Bayesian Models for More Complex Data

Recall: Regression

Regression modeling is concerned with describing how the sampling distribution of one random variable

$$Y$$

varies with another variable or set of variables

$$x = (x_1, \ldots, x_p)$$

Recall: Regression

Regression modeling is concerned with describing how the sampling distribution of one random variable

Y response variable

varies with another variable or set of variables

$x = (x_1, \ldots, x_p)$ explanatory variable(s) or "covariates"

Recall: Regression

Regression modeling is concerned with describing how the sampling distribution of one random variable

Y response variable

varies with another variable or set of variables

$x = (x_1, \ldots, x_p)$ explanatory variable(s) or "covariates"

Specifically, a regression model postulates a form for $p(y|x)$, the conditional distribution for Y given x

Recall: Regression

Regression modeling is concerned with describing how the sampling distribution of one random variable

Y response variable

varies with another variable or set of variables

$x = (x_1, \ldots, x_p)$ explanatory variable(s) or "covariates"

Specifically, a regression model postulates a form for $p(y|x)$, the conditional distribution for Y given x

Estimation of $p(y|x)$ is made using data $y_1, \ldots, y_n$ gathered under a variety of conditions $x_1, \ldots, x_n$

Linear model

One simple but flexible approach to regression is via the linear (sampling) model (LM)

Linear model

One simple but flexible approach to regression is via the linear (sampling) model (LM)

The LM treats responses Y_i as independent (but not identically distributed) realizations of a process that is linear in explanatory variables $x_i^\top = (x_{i,1}, \ldots, x_{i,p})$, observed with Gaussian noise

$$Y_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu_i, \sigma^2) \quad \text{where} \quad \mu_i = \mathbb{E}\{Y|x_i\} = x_i^\top \beta$$

Linear model

One simple but flexible approach to regression is via the linear (sampling) model (LM)

The LM treats responses Y_i as independent (but not identically distributed) realizations of a process that is linear in explanatory variables $x_i^\top = (x_{i,1}, \ldots, x_{i,p})$, observed with Gaussian noise

$$Y_i \stackrel{\text{ind}}{\sim} \mathcal{N}(\mu_i, \sigma^2) \quad \text{where} \quad \mu_i = \mathbb{E}\{Y|x_i\} = x_i^\top \beta$$

where the x_i are known, β is an unknown p-dimensional parameter vector of regression coefficients, and σ^2 is an unknown variance parameter

Compact notation

The LM is usually written as $Y = X\beta + \varepsilon$, where

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_1^\top \\ \vdots \\ x_n^\top \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

and $\{\varepsilon_1, \ldots, \varepsilon_n\} \overset{\text{iid}}{\sim} N(0, \sigma^2)$

Compact notation

The LM is usually written as $Y = X\beta + \varepsilon$, where

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_1^\top \\ \vdots \\ x_n^\top \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

and $\{\varepsilon_1, \ldots, \varepsilon_n\} \overset{\text{iid}}{\sim} N(0, \sigma^2)$

Even more compact notation is

$$Y \sim \mathcal{N}_n(X\beta, I_n\sigma^2)$$

where I_n is a $n \times n$ identity matrix

Compact notation

The LM is usually written as $Y = X\beta + \varepsilon$, where

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_1^\top \\ \vdots \\ x_n^\top \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

and $\{\varepsilon_1, \ldots, \varepsilon_n\} \overset{\text{iid}}{\sim} N(0, \sigma^2)$

Even more compact notation is

design matrix

$$Y \sim \mathcal{N}_n(X\beta, I_n\sigma^2)$$

where I_n is a $n \times n$ identity matrix

Linear model assumptions

The linear model rests on some important assumptions:

- Errors are additive and normally distributed
- Errors are homoskdastic (don't vary across Xs)
- Observations are independent (conditional on the linear predictor)
- Linear (in covariates) mean function
- All error/randomness is in the value of the response (i.e., the X values are precisely known)
- There is no (systematic) missing data

Linear model assumptions

The linear model rests on some important assumptions:

- Errors are additive and normally distributed
- Errors are homoskdastic (don't vary across Xs)
- Observations are independent (conditional on the linear predictor)
- Linear (in covariates) mean function
- All error/randomness is in the value of the response (i.e., the X values are precisely known)
- There is no (systematic) missing data

Ecological data rarely conform to these assumptions!

Non-normal distributions

The most common deviation from these assumptions is that data are non-normal, and especially are not continuous:

- **Binary Data** (0 or 1)
  - Sick or Healthy
  - Yes or No

- **Count data** (1, 2, 3, 4...)
  - number of animals observed
  - number of people ill

Example: Estimating the probability of a rare event

Suppose we are interested in the prevalence of an infectious disease in a small city. A small random sample of 20 individuals will be checked for infection.

