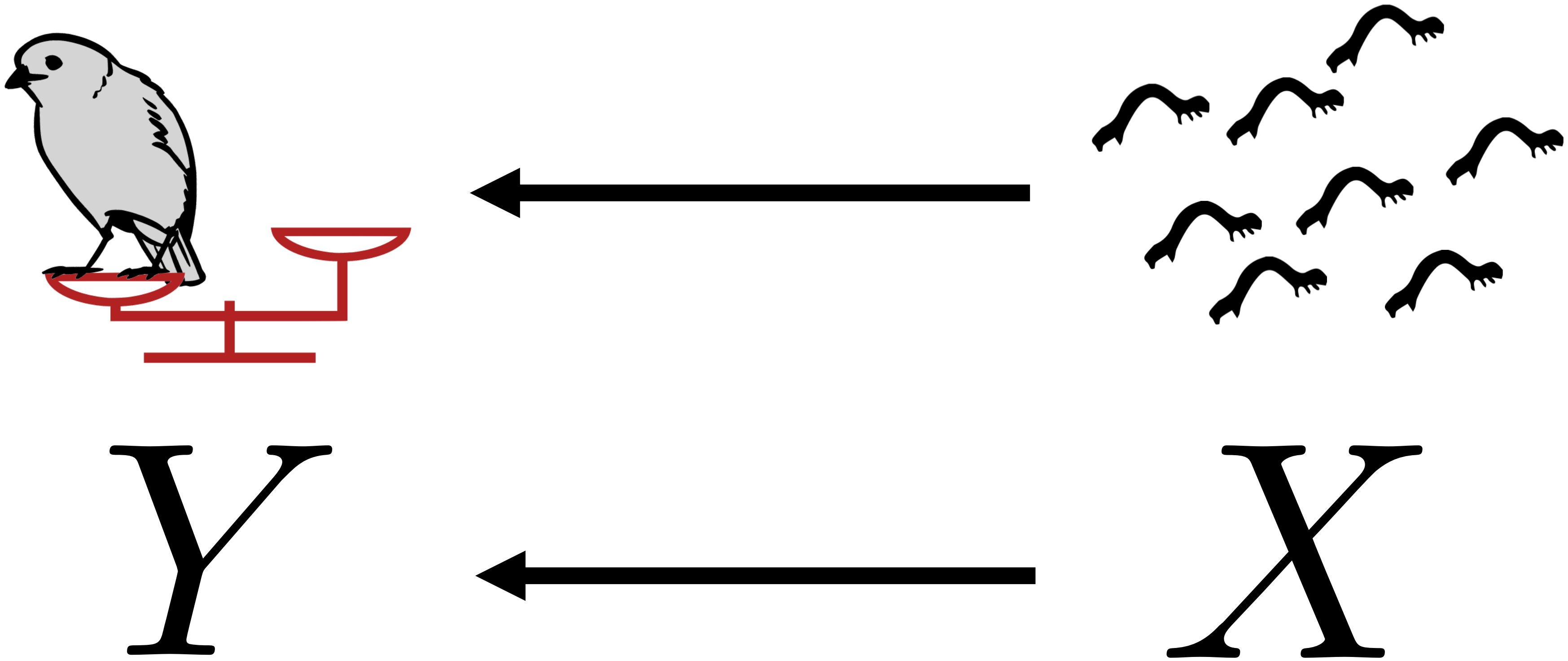
