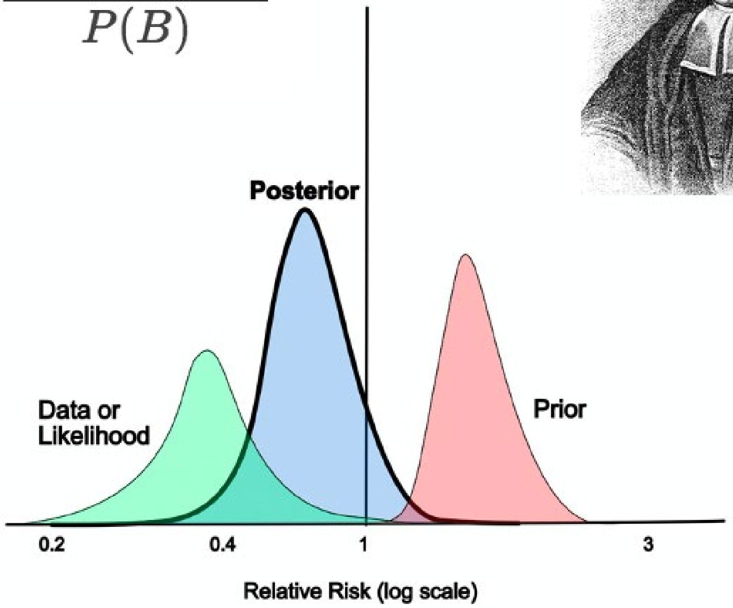
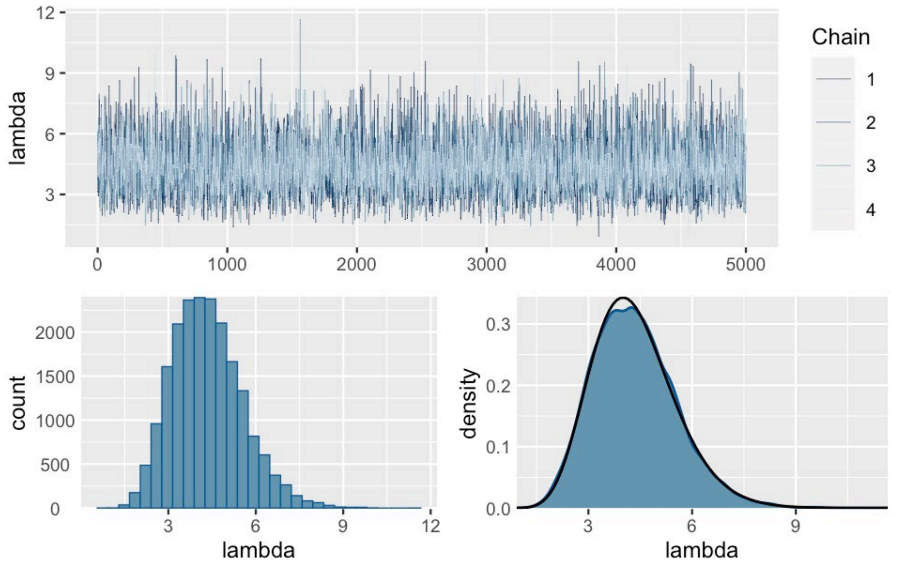
Bayesian Analysis Workshop:
Introduction to Bayesian Framework

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$





Overview

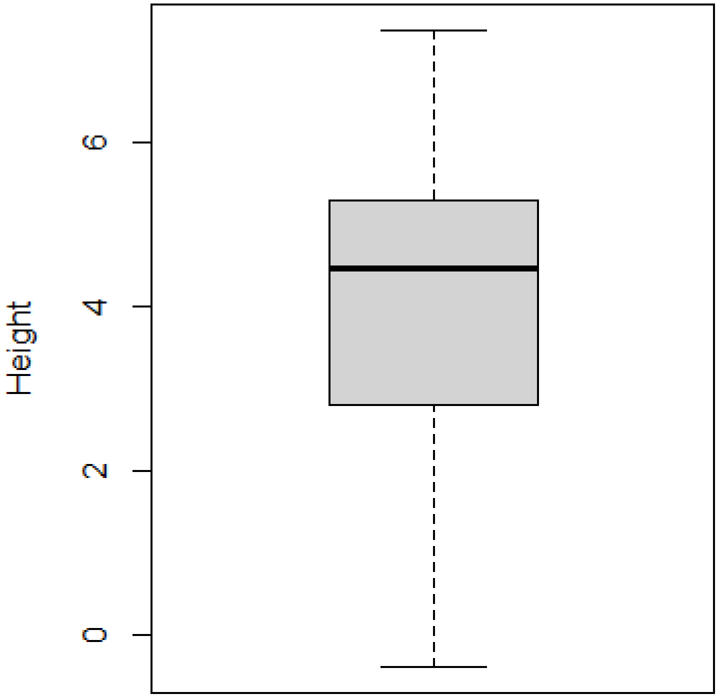
- What’s a statistical model?
- Frequentist vs. Bayesian
- Bayes Theorem
- Review of (generalized) linear models
- Side-by-side comparison of frequentist and Bayesian analysis
- Steps of Bayesian data analysis