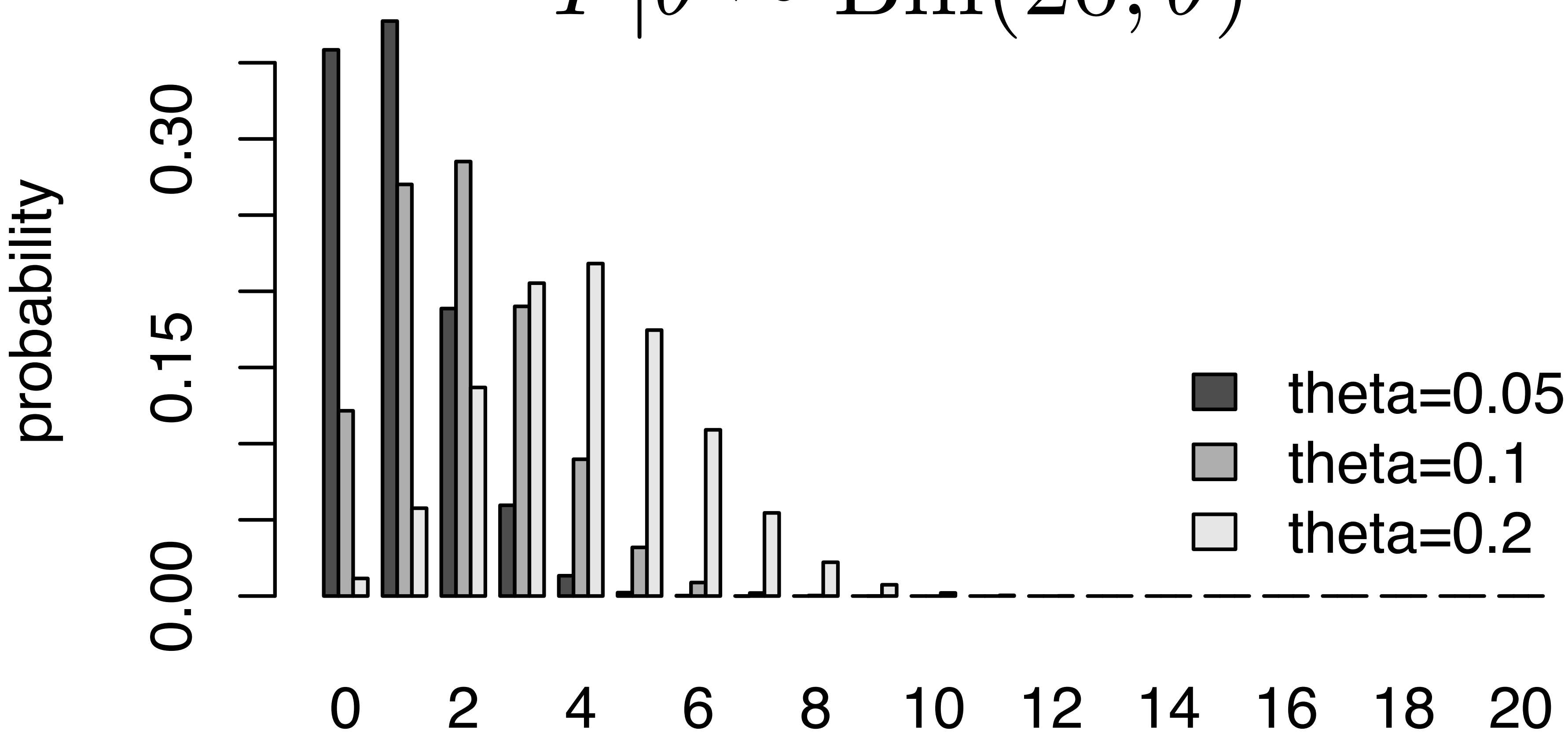
- We want to estimate the fraction of infected individuals in the population: $\theta \in \Theta = [0, 1]$
- The data records the number of infected individuals: $y \in \mathcal{Y} = \{0, 1, \ldots, 20\}$

Example: Likelihood/sampling model

Before the sample is obtained, the number of infected individuals is unknown.

- Let Y denote this to-be-determined value
- If θ were known, a sensible sampling model is

