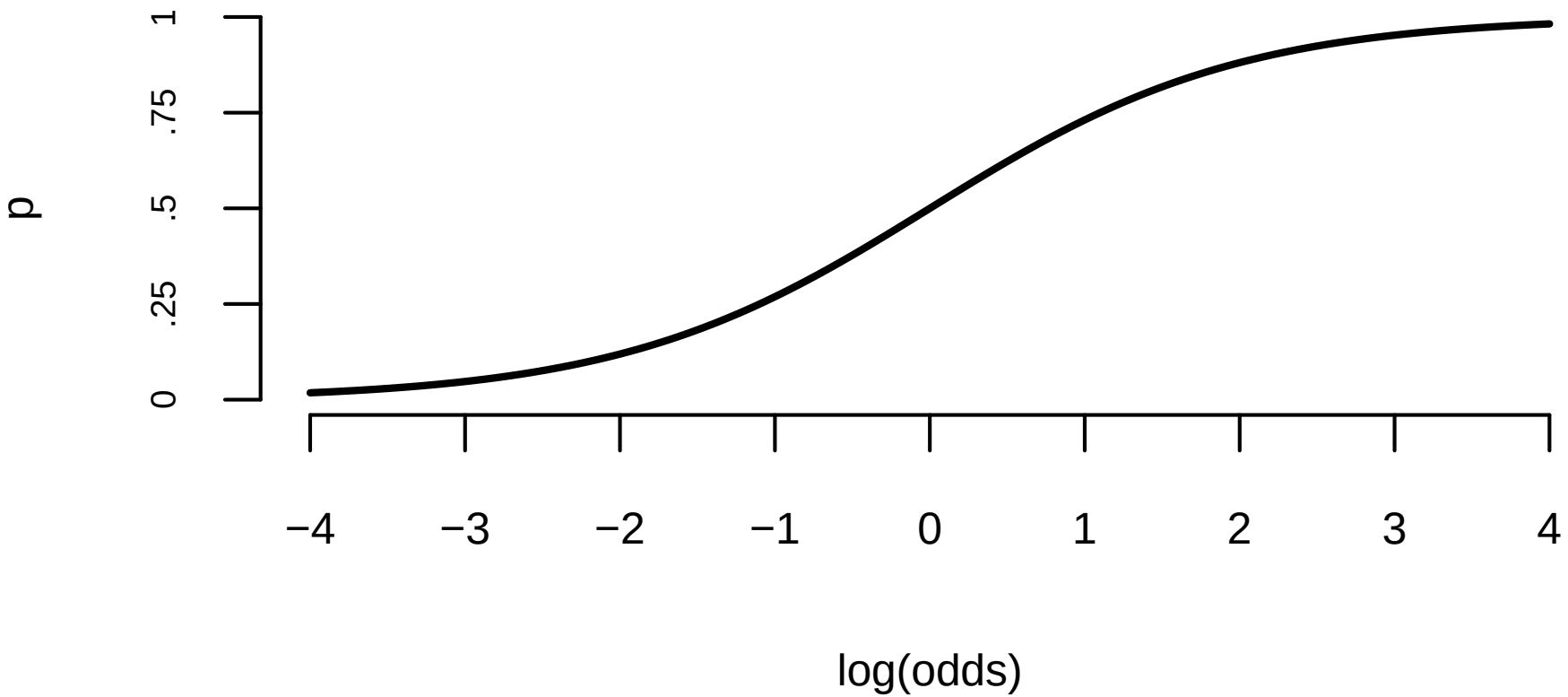
- odds: $\frac{p}{1-p} \in [0,\infty)$
- log odds: $\log(\text{odds}) = \log\left(\frac{p}{1-p}\right) \in (-\infty,\infty)$

Moving between probability and log odds and relating to x:

- $\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = x$
  - input to logit() is probability p, output is log odds x
- Inverting the above, we obtain $p = \frac{e^x}{e^x+1} =$ inverse $\text{logit}(x) = \text{expit}(x)$
  - input to inverse logit() is $x =$ log odds, output is probability p

Binary data, 0 or 1 (aka logistic)

Inverse logit mapping: input is log odds $= \log\left(\frac{p}{1-p}\right)$, output is probability



Binary data, 0 or 1 (aka logistic)

You have collected some binary data ($y = 1, 0, 0, 1, 1, 0, 1, \ldots, 1$) at n sites, along with a covariate x_i measured at each location which you believe is likely to affect these counts. Write a model regressing y on x.