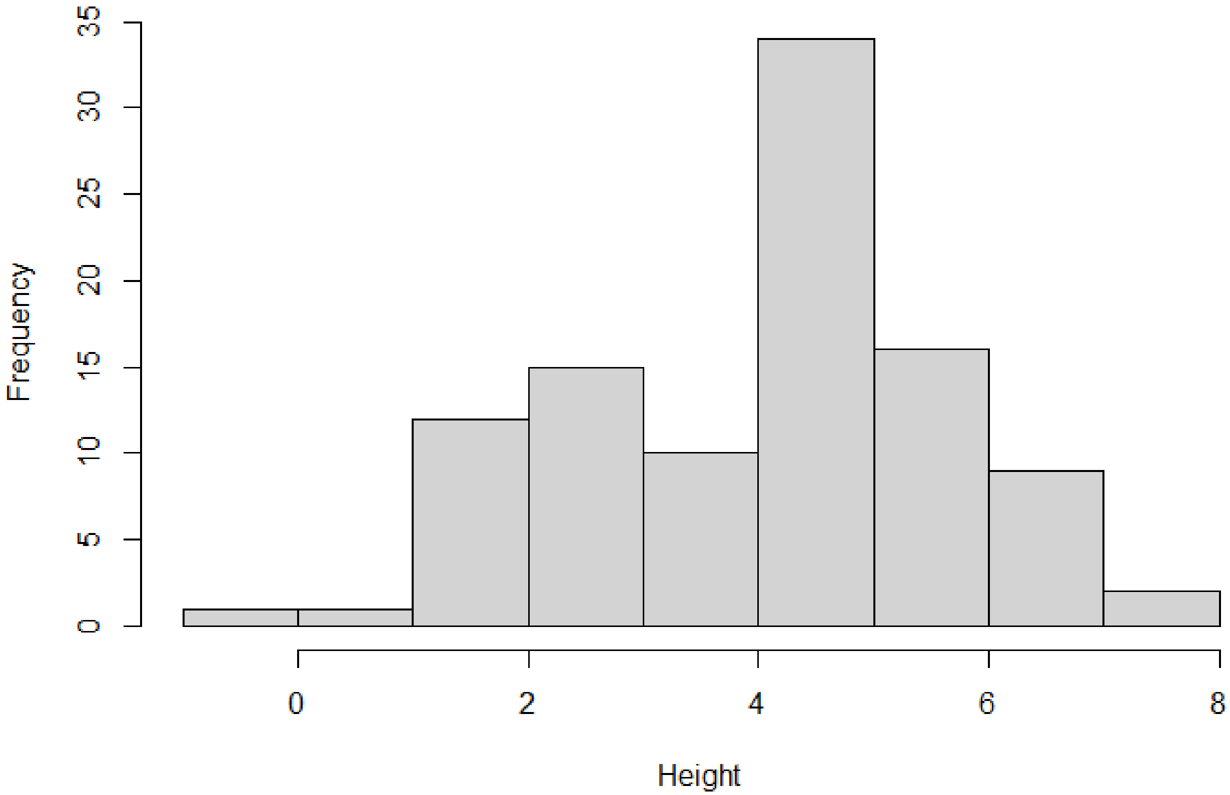
Acknowledgements

- Ohio Division of Wildlife – Sponsor
- Federal Aid in Wildlife Restoration (PR)
- Dennis Hull – Support, coordination, logistics
- Trisha Taylor – Catering support

Purpose of statistical models

- Help describe how we *think* a system works
- Can summarize data (e.g., mean, SD)





$Normal(\mu, \sigma)$

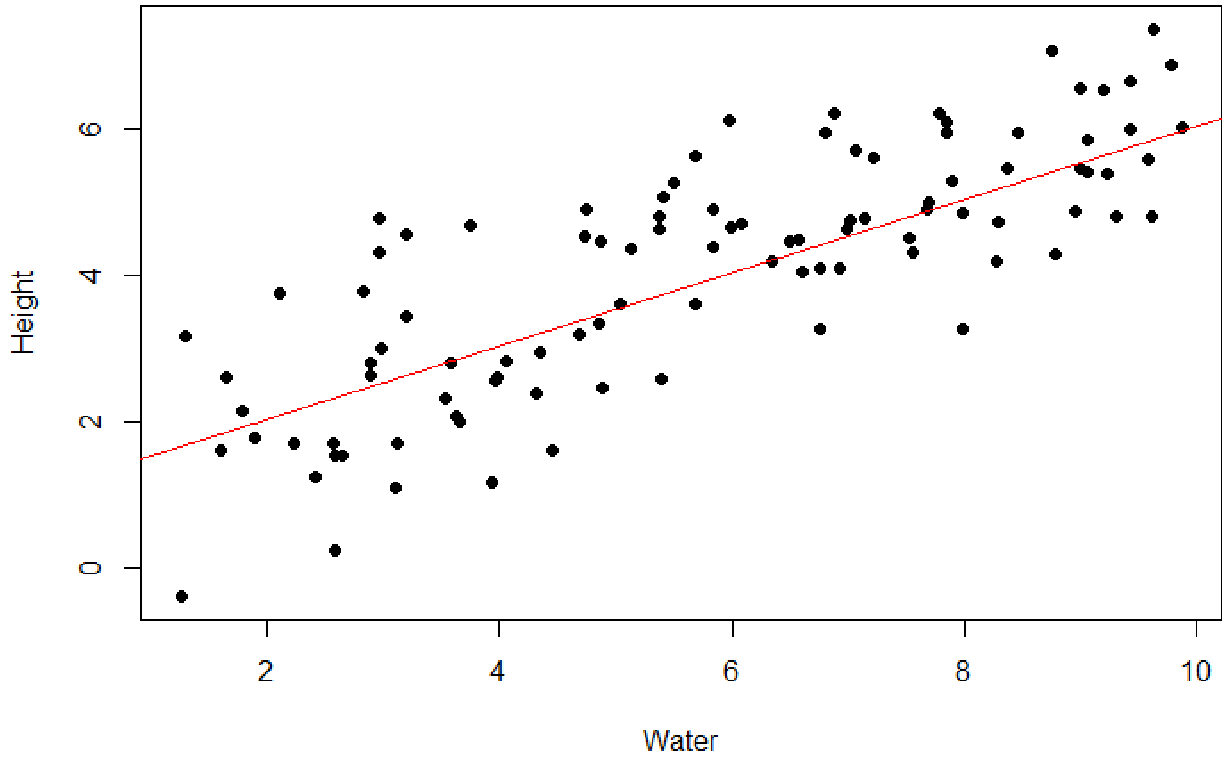
$Normal(4.07, 1.64)$

Purpose of statistical models

- Help describe how we *think* a system works
- Can summarize data (e.g., mean, SD)
- Comparing model predictions with data helps our understanding of the system
- Models allow for predictions, including estimates of uncertainty
  - Help with decision making
- Statistics help us make probabilistic conclusions about parameters, based on a model and observations

What’s a statistical model?

- (simplified) mathematical expression of reality
- Data are observations of a system
- Express what we know about a system
    - No model is perfect, but a good model is useful!



Statistical Approaches

* Data = observed realization of stochastic systems containing random processes

* Differ in their definition of probability

Classical (Frequentist)

- Parameters (random processes) are fixed and unknown constants
- Uncertainty evaluated in terms of *frequency* of hypothetical data sets
- Relative frequency of a feature of observed data

Bayesian

- Parameters are viewed as unobserved realizations of random processes
- Uncertainty evaluated using **posterior distribution** of parameter
- Probability used to express uncertainty of estimated parameters

Bayesian vs. Frequentist

- Both rely on defining a likelihood function
  - $p(x|\theta)$
  - Reads: "Probability of data, given theta"
- Bayesian distinguishes between observable (x) and unobservable (θ)
  - $\theta \rightarrow$ Random quantities that can only be determine probabilistically
  - Statistical parameters, missing data, mismeasured data, future outcomes (predictions)
  - Makes inference about entire distribution of possibilities
- Frequentist approach seeks point estimate (maximum likelihood estimate) of θ, which is an unknown constant

Posterior Distribution

- The conditional probability distribution of all unknown quantities (parameters), given the **data**, the **model** , and what we know about these quantities (**priors**) before conducting the analysis