$$Y|\theta \sim \text{Bin}(20, \theta)$$

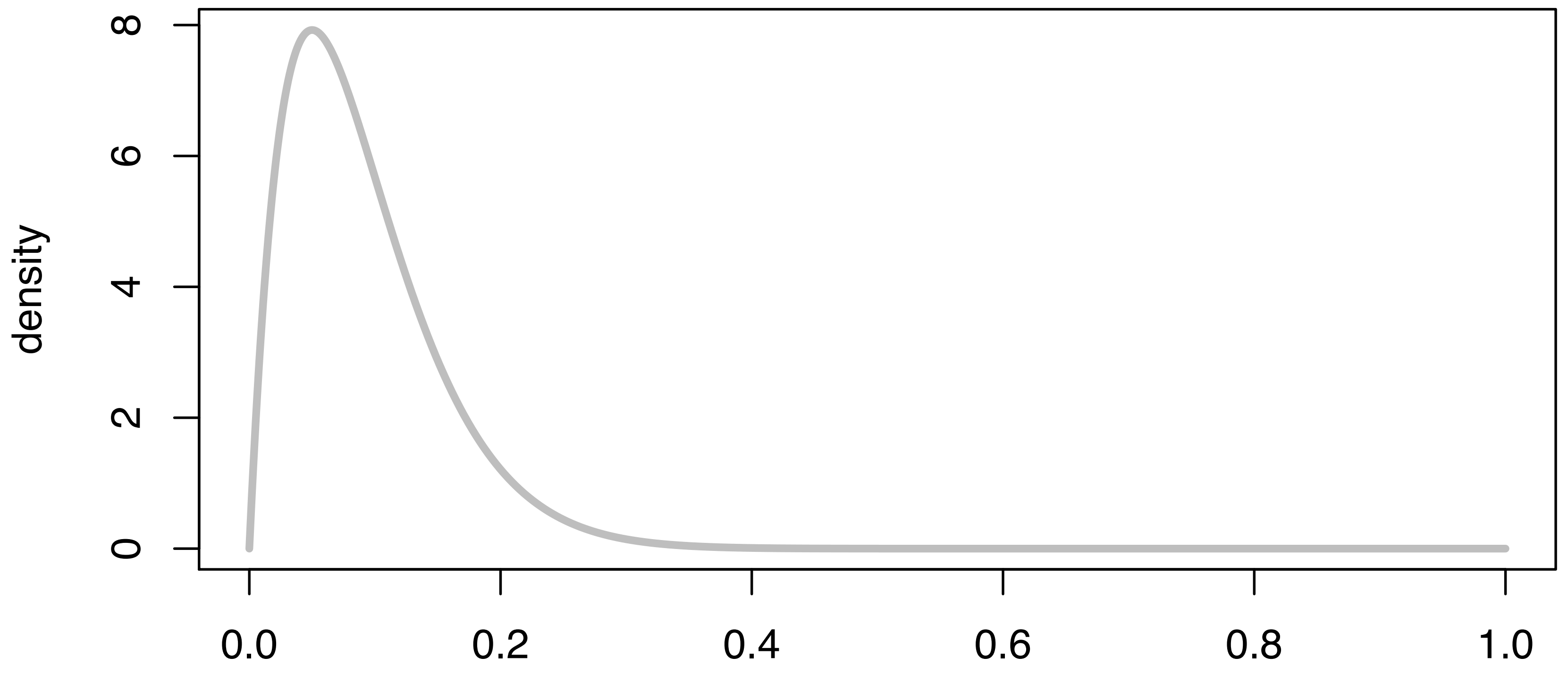


Example: Prior

Other studies from various parts of the country indicate that the infection rate ranges from about 0.05 to 0.20 with an average prevalence of 0.1

- Moment matching from a beta distribution (a convenient choice, as we'll see) give the prior:

$$\theta \sim \text{Beta}(2, 20)$$



Example: Posterior

The prior and sample model combination:

$$\theta \sim \mathrm{Beta}(a, b)$$
$$Y|\theta \sim \mathrm{Bin}(n, \theta)$$

With observed data y, we can obtain an analytic expression for the posterior:

$$p(\theta|y) = \mathrm{Beta}(a + y, b + n - y)$$

Example: Posterior

The prior and sample model combination:

$$\theta \sim \mathrm{Beta}(a, b)$$
$$Y|\theta \sim \mathrm{Bin}(n, \theta)$$

With observed data y, we can obtain an analytic expression for the posterior:

$$p(\theta|y) = \mathrm{Beta}(a + y, b + n - y)$$

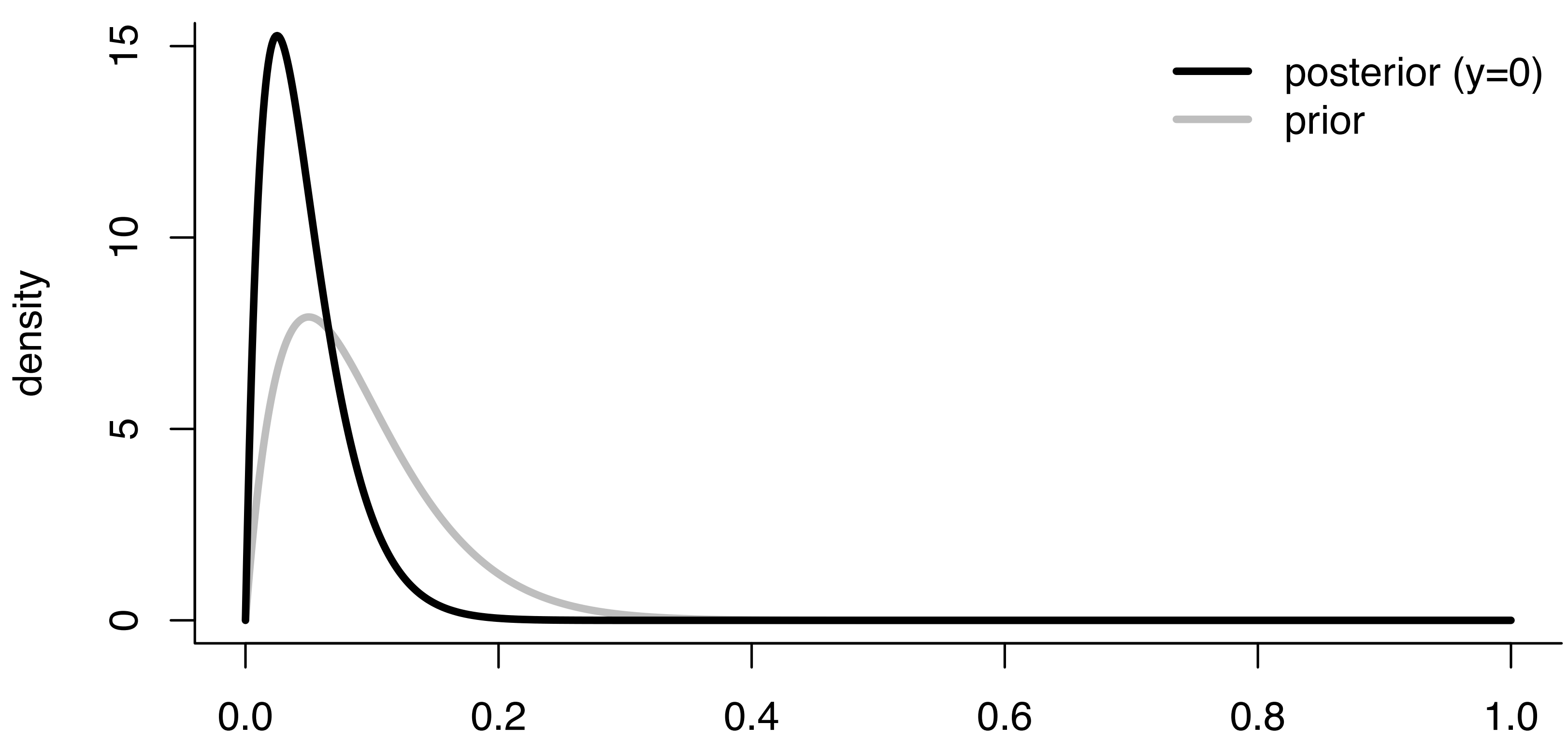
This is an example of a conjugate Bayesian model.