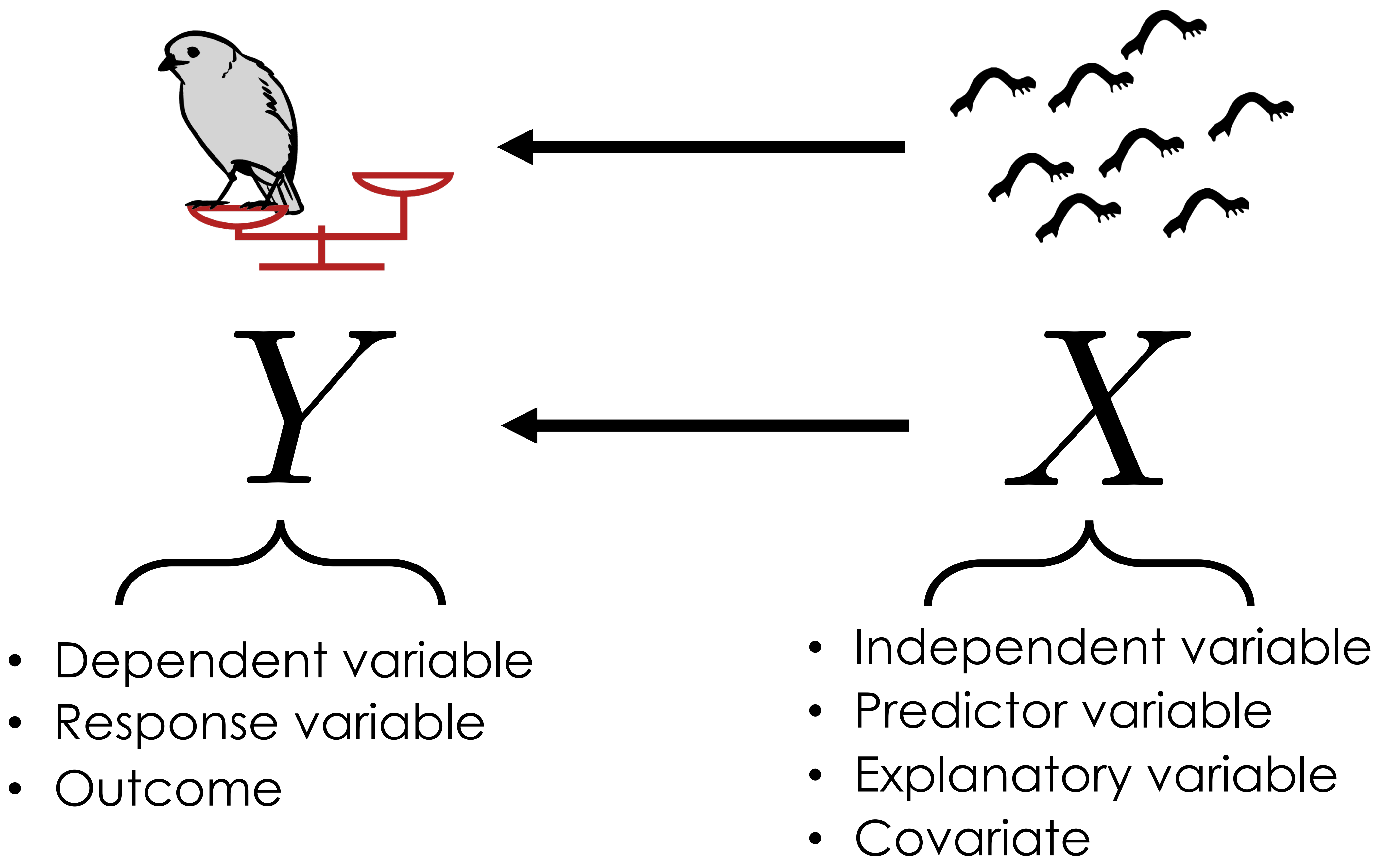