- Bayes' Theorem

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

LIKELIHOOD
The probability of the data given the parameters

PRIOR
Probability of theta (defined as a distribution)

$$\boxed{p(\theta|x)} = \frac{\boxed{p(x|\theta)}\,\boxed{p(\theta)}}{\boxed{p(x)}}$$

POSTERIOR
Probability of given the data are true

CONSTANT
Probability of observing the data

Posterior distribution $\propto$ Likelihood $\times$ Prior distribution

Struggles with Priors

- Perhaps the most ‘contentious’ aspect of Bayesian modeling
- Prior represents our understanding and/or assumptions about a parameter before modeling

$$\text{Posterior distribution} \propto \text{Likelihood} \times \text{Prior distribution}$$

Benefits of Bayes

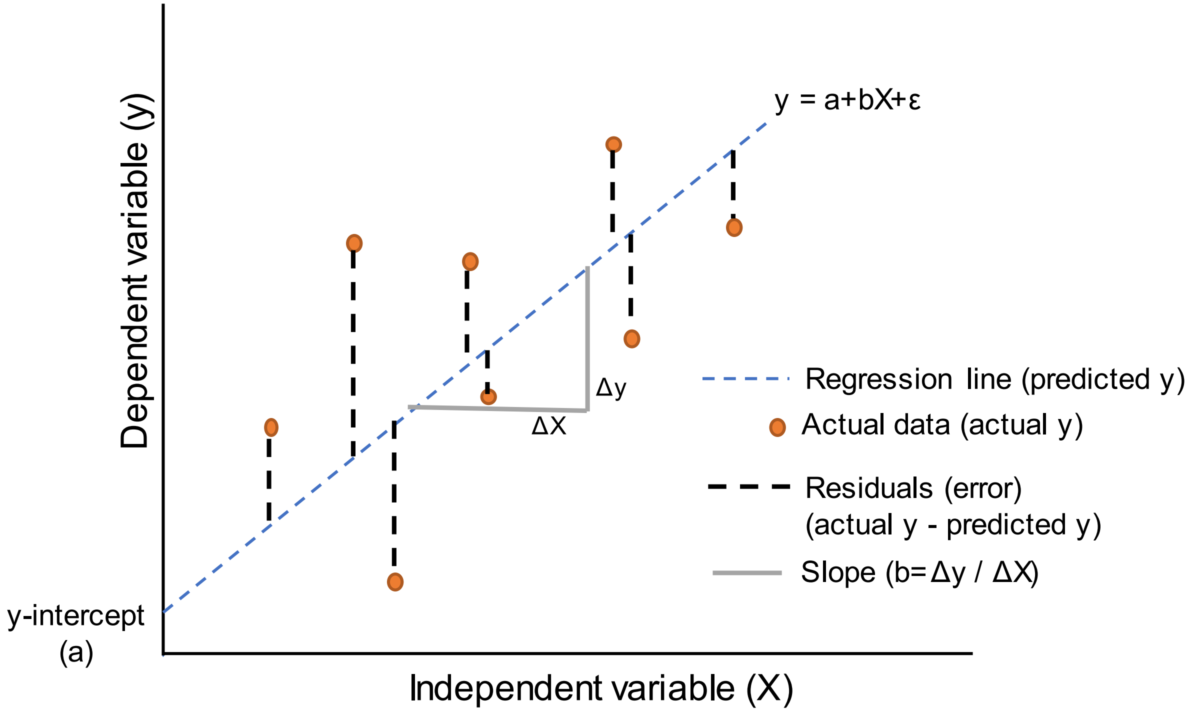
- Valid with small sample size
  - More reliance placed on prior information
- Principled way of combining prior information with data
- Flexible and customizable…virtually anything is possible if you can code it!
- With large sample sizes and/or uninformed priors, Bayesian models give similar point estimates to frequentist MLE models
- Provides interpretable answers
  - Bayes: “The true parameter θ has a 0.95 probability of falling within the estimated 95% credible interval”
  - Frequentist: “Given an infinite number of calculated intervals, the true parameter θ will be within the confidence interval 95% of the time

Worked example

- Overview / Description of Data
  - Data summary, descriptive stats, scatterplot
- Frequentist
  - Linear model estimates
- Bayesian (brms)
  - uninformed priors
  - informed priors

(Generalized) Linear Models – Refresher

- Broadly and generally used
- Two parts
  - Systematic
    - Describes how predictors (*x*) relate to observations (*y*)
    - Produces fitted value (i.e. **slope** and **intercept**)
  - Stochastic
    - Describes the uncertainty/scatter around our estimates
    - **Residuals**



Linear Model Assumptions

- The observed y-values are *independent*, conditional on x.
- The y-values are *normally distributed* with *constant variance*
  - $y \sim N(\mu y, \sigma^2)$
- There is a *straight-line relationship* between the mean of y and each x:
  - $\mu y = \beta 0 + \mathbf{x}T\ \beta$
- There is a *straight-line relationship* between *some known function of the mean* of y and each x
  - $g(\mu y) = \beta 0 + \mathbf{x}T\ \beta$

<u>Generalized</u> Linear Model Assumptions

- The observed y-values are *independent*, conditional on x.
- ~~The y-values are *normally distributed* with *constant variance*~~
  - ~~$y \sim N(\mu_y, \sigma^2)$~~
- ~~There is a *straight-line relationship* between the mean of y and each x:~~
  - ~~$\mu_y = \beta_0 + \mathbf{x}^T \beta$~~
- There is a *straight-line relationship* between *some known function of the mean* of y and each x
  - $g(\mu_y) = \beta_0 + \mathbf{x}^T \beta$

$$g(\mu_y) = \beta_0 + X^T\beta$$

- Link function
  - Transforms predicted values into the range of the *linear* predictor ($-\infty$ to $+\infty$)