Example: Posterior

For our case, we have $a = 2, \ b = 20, \ n = 20$

If we don't find any infections $(y = 0)$ our posterior is

$$p(\theta|y = 0) = \text{Beta}(2, 40)$$



Example: Prior Sensitivity

How influential is our prior?

The posterior expectation can be written as

$$\mathrm{E}\{\theta|Y = y\} = \frac{n}{w+n}\bar{y} + \frac{w}{w+n}\theta_0$$

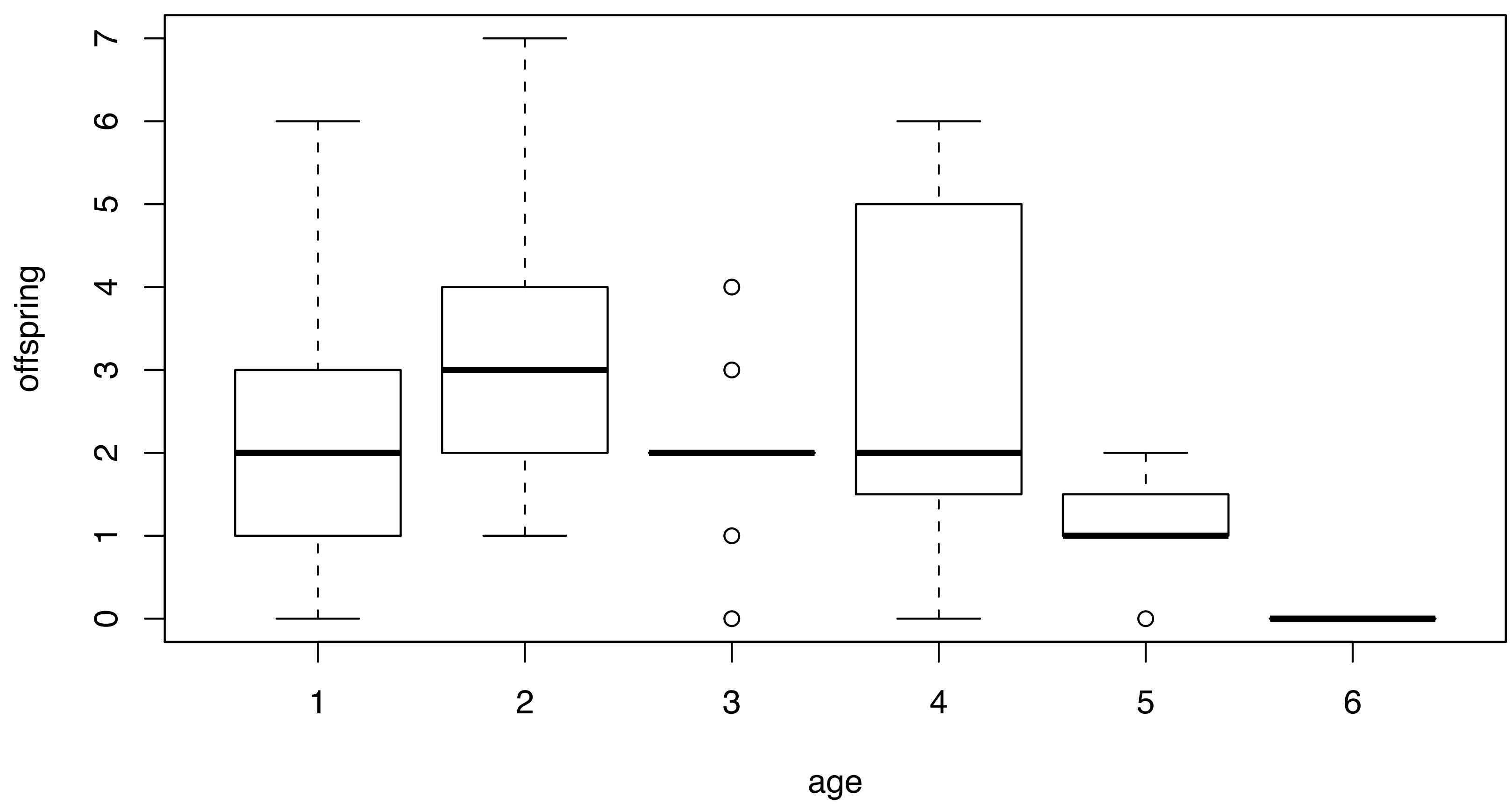
a weighted average of the sample mean and prior expectation:

$$\theta_0 = \frac{a}{a+b}$$ → prior expectation (or guess)