Y
X


Y
X
• Dependent variable
• Response variable
• Outcome
• Independent variable
• Predictor variable
• Explanatory variable
• Covariate


BIRDS
AREN'T
REAL
WAKE UP
LIES
birds charge on power lines
PIGEONS ARE LIARS

```r
#set seed for reproducing results
set.seed(1)
#number of data points
N <- 30
```

3-bird-simulation.R

```r
#set seed for reproducing results
set.seed(1)
#number of data points
N <- 30
#simulate predictors (standardized)
food_std <- rnorm(N, 0, 1)
```

3-bird-simulation.R

```r
#set seed for reproducing results
set.seed(1)
#number of data points
N <- 30
#simulate predictors (standardized)
food_std <- rnorm(N, 0, 1)
#generating intercept and slope values
alpha <- 40
beta <- 3
#simulate linear predictor
mu <- alpha + beta * food_std
```

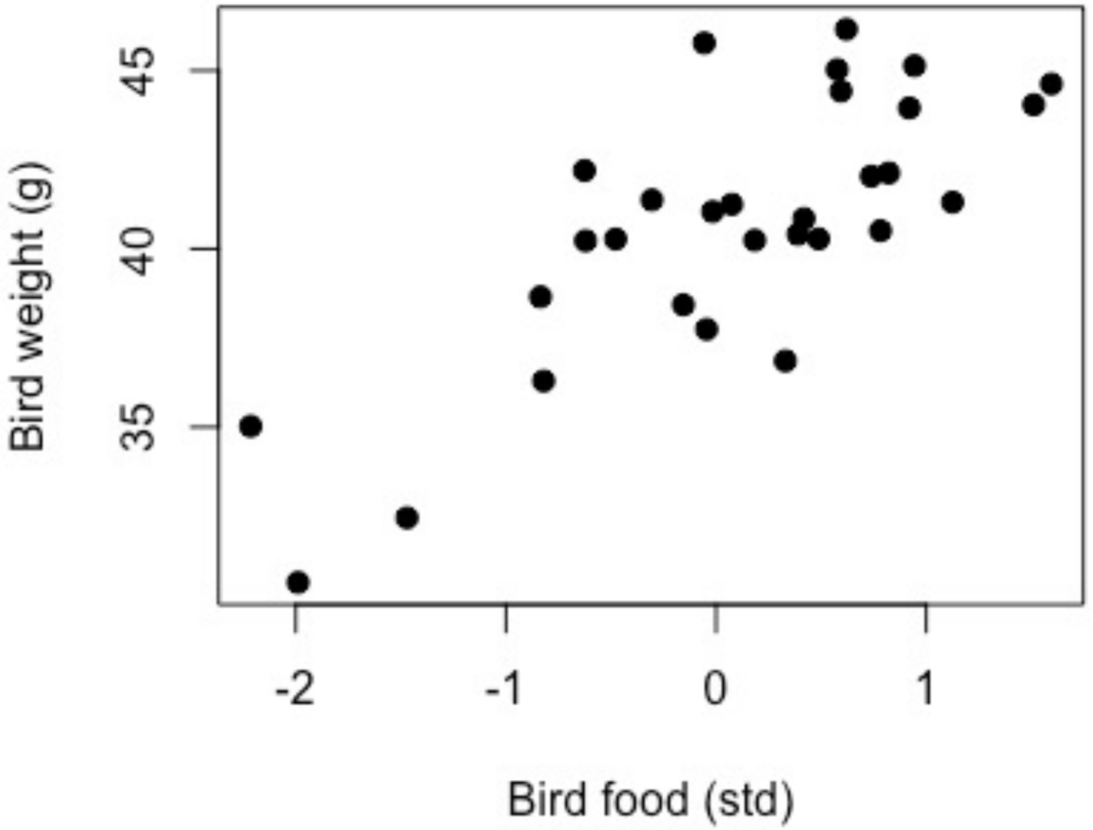
3-bird-simulation.R

```r
#set seed for reproducing results
set.seed(1)
#number of data points
N <- 30
#simulate predictors (standardized)
food_std <- rnorm(N, 0, 1)
#generating intercept and slope values
alpha <- 40
beta <- 3
#simulate linear predictor
mu <- alpha + beta * food_std
#process error (wrt relationship
between food and bird weight)
sigma <- 3
#simulate y values
bird_weight <- rnorm(N, mu, sigma)
```