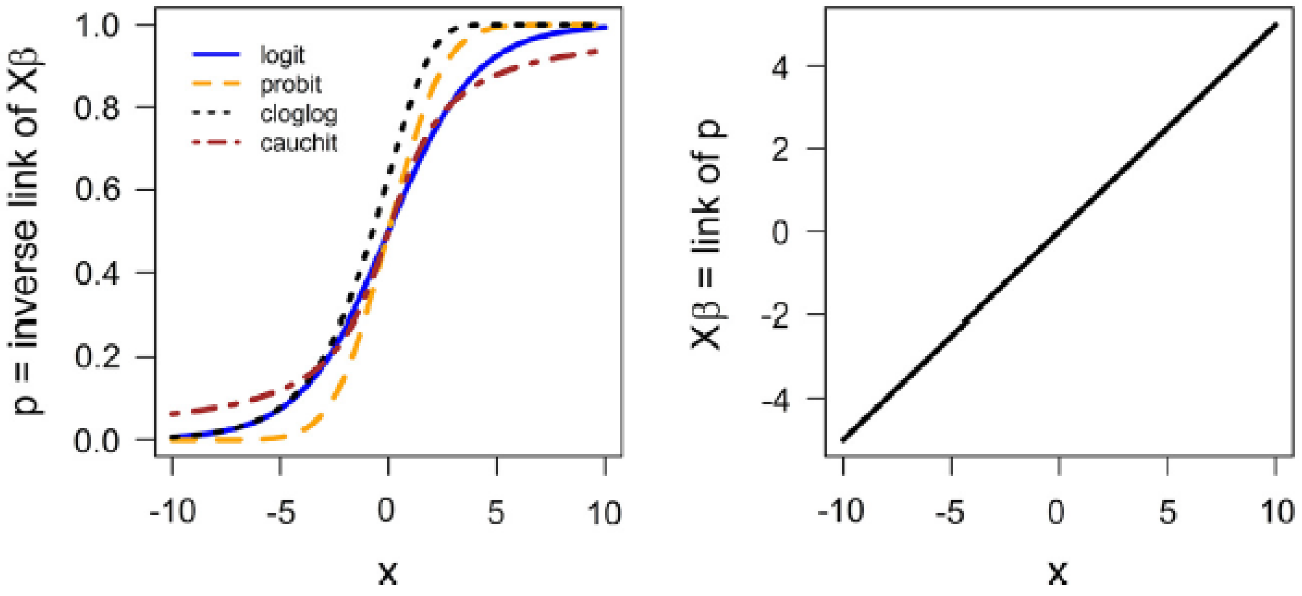


FIGURE 8-3 Left panel: Shape of different link functions commonly used in binomial models. Right panel: The relationship between the predictor X (x-axis) and p on the scale of the link function (y-axis) is assumed to be linear.

Table 10.1: Common choices of distribution and suggested link functions $g(\mu)$ in generalised linear models. Each distribution implies a particular mean–variance assumption $V(\mu)$. The required `family` argument to use each of these in `R` is also given

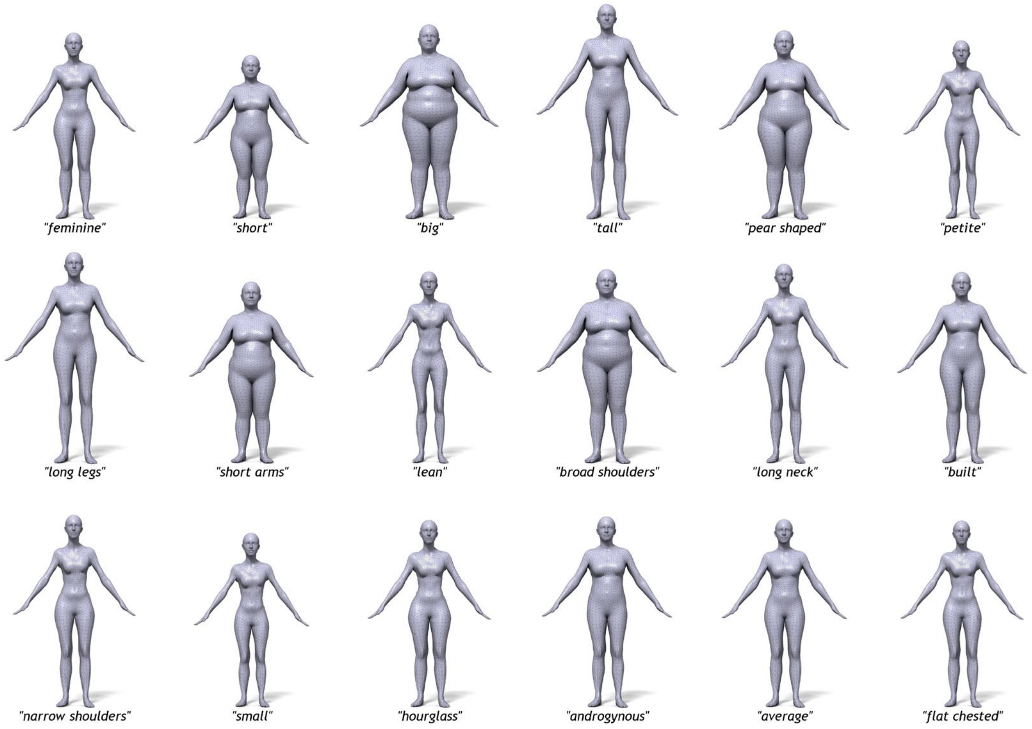
Distribution	$V(\mu)$	Good for...	Link, $g(\mu)$	`family=...`
Binomial	$n\mu(1-\mu)$	Binary responses (e.g. presence–absence)	$\log\left(\frac{\mu}{1-\mu}\right)$	`binomial`
			Probit	`binomial("probit")`
			$\log(-\log(1-\mu))$	`binomial("cloglog")`
Poisson	μ	Counts[a]	$\log(\mu)$	`poisson`
Negative binomial	$\mu+\phi\mu^2$	Counts	$\log(\mu)$	`"negative.binomial"` (in `mvabund`)
Tweedie	$a\mu^p$	Biomass	$\log(\mu)$	`tweedie(p,link=0)` (in `statmod`)
Normal	σ^2	Continuous responses	μ	`gaussian`

[a] But does not account for overdispersion (i.e. all individuals being counted need to be independent, no missing predictors in the model.)