1. Choose a specific stochastic and deterministic model.
2. Specify (vague) priors for your parameters.
3. Write out the DAG and express posterior distribution as proportional to joint distribution for your model.
4. Write the JAGS code for the model.
5. Interpret the coefficients of your model.

Bernoulli, 0 or 1 (aka logistic)

Stochastic model:

$$y_i|\beta_0, \beta_1 \overset{\perp}{ind}.\text{Bernoulli}(g(\beta_0, \beta_1, x_i)) \qquad i = 1, ..., n$$

Deterministic model:

$$\mu_i = p_i = g(\beta_0, \beta_1, x_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{e^{\beta_0 + \beta_1 x_i} + 1}$$

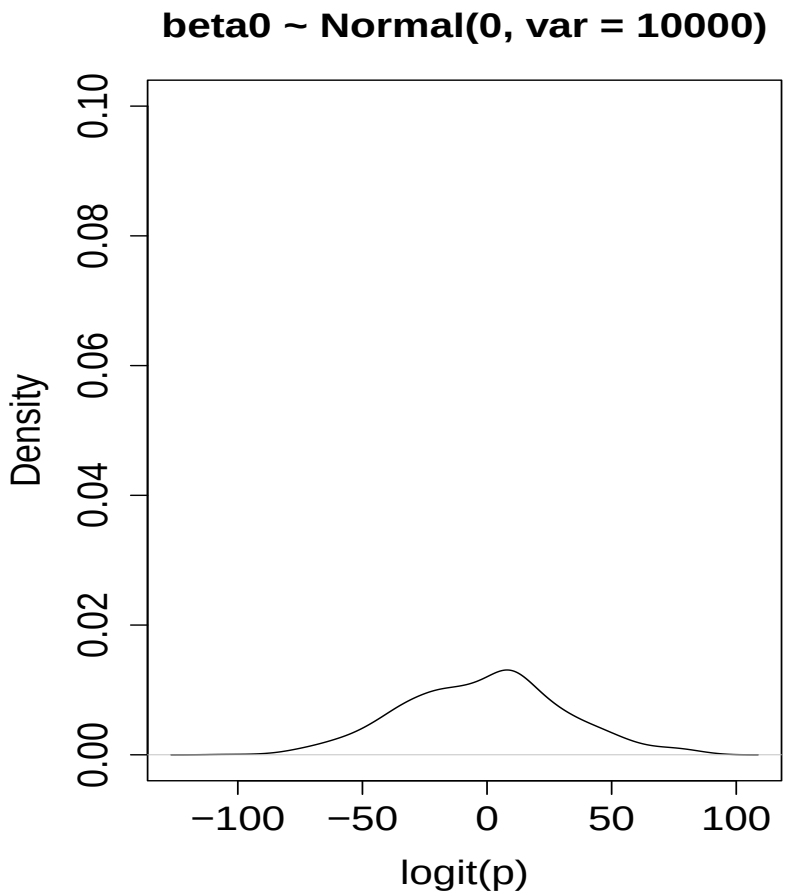
Priors:

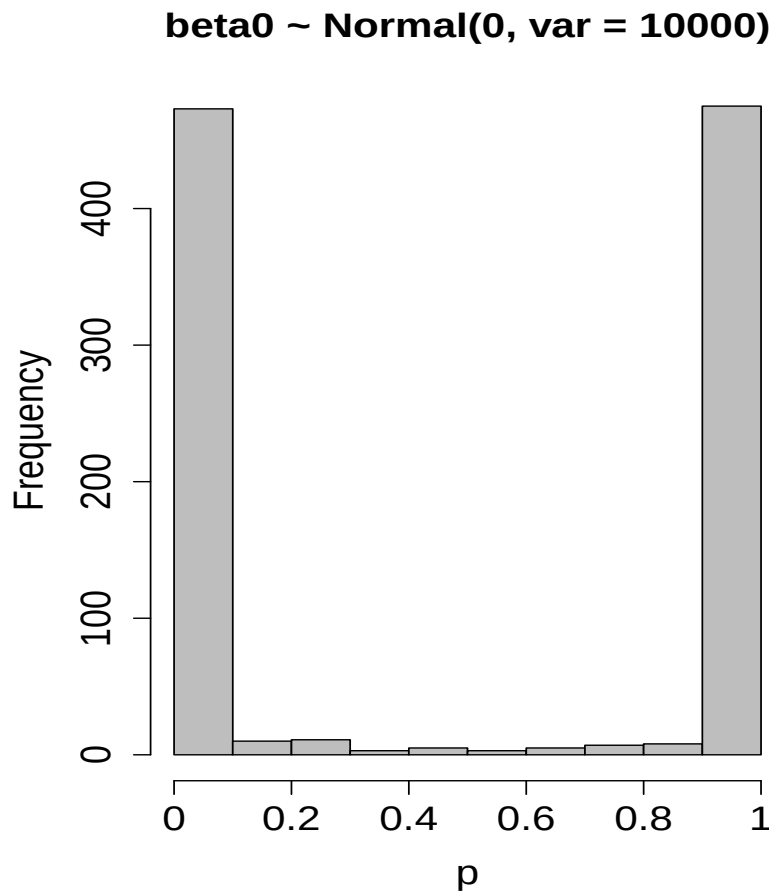
$$\beta_0 \sim ? \qquad [\beta_0] = ?$$
$$\beta_1 \sim ? \qquad [\beta_1] = ?$$

Choosing reasonable flat priors on logit intercept

Imagine for now that we have no predictor, so $p_i = \mu_i = \frac{e^{\beta_0}}{e^{\beta_0}+1}$.

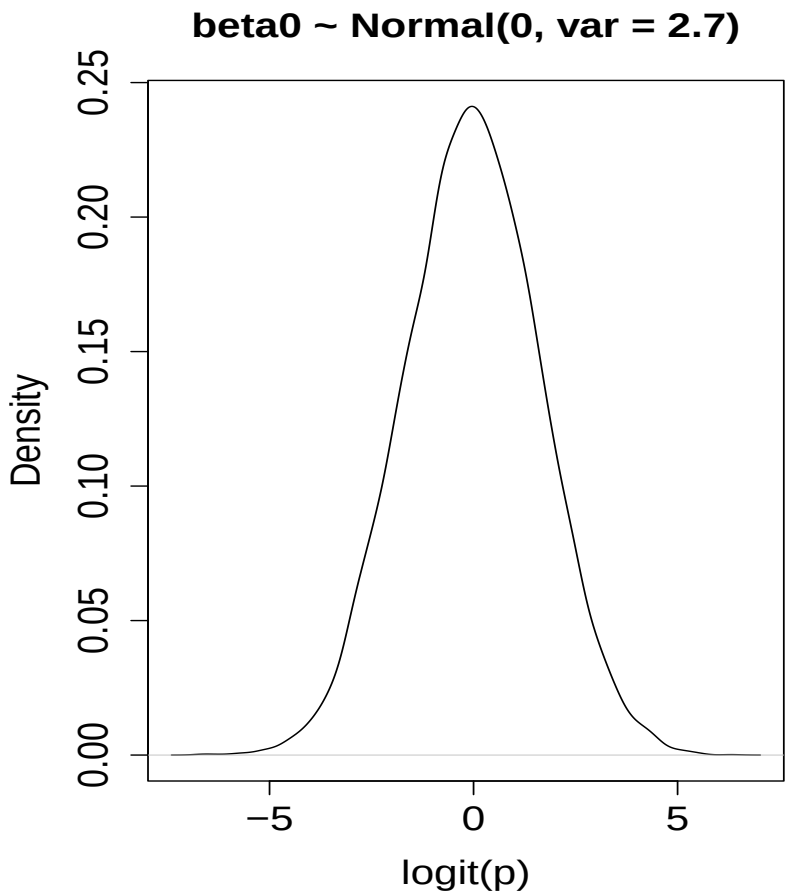
If we use the same Normal(0, large variance) prior as before, what is the induced prior for the success probability p?