$$w = a + b$$ → prior confidence/ sample size

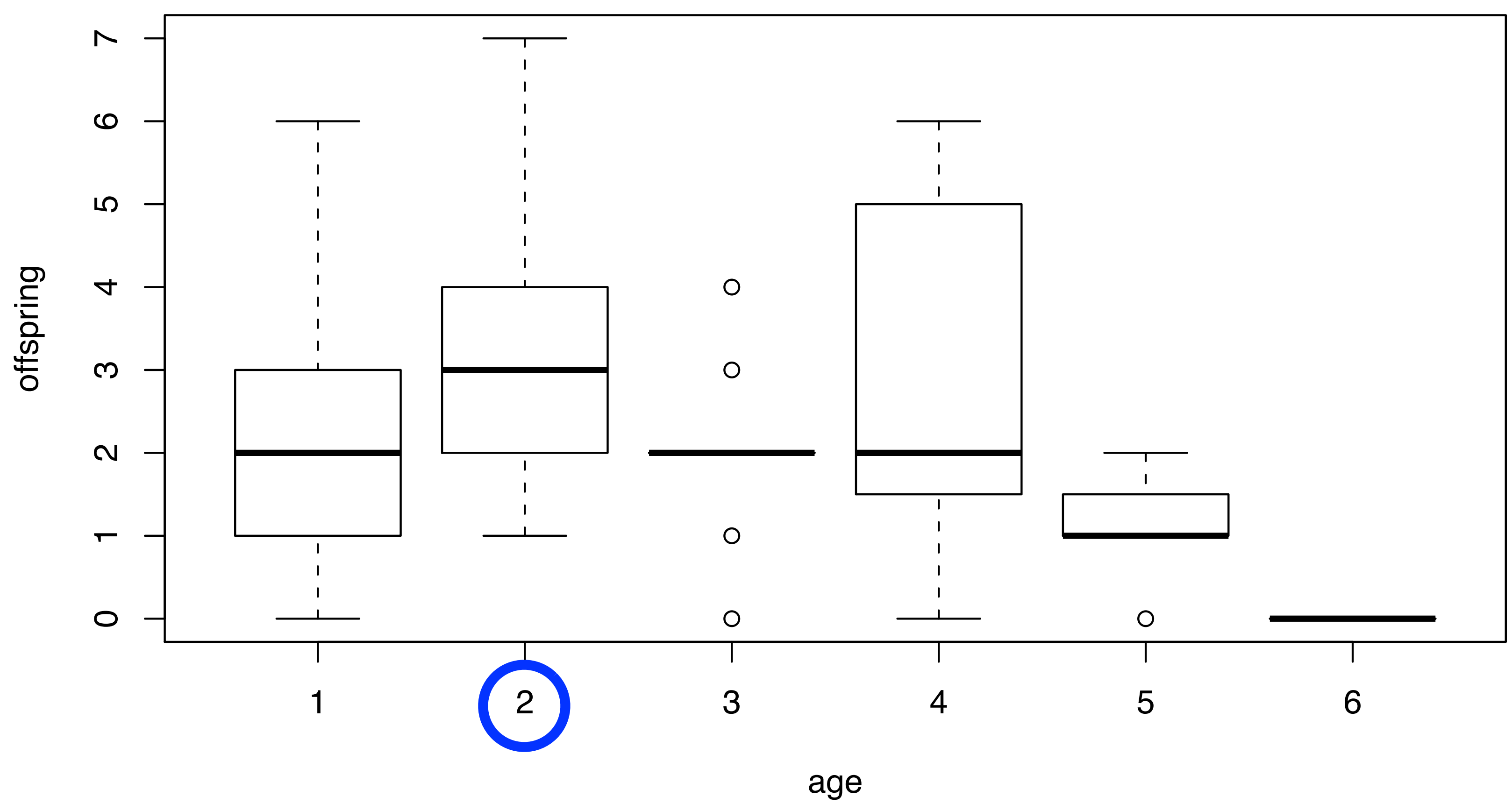
Example: Song sparrow reproductive success

Arcese et al., (1992) provide data on a sample from a population of 52 female song sparrows studied over the course of a summer, during which their reproductive activities were recorded



Example: Song sparrow reproductive success

Arcese et al., (1992) provide data on a sample from a population of 52 female song sparrows studied over the course of a summer, during which their reproductive activities were recorded



2-year-old birds had the highest median reproductive success, declining thereafter

Example: Typical biology

This is not surprising from a biological point of view:

Example: Typical biology

This is not surprising from a biological point of view:

- 1-year-old birds are in their first mating season and are relatively inexperienced compared to two-year-old birds

Example: Typical biology

This is not surprising from a biological point of view:

- 1-year-old birds are in their first mating season and are relatively inexperienced compared to two-year-old birds
- As birds age beyond two years they experience a general decline in health and activity

Example: Typical biology

This is not surprising from a biological point of view:

- 1-year-old birds are in their first mating season and are relatively inexperienced compared to two-year-old birds
- As birds age beyond two years they experience a general decline in health and activity

We wish to fit a probability model to these data

Example: Typical biology

This is not surprising from a biological point of view:

- 1-year-old birds are in their first mating season and are relatively inexperienced compared to two-year-old birds
- As birds age beyond two years they experience a general decline in health and activity

We wish to fit a probability model to these data

- perhaps to understand the relationship between age and reproductive success

Example: Typical biology

This is not surprising from a biological point of view:

- 1-year-old birds are in their first mating season and are relatively inexperienced compared to two-year-old birds
- As birds age beyond two years they experience a general decline in health and activity

We wish to fit a probability model to these data

- perhaps to understand the relationship between age and reproductive success
- or to make population forecasts for this group of birds

Example: Poisson model

Since the number of offspring for each bird is a non-negative integer $\{0, 1, 2, \ldots\}$, a simple probability model for

Example: Poisson model

Since the number of offspring for each bird is a non-negative integer $\{0, 1, 2, \dots\}$, a simple probability model for

$$Y = \text{the number of offspring}$$

conditional on $x = \text{age}$

Example: Poisson model

Since the number of offspring for each bird is a non-negative integer $\{0, 1, 2, \ldots\}$, a simple probability model for

$$Y = \text{the number of offspring}$$

conditional on $x = \text{age}$

would be a Poisson model

$$\{Y|x\} \sim \text{Pois}(\theta_x)$$

Example: Poisson model

Since the number of offspring for each bird is a non-negative integer $\{0, 1, 2, \ldots\}$, a simple probability model for

$$Y = \text{the number of offspring}$$

conditional on $x = \text{age}$

would be a Poisson model

$$\{Y|x\} \sim \text{Pois}(\theta_x)$$

One possibility would be to estimate θ_x separately for each age group

Example: Adding stability

However, the number of birds of each age is small and so the estimates of θ_x would be imprecise