Height Weight example

Goal: estimate the relationship between height and weight

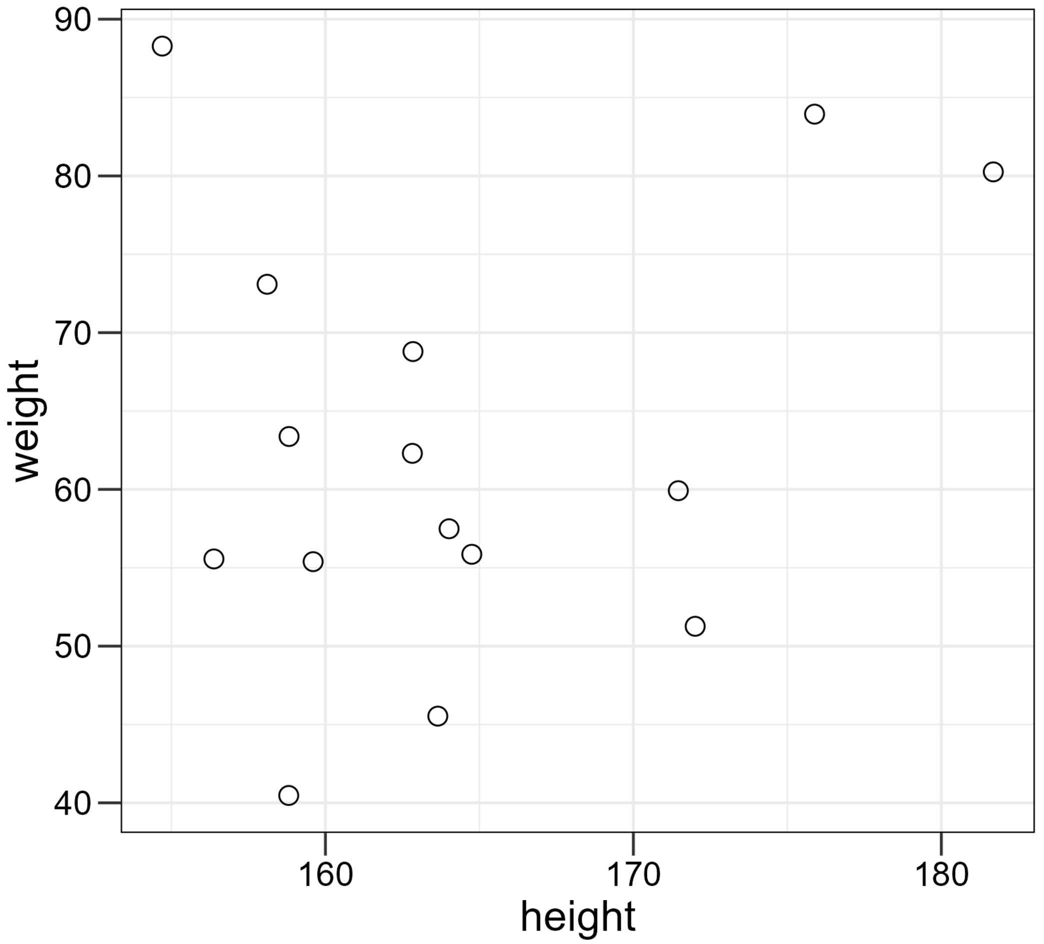
1. Simulate heights and weights
2. Fit three models:
    1. Maximum likelihood estimation
    2. Bayesian estimation (flat priors)
    3. Bayesian estimation (informed priors)
3. Compare estimates



Simulated dataset

```
# Simulation parameters
N = 15
bH = 0.5
b_intercept = 66
error = 10
```

$$Y_i = 66 + 0.5*X_i + error$$



Three models

1. Maximum likelihood estimator

   Parameters based on theoretical t-distribution

2. Bayesian estimator

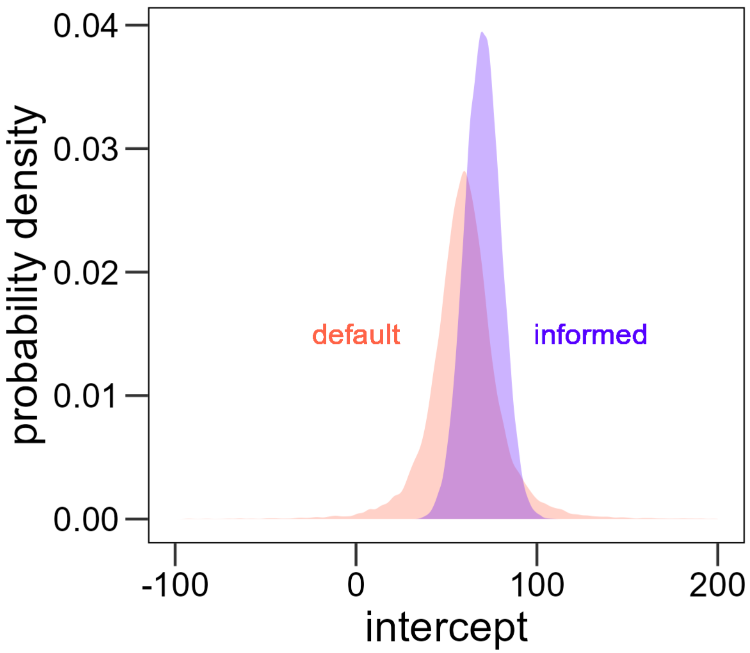
   brms default: Flat priors for slopes (uniform), t-distribution for intercept

3. Bayesian estimator

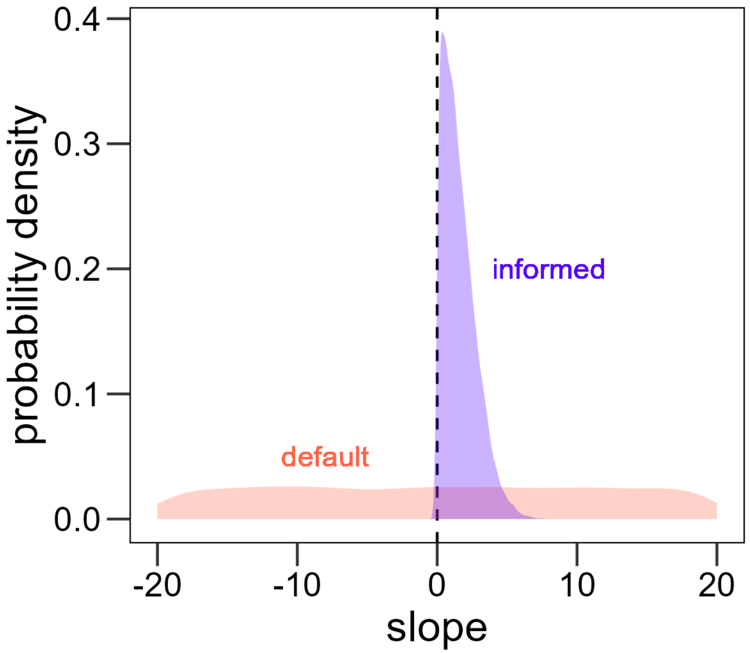
   brms: informed priors (next slide)

Priors for models 2 & 3

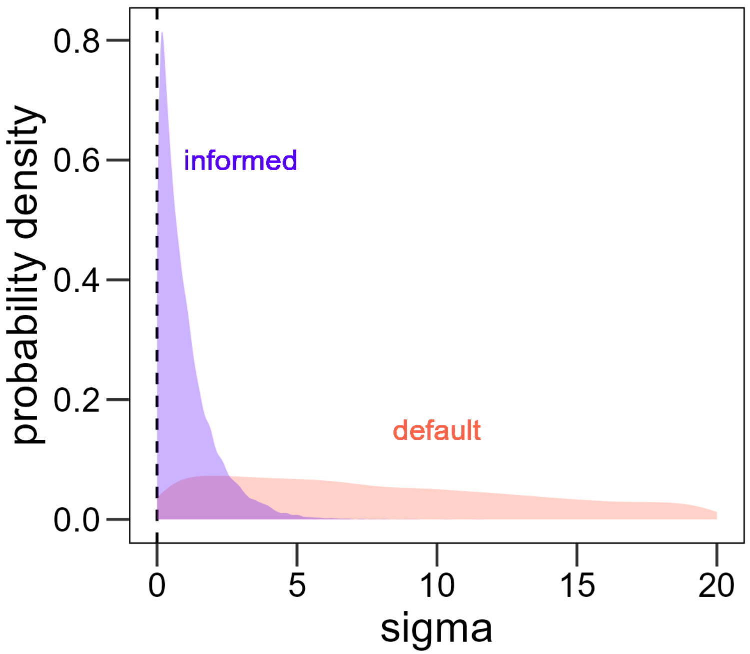
student-t(3, 0, 12.8)
normal(70,10)