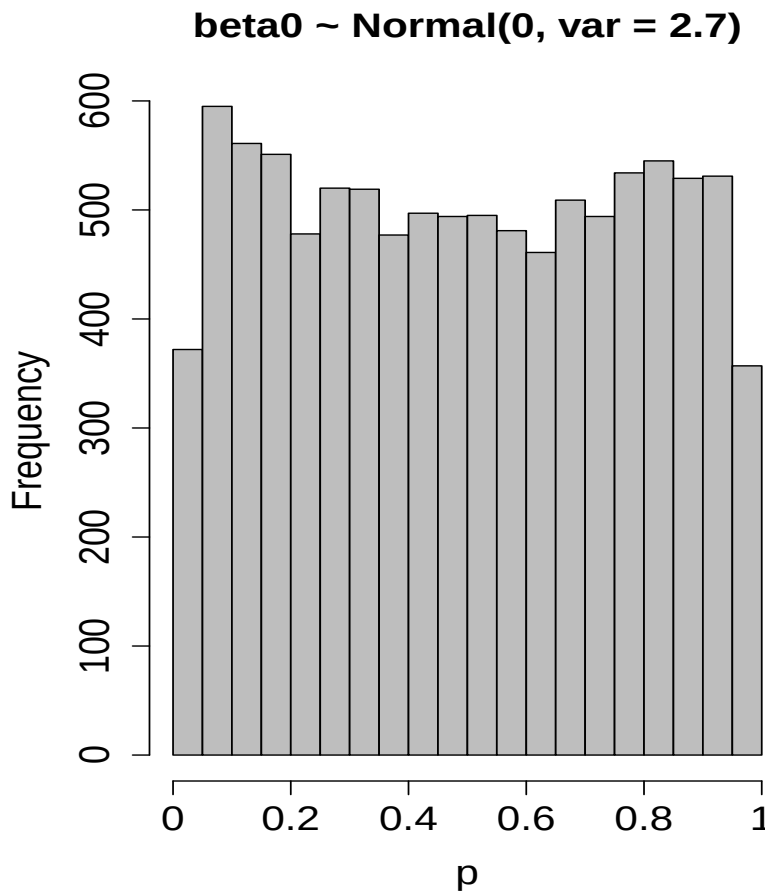




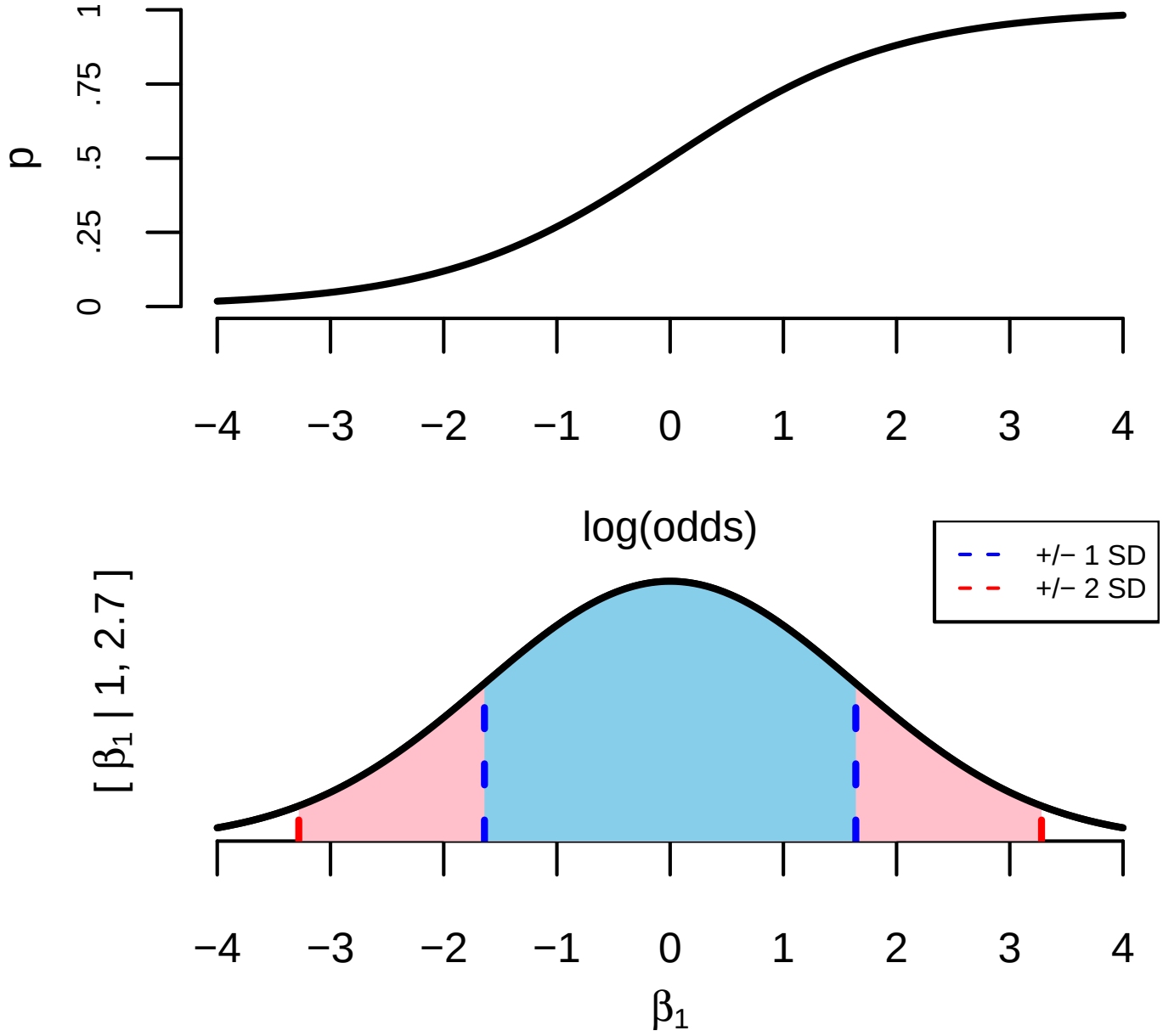
Choosing reasonable flat priors on logit intercept