Example: Adding stability

However, the number of birds of each age is small and so the estimates of θ_x would be imprecise

To add stability to the estimation we will assume that the mean number of offspring is a smooth function of age

Example: Adding stability

However, the number of birds of each age is small and so the estimates of θ_x would be imprecise

To add stability to the estimation we will assume that the mean number of offspring is a smooth function of age

We will want to allow this function to be quadratic so that we can represent

Example: Adding stability

However, the number of birds of each age is small and so the estimates of θ_x would be imprecise

To add stability to the estimation we will assume that the mean number of offspring is a smooth function of age

We will want to allow this function to be quadratic so that we can represent

- the increase in mean offspring while birds mature

Example: Adding stability

However, the number of birds of each age is small and so the estimates of θ_x would be imprecise

To add stability to the estimation we will assume that the mean number of offspring is a smooth function of age

We will want to allow this function to be quadratic so that we can represent

- the increase in mean offspring while birds mature
- and the decline they experience thereafter

Example: A linear model?

One possibility would be to express θ_x as

$$\theta_x = \beta_1 + \beta_2 x + \beta_3 x^2$$

Example: A linear model?

One possibility would be to express θ_x as

$$\theta_x = \beta_1 + \beta_2 x + \beta_3 x^2$$

However, this might allow some values of θ_x to be negative, which is not physically possible

Example: A linear model?

One possibility would be to express θ_x as

$$\theta_x = \beta_1 + \beta_2 x + \beta_3 x^2$$

However, this might allow some values of θ_x to be negative, which is not physically possible

As an alternative, we will model the log-mean of Y in terms of this regression so that

$$\log \mathbb{E}\{Y|x\} = \log \theta_x = \beta_1 + \beta_2 x + \beta_3 x^2$$

Example: A linear model?

One possibility would be to express θ_x as

$$\theta_x = \beta_1 + \beta_2 x + \beta_3 x^2$$

However, this might allow some values of θ_x to be negative, which is not physically possible

As an alternative, we will model the log-mean of Y in terms of this regression so that

$$\log \mathbb{E}\{Y|x\} = \log \theta_x = \beta_1 + \beta_2 x + \beta_3 x^2$$

which means that, for all x and β

$$\mathbb{E}\{Y|x\} = \exp\{\beta_1 + \beta_2 x + \beta_3 x^2\} > 0$$

Poisson regression

The resulting model

$$\{Y|x\} \sim \mathrm{Pois}(\exp\{x^\top \beta\})$$

is called a Poisson regression model, or log-linear model

Poisson regression

The resulting model

$$\{Y|x\} \sim \text{Pois}(\exp\{x^\top \beta\})$$

is called a Poisson regression model, or log-linear model

The term $x^\top \beta$ is called the linear predictor

Poisson regression

The resulting model

$$\{Y|x\} \sim \mathrm{Pois}(\exp\{x^\top \beta\})$$

is called a Poisson regression model, or log-linear model

The term $x^\top \beta$ is called the linear predictor

In the regression model the linear predictor is linked to $\mathbb{E}\{Y|x\}$ via the $\log$ function, and so we say that this model has a $\log$ link

Generalized linear model

The Poisson regression/log-linear model is a type of generalized linear model (GLM), a model which

Generalized linear model

The Poisson regression/log-linear model is a type of generalized linear model (GLM), a model which

- allows more general response distributions for Y than the normal distribution
- relates a function of the expectation $\mu = \mathbb{E}\{Y\}$ to a linear predictor $\eta = x^\top \beta$ through the link $g(\mu) = \eta$

Generalized linear model

The Poisson regression/log-linear model is a type of generalized linear model (GLM), a model which

- allows more general response distributions for Y than the normal distribution

- relates a function of the expectation $\mu = \mathbb{E}\{Y\}$ to a linear predictor $\eta = x^\top \beta$ through the link $g(\mu) = \eta$

These two choices define the GLM

Example: Prior specification

Unlike the normal or Beta-Binomial models, there is not a conjugate (or semi-conjugate) prior for the Poisson regression model.

$$\beta_0, \beta_1, \beta_2 \sim \mathcal{N}(0, 100^2)$$

Example: Prior specification

Unlike the normal or Beta-Binomial models, there is not a conjugate (or semi-conjugate) prior for the Poisson regression model.

So we can choose what we'd like! Since the coefficients of the linear predictors can be anywhere along the real line, but we don't want to force them to be any particular value, I choose:

$$\beta_0, \beta_1, \beta_2 \sim \mathcal{N}(0, 100^2)$$

Example: Prior specification

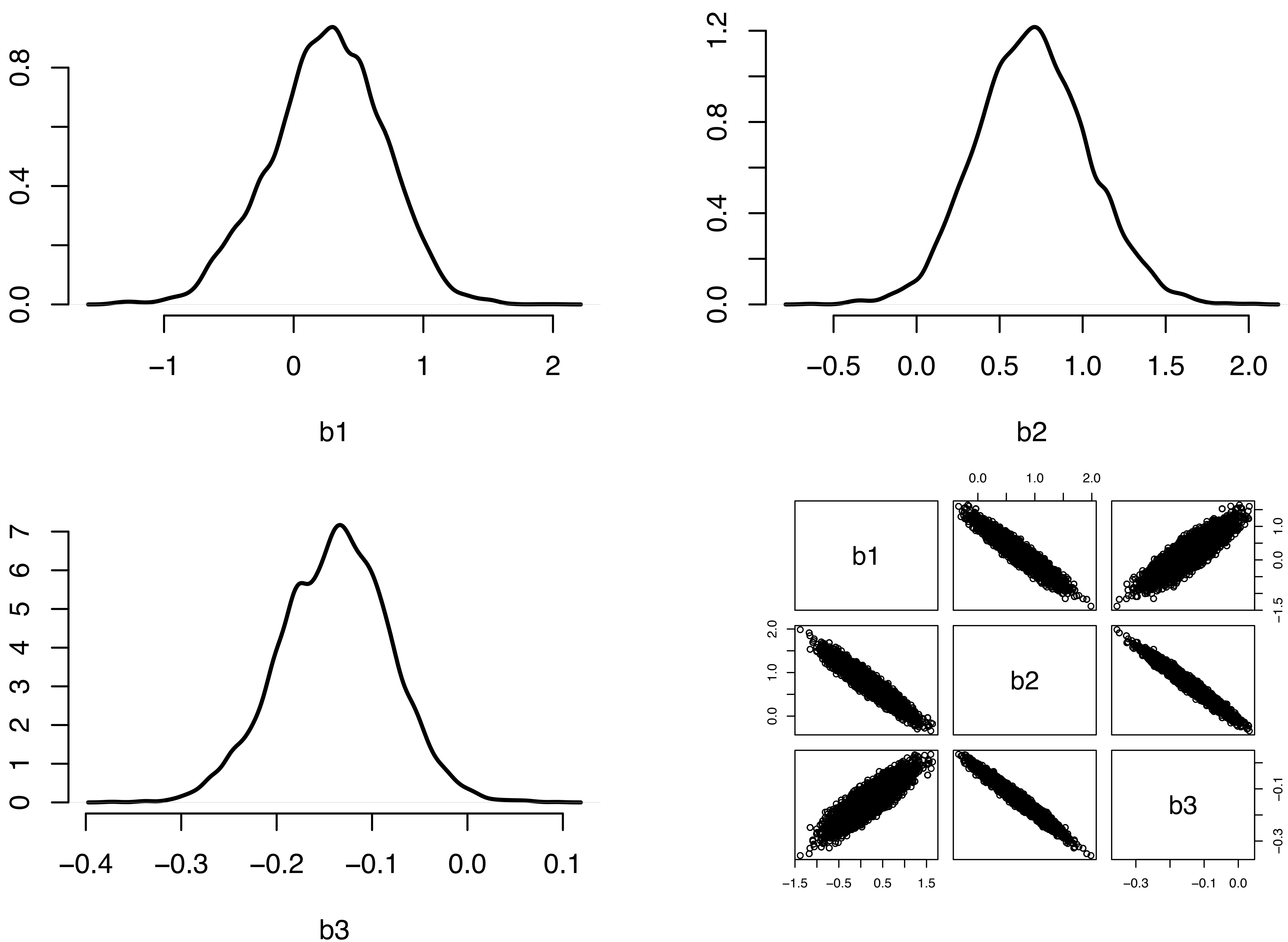
Unlike the normal or Beta-Binomial models, there is not a conjugate (or semi-conjugate) prior for the Poisson regression model.

So we can choose what we'd like! Since the coefficients of the linear predictors can be anywhere along the real line, but we don't want to force them to be any particular value, I choose:

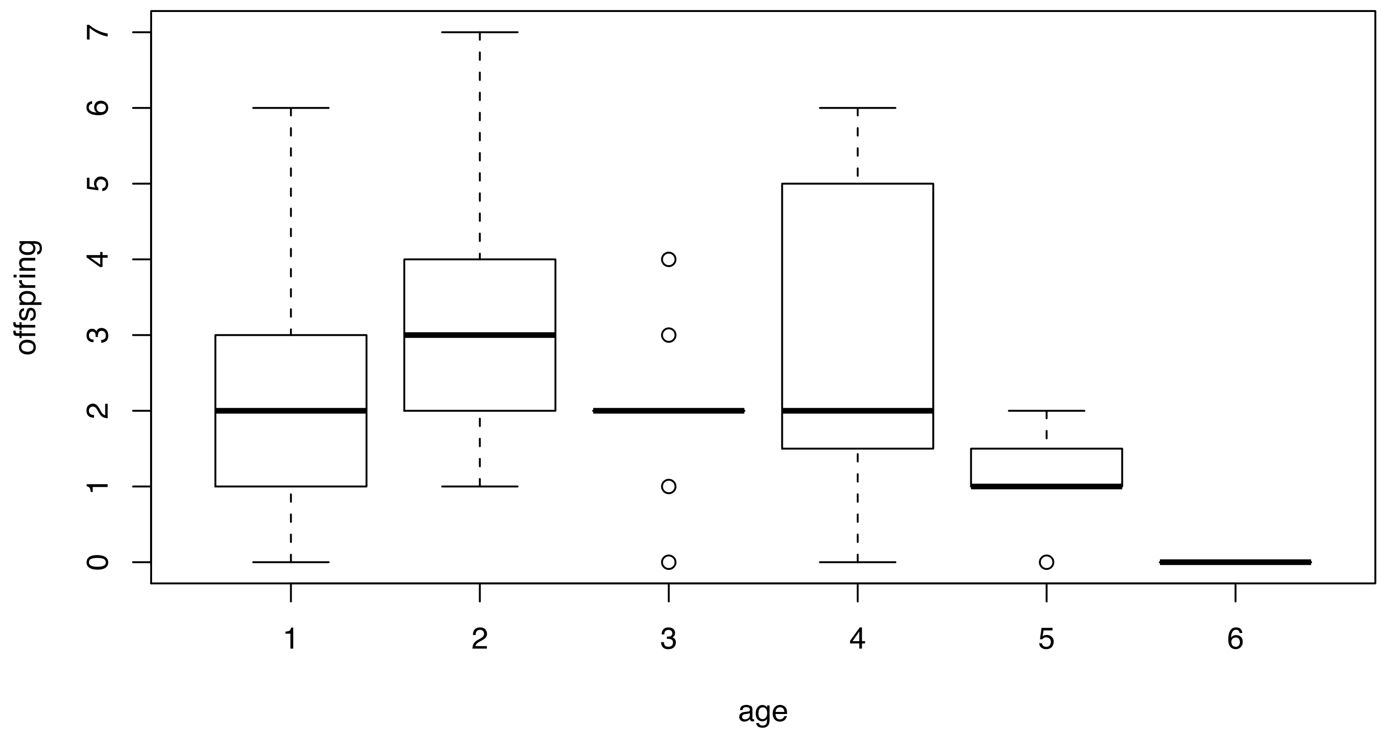
$$\beta_0, \beta_1, \beta_2 \sim \mathcal{N}(0, 100^2)$$