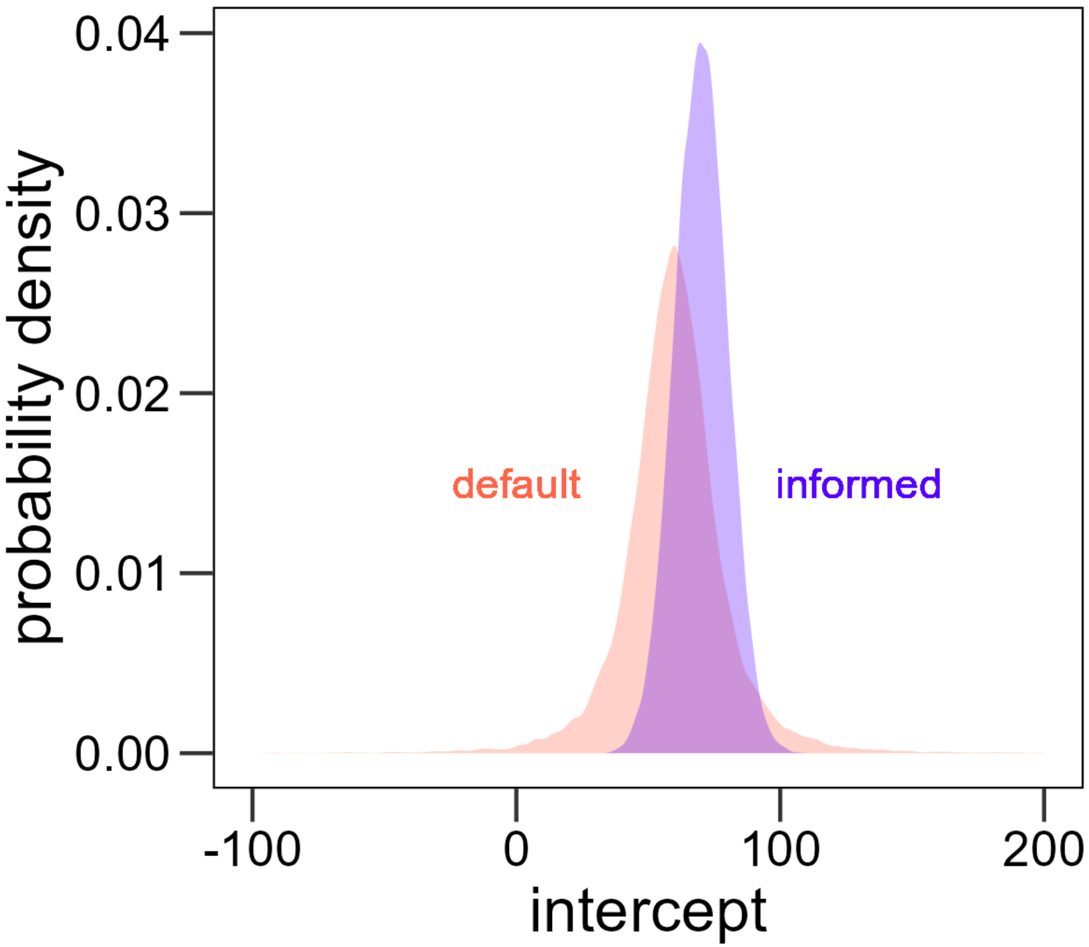
uniform(-Inf, Inf)
half-normal(70,10)



half-student-t(3, 0, 12.8)
exponential(1)