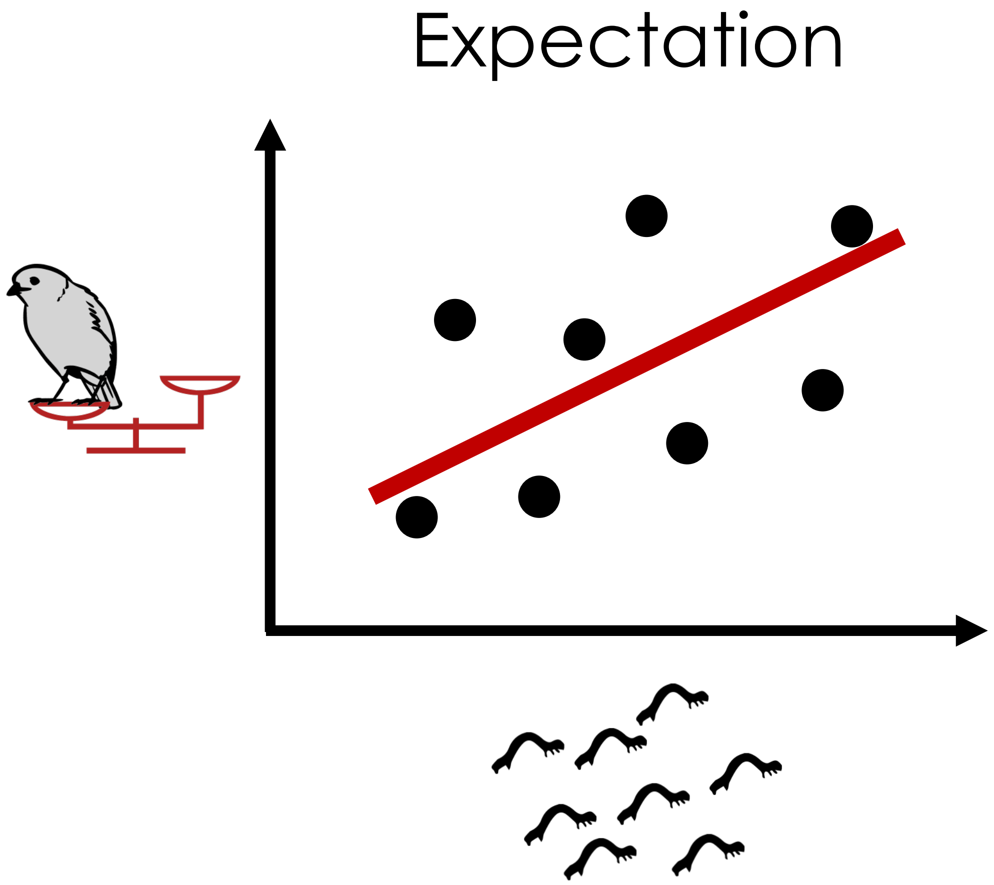
3-bird-simulation.R

```r
#set seed for reproducing results
set.seed(1)
#number of data points
N <- 30
#simulate predictors (standardized)
food_std <- rnorm(N, 0, 1)
#generating intercept and slope values
alpha <- 40
beta <- 3
#simulate linear predictor
mu <- alpha + beta * food_std
#process error (wrt relationship
between food and bird weight)
sigma <- 3
#simulate y values
bird_weight <- rnorm(N, mu, sigma)

#plot
plot(food_std, bird_weight,
     pch = 19,
     xlab = 'Bird food (std)',
     ylab = 'Bird weight (g)')
```