Now instead consider the prior $\beta_0 \sim \text{Normal}(0, 2.7)$.





Choosing reasonable flat priors on logit effects



Bernoulli, 0 or 1 (aka logistic)

Returning to case with a single predictor, the posterior distribution is:

$$
\begin{aligned}
\left[\beta_0, \beta_1 \mid \mathbf{y}\right] &\propto \prod_{i=1}^{n} \text{Bernoulli}\left(y_i \mid g\left(\beta_0, \beta_1, x_i\right)\right) \\
&\quad \times \text{Normal}\left(\beta_0 \mid 0, 2.7\right) \text{Normal}\left(\beta_1 \mid 0, 2.7\right) \\
g\left(\beta_0, \beta_1, x_i\right) &= \frac{e^{\beta_0 + \beta_1 x_i}}{e^{\beta_0 + \beta_1 x_i} + 1}
\end{aligned}
$$

Bernoulli, 0 or 1 (aka logistic)

JAGS code for the model:

```
b0 ~ dnorm(0, 1/2.7)
b1 ~ dnorm(0, 1/2.7)
for(i in 1:length(y)){
  logit(p[i]) <- b0 + b1 * x[i]
  y[i] ~ dbern(p[i])
}
```

or

```
  p[i] <- inv.logit(b0 + b1 * x[i])
  y[i] ~ dbern(p[i])
```

Bernoulli, 0 or 1 (aka logistic)

$$\begin{aligned}[\beta_0, \beta_1 \mid \mathbf{y}] \quad &\propto \quad \prod_{i=1}^{n} \text{Bernoulli}\left(y_i \mid g\left(\beta_0, \beta_1, x_i\right)\right) \\ &\quad \times \text{Normal}\left(\beta_0 \mid 0, 2.7\right) \text{Normal}\left(\beta_1 \mid 0, 2.7\right)\end{aligned}$$

$$p_i = g\left(\beta_0, \beta_1, x_i\right) = \frac{e^{\beta_0 + \beta_1 x_i}}{e^{\beta_0 + \beta_1 x_i} + 1} \qquad \Longleftrightarrow \qquad \frac{p_i}{1 - p_i} = e^{\beta_0 + \beta_1 x_i} = e^{\beta_0} e^{\beta_1 x_i}$$