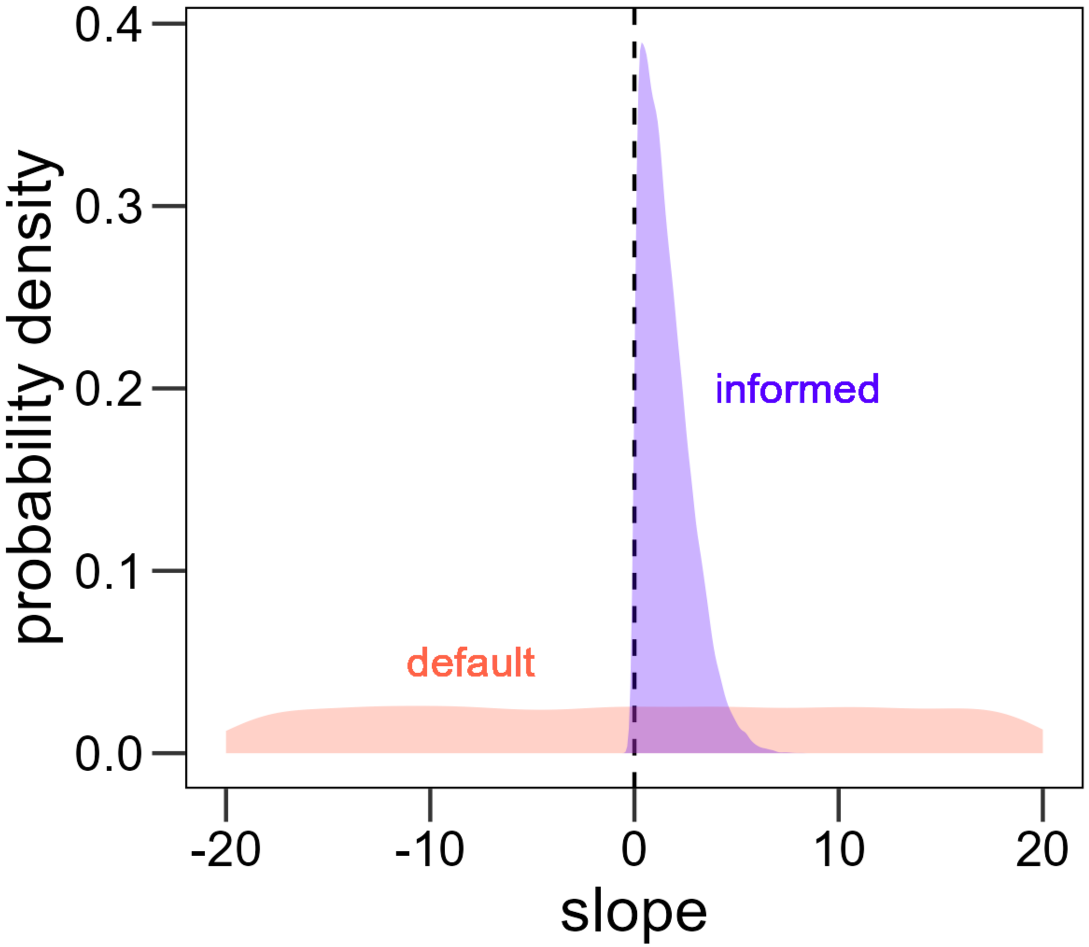


Intercept priors



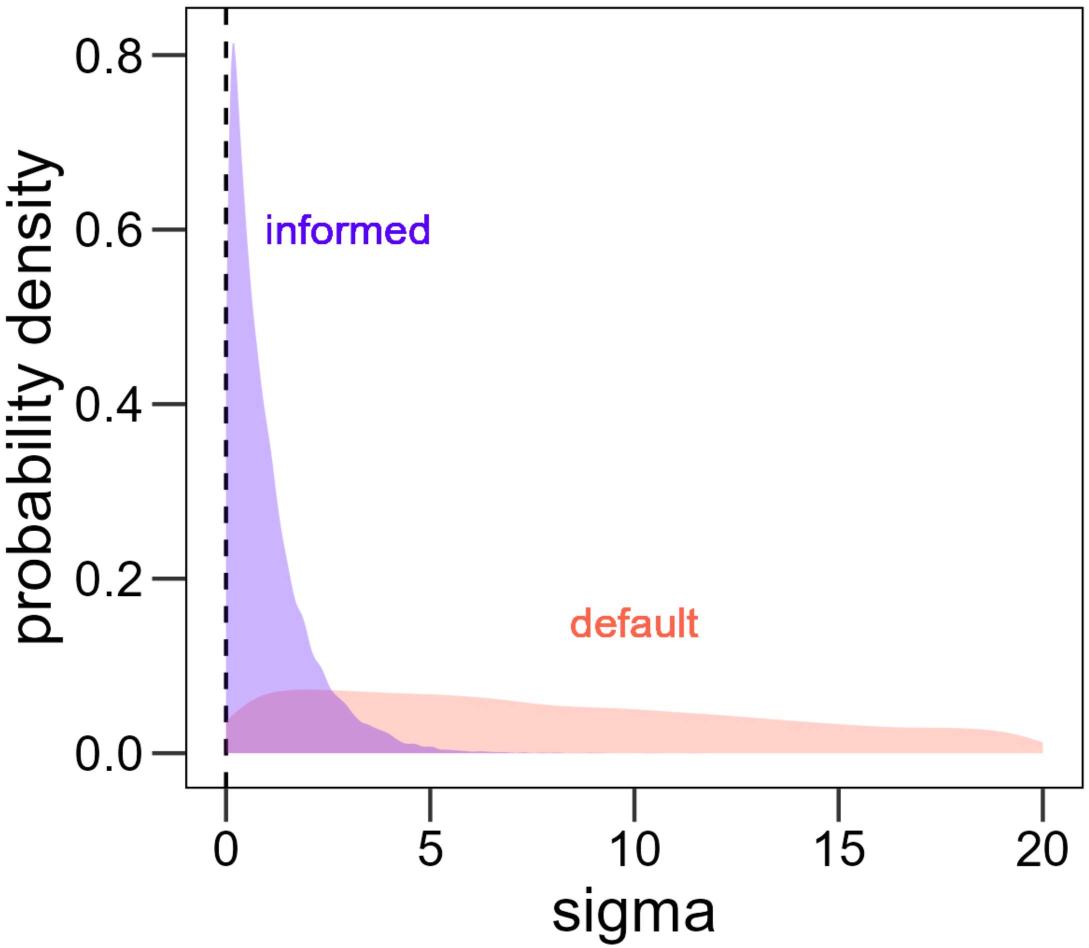
- Default is a student-t with long tails that allows *negative weight*
- Informed uses a normal with a mean of 70 and SD of 10
- Informed prior encodes our expectation that average weight will be *normally distributed and greater than 0.*

Slope priors



- Default is a flat uniform; any slope value equally likely
- Informed uses a half-normal with a mean of 0 and SD of 2
- Informed prior encodes our expectation that *the relationship between height and weight must be positive*

Sigma priors



- Default is a half-student-t, again with a very long tail
- Informed uses an exponential with rate of 1.
- Informed prior is skeptical of extreme levels of error