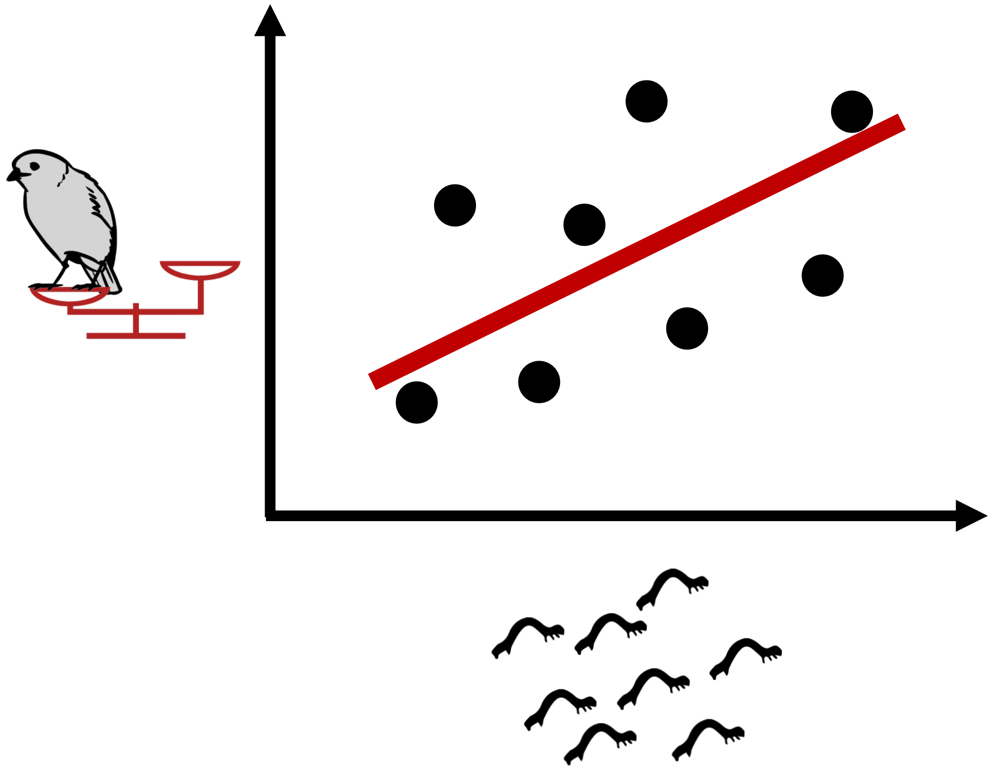
3-bird-simulation.R


Expectation

Response

$$y_i \sim Normal(\mu_i, \sigma)$$

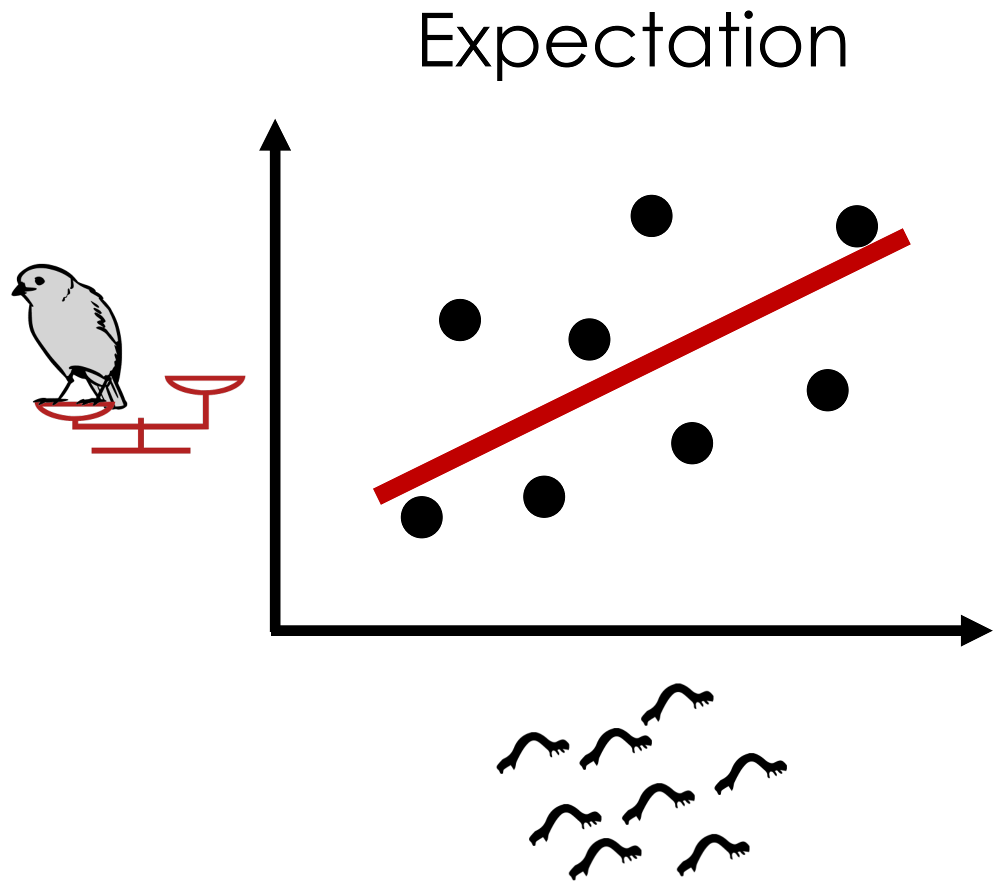
Expectation



Response

Expected value

$$y_i \sim Normal(\mu_i, \sigma)$$



Response

Expected value

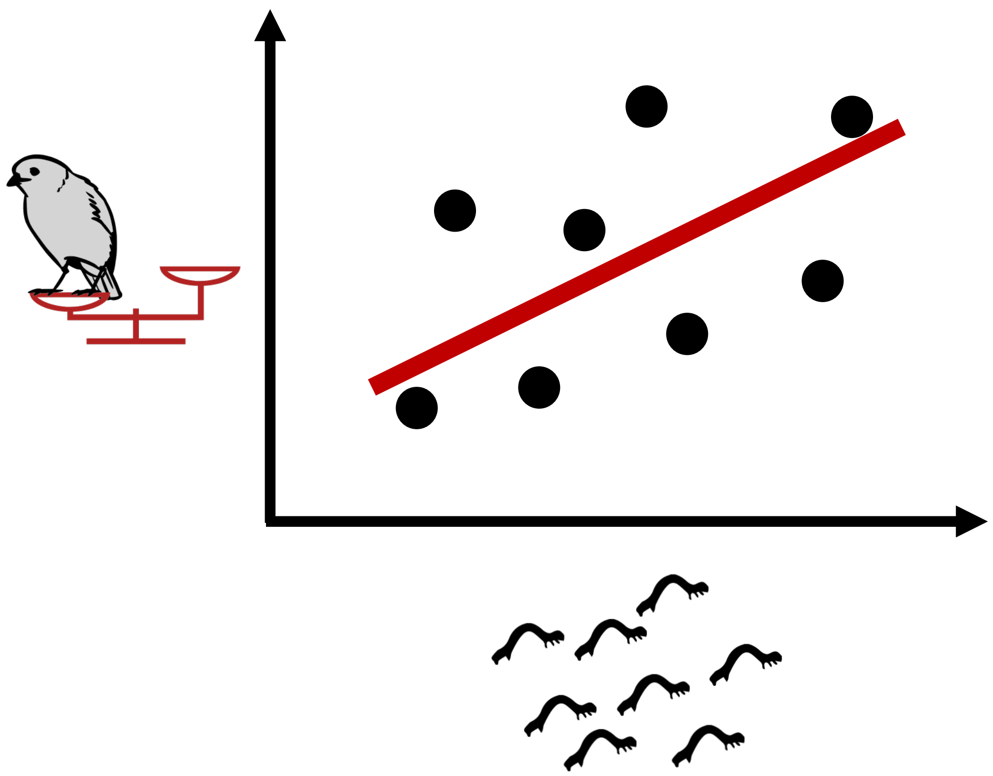