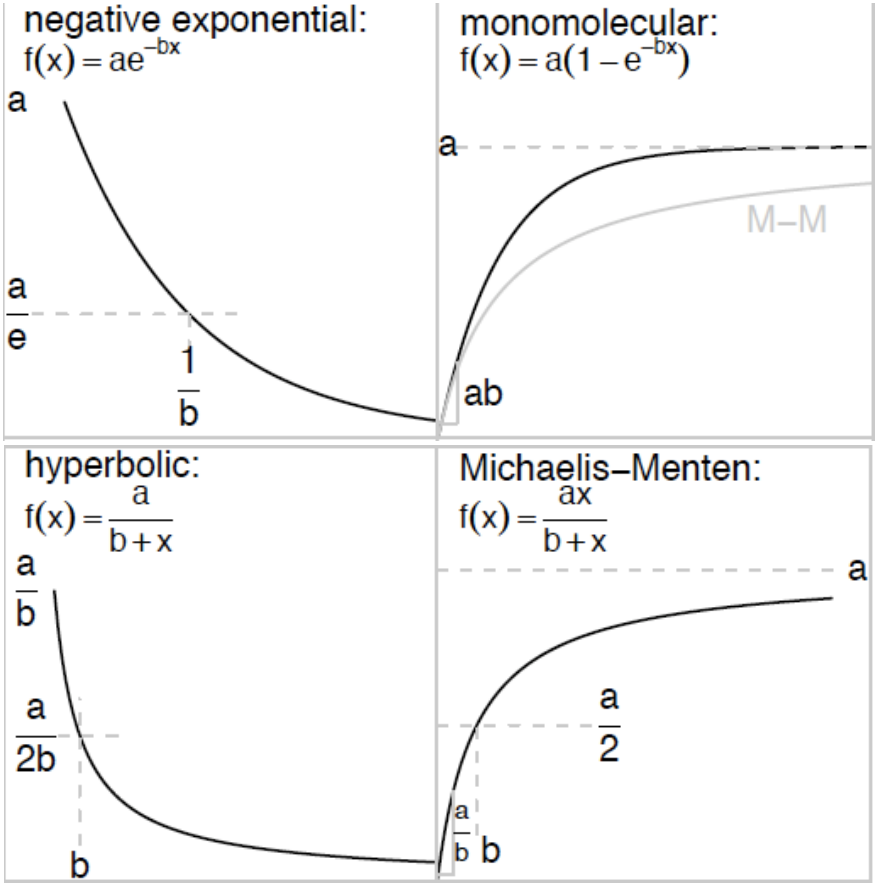
Interpretation: (Exponentiate coef. and report odds and odds ratios.)

- e^{β_0} : odds when $x = 0$
- e^{β_1} : multiplicative change in the odds for a one unit change in x

For example: "the odds of detecting weevils in upland willow stems were 3.2 (95% CI: 2.3, 4.8) times greater than detecting them in riparian willow stems."

What was the estimate of β_1? (Might be helpful to identify/define x_i).

Nonlinear regression



{Figures c/o Bolker, B. 2008. *Ecological Models and Data in R*. Princeton University Press, Princeton, NJ. USA.}

Centering and standardizing

The remainder of the slides apply to all of the general linear models, but we will use a simple linear model for Normally distributed data as an example.

Centering predictor data

Rather than the usual linear model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ where $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, consider the following linear model:

$$y_i = \beta_0 + \beta_1 \left(x_i - \bar{x} \right) + \varepsilon_i$$

where $\bar{x} = \sum_{i=1}^{n} x_i$ is the sample mean of the predictor.

Why complicate things?

- To reduce autocorrelation in MCMC chain and speed convergence.
- To make the intercept more easily interpretable.