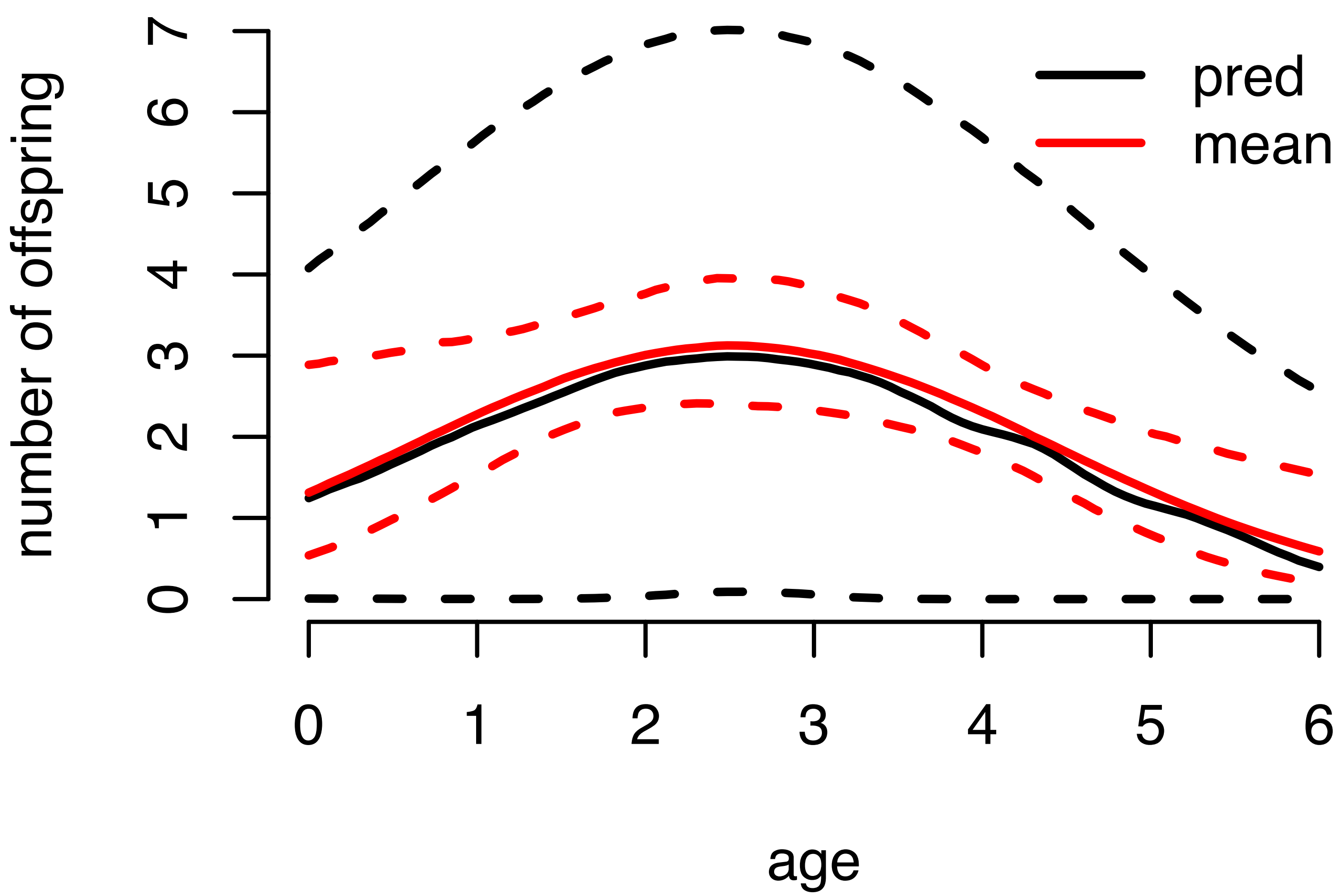
Then turn the Bayesian crank...

Example: Posterior marginals/joint



Example: Posterior predictive





Linear model assumptions

The linear model rests on some important assumptions:

- Errors are additive and ~~normally distributed~~
- Errors are homoskdastic (don't vary across Xs)
- Observations are independent (conditional on the linear predictor)
- Linear (in covariates) mean function
- All error/randomness is in the value of the response (i.e., the X values are precisely known)
- There is no (systematic) missing data

Linear model assumptions

The linear model rests on some important assumptions:

- Errors are additive and ~~normally distributed~~
- Errors are homoskdastic (don't vary across Xs)
- Observations are independent (conditional on the linear predictor)
- Linear (in covariates) mean function
- All error/randomness is in the value of the response (i.e., the X values are precisely known)
- There is no (systematic) missing data

How can relax some of these other assumptions?