$$y_i \sim Normal(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \times x_i$$

Intercept

Slope

Predictor

Expectation



Response

Expected value

Residual Error

$$y_i \sim Normal(\mu_i, \sigma)$$

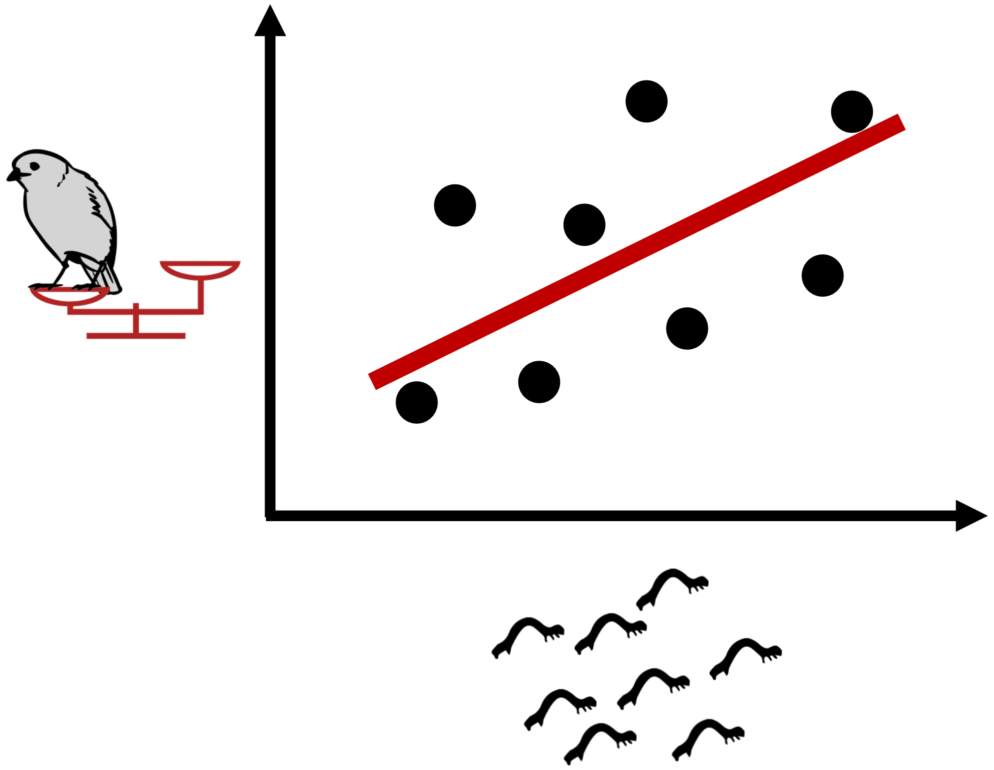
$$\mu_i = \alpha + \beta \times x_i$$

Intercept

Slope

Predictor