Centering predictor data

$$
\begin{aligned}
[\beta_0, \beta_1, \sigma \mid \mathbf{y}] &\propto \prod_{i=1}^{n} \text{Normal}\left(y_i \mid g\left(\beta_0, \beta_1, x_i\right), \sigma^2\right) \times \\
&\quad \text{Normal}\left(\beta_0 \mid 0, 1000\right) \text{Normal}\left(\beta_1 \mid 0, 1000\right) \times \\
&\quad \text{uniform}\left(\sigma \mid 0, 100\right) \\
g\left(\beta_0, \beta_1, x_i\right) &= \beta_0 + \beta_1 \left(x_i - \bar{x}\right)
\end{aligned}
$$

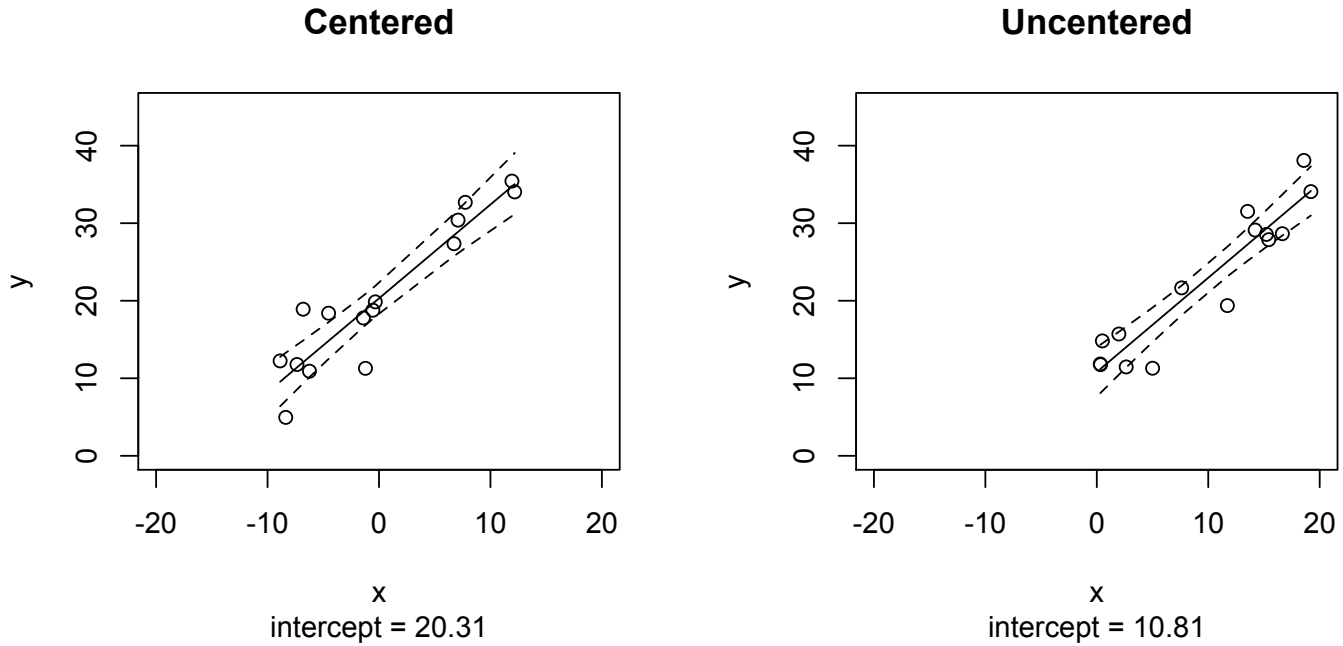
```
b0 ~ dnorm(0, .001)
b1 ~ dnorm(0, .001)
sigma ~ dunif(0, 100)
tau <- 1/sigma^2
xBar <- mean(x)
for (i in 1:length(y)){
  mu[i] <- b0 + b1 * (x[i] - xBar)
  y[i] ~ dnorm(mu[i], tau)
}
b0_UC <- b0 - b1 * xBar
```

Exercise

What is the interpretation of β_0 in this model?

What is the interpretation of β_1 in this model?

Recovering uncentered parameters



$$\begin{aligned} \beta_0^* &= \beta_0 - \beta_1 \bar{x} \\ \beta_1^* &= \beta_1 \end{aligned}$$

- For this to work properly, all the coefficients in the model must be *added*.
- Slopes will not be the same if there is an interaction term or quadratic. In these cases, back transforming is not simple.

Standardizing predictor data

$$y_i = \beta_0 + \beta_1 \left(\frac{x_i - \bar{x}}{s_x} \right)$$

where s_x is sample standard deviation of predictor
($s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$)

Why complicate things?