Results

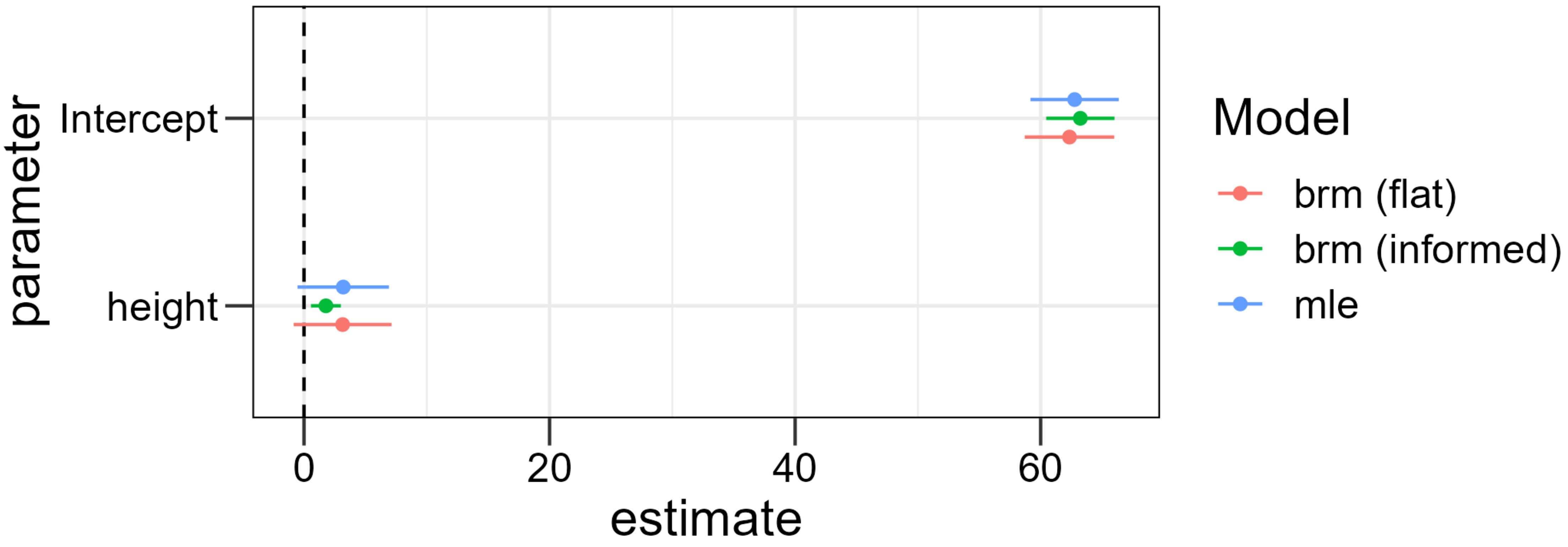
True values

bH = 0.5

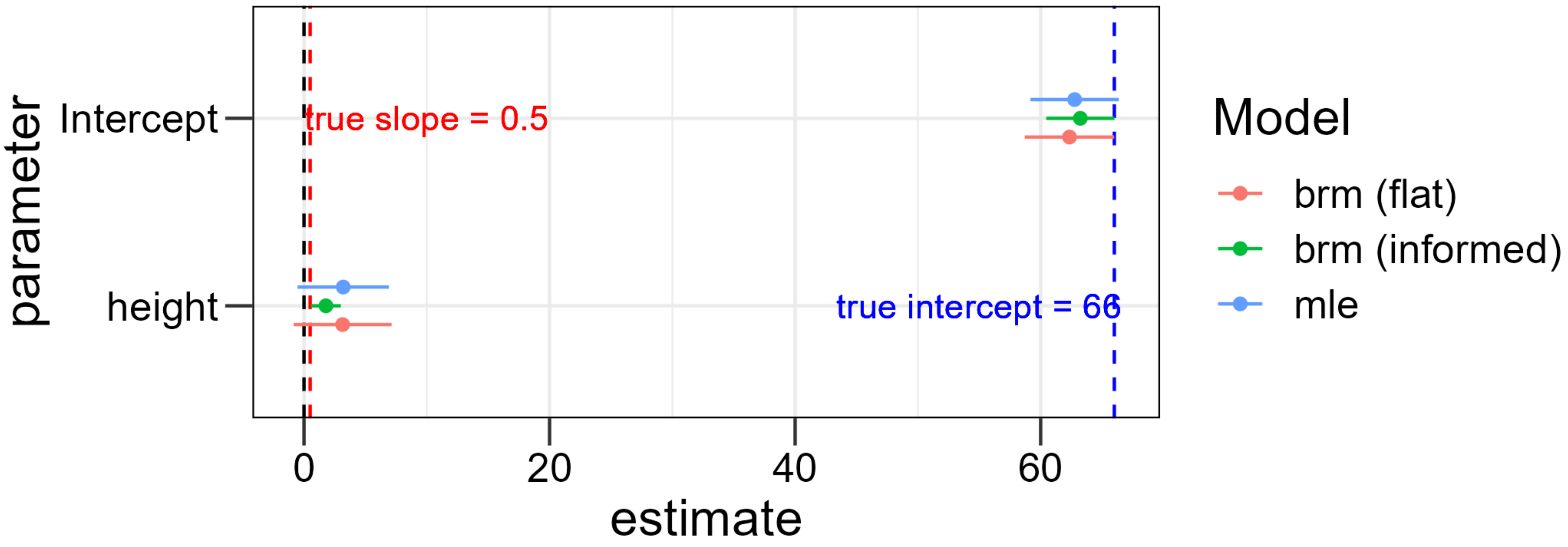
b_intercept = 66

Model	**Intercept (SE)**	**Height (SE)**
MLE	62.77 (3.60)	3.19 (3.72)
Bayesian (flat)	62.35 (3.65)	3.14 (3.99)
Bayesian (informed)	63.23 (2.78)	**1.78 (1.22)**

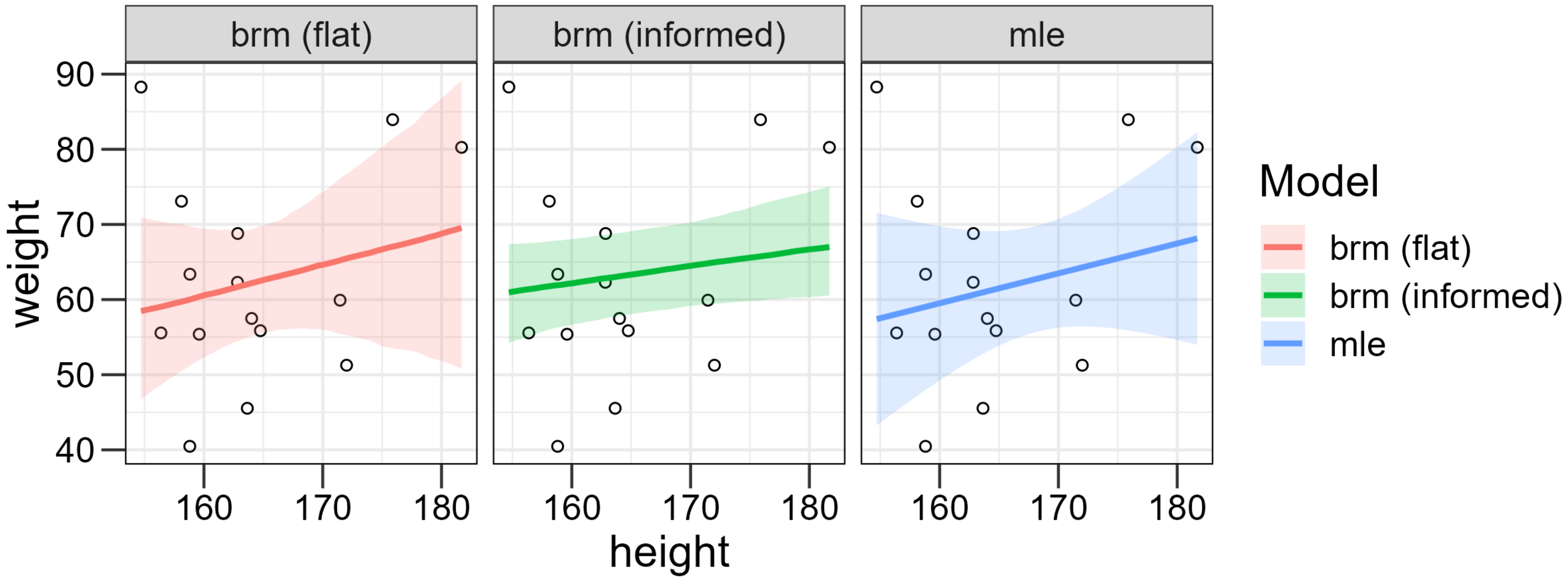
Informed models performs best



Informed models performs best



Predictions



Steps to Bayesian Data Analysis

- Data: Identify variables to be predicted and variables that act as predictors
- Define a descriptive model
- Specify a prior distribution on the parameters
- Use Bayesian inference to re-allocate credibility across parameter values.
- Check that posterior predictions mimic the data with reasonable accuracy (i.e., posterior predictive check)