- To reduce autocorrelation in MCMC chain and speed convergence.
- To make the intercept more easily interpretable.
- To make coefficients more easily comparable

Standardizing predictor data

$$
\begin{aligned}
[\beta_0, \beta_1, \sigma \mid \mathbf{y}] &\propto \prod_{i=1}^{n} \text{Normal}\left(y_i \mid g\left(\beta_0, \beta_1, x_i\right), \sigma^2\right) \times \\
&\quad \text{Normal}\left(\beta_0 \mid 0, 1000\right) \text{Normal}\left(\beta_1 \mid 0, 1000\right) \times \\
&\quad \text{uniform}\left(\sigma \mid 0, 100\right) \\
g\left(\beta_0, \beta_1, x_i\right) &= \beta_0 + \beta_1 \left(\frac{x_i - \bar{x}}{s_x}\right)
\end{aligned}
$$

```
b0 ~ dnorm(0, .001)
b1 ~ dnorm(0, .001)
sigma ~ dunif(0, 100)
tau <- 1/sigma^2
xBar <- mean(x)
xSD <- sd(x)
for (i in 1:length(y)){
  mu[i] <- b0 + b1 * ((x[i] - xBar)/xSD
  y[i] ~ dnorm(mu[i], tau)
}
```

Recovering unstandardized parameters

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 \left(\frac{x_i - \bar{x}}{s_x} \right) \\
&= \beta_0 + \frac{\beta_1}{s_x} x_i - \frac{\beta_1 \bar{x}}{s_x} \\
&= \beta_0^* + \beta_1^* x_i \\
\beta_0^* &= \beta_0 - \frac{\beta_1 \bar{x}}{s_x} \\
\beta_1^* &= \frac{\beta_1}{s_x}
\end{aligned}
$$

- This only works if there are not squared values or interactions.
- It is fine to make predictions using $\hat{y}_i = \beta_0 + \beta_1 \frac{x_i - \bar{x}}{s_x}$ and plot $\hat{y}_i$ against x_i and the observed y_i

lognormal, data continuous and > 0 (log link)

$$
\begin{aligned}
[\beta_0, \beta_1, \sigma \mid \mathbf{y}] &\propto \prod_{i=1}^{n} \text{lognormal}\left(y_i \mid \log\left(g\left(\beta_0, \beta_1, x_i\right)\right), \sigma^2\right) \\
&\quad \times \text{Normal}\left(\beta_0 \mid 0, 1000\right) \text{Normal}\left(\beta_1 \mid 0, 1000\right) \\
&\quad \times \text{uniform}\left(\sigma \mid 0, 5\right) \\
g\left(\beta_0, \beta_1, x_i\right) &= e^{\beta_0 + \beta_1 x_i}
\end{aligned}
$$

Talk about the interpretation of σ.

```
b0 ~ dnorm(0, .001)
b1 ~ dnorm(0, .001)
sigma ~ dunif(0, 5)
tau <- 1/sigma^2
for(i in 1:length(y)){
  mu[i] <- exp(b0 + b1 * x[i])
  y[i] ~ dlnorm(log(mu[i]), tau)
}
```

lognormal, data continuous and > 0 (not log link)

$$
\begin{aligned}
\left[\beta_0, \beta_1, \sigma \mid \mathbf{y}\right] &\propto \prod_{i=2}^{n} \text{lognormal}\left(y_i \mid \log\left(g\left(\beta_0, \beta_1, y_{i-1}\right)\right), \sigma^2\right) \\
&\quad \times \text{Normal}\left(\beta_0 \mid 0, 1000\right) \text{Normal}\left(\beta_1 \mid 0, 1000\right) \\
&\quad \times \text{uniform}\left(\sigma \mid 0, 5\right) \text{uniform}\left(y_1 \mid 1, 1E6\right) \\
g\left(\beta_0, \beta_1, y_{i-1}\right) &= y_{i-1} e^{\beta_0 + \beta_1 y_{i-1}}
\end{aligned}
$$

```
b0 ~ dnorm(0, .001)
b1 ~ dnorm(0, .001)
sigma ~ dunif(0, 5); tau <- 1/sigma^2
y[1] ~ dunif(1, 1E6)
for(i in 2:length(y)){
  mu[i] <- y[i-1] * exp(b0 + b1 * y[i-1])
  y[i] ~ dlnorm(log(mu[i]), tau)
}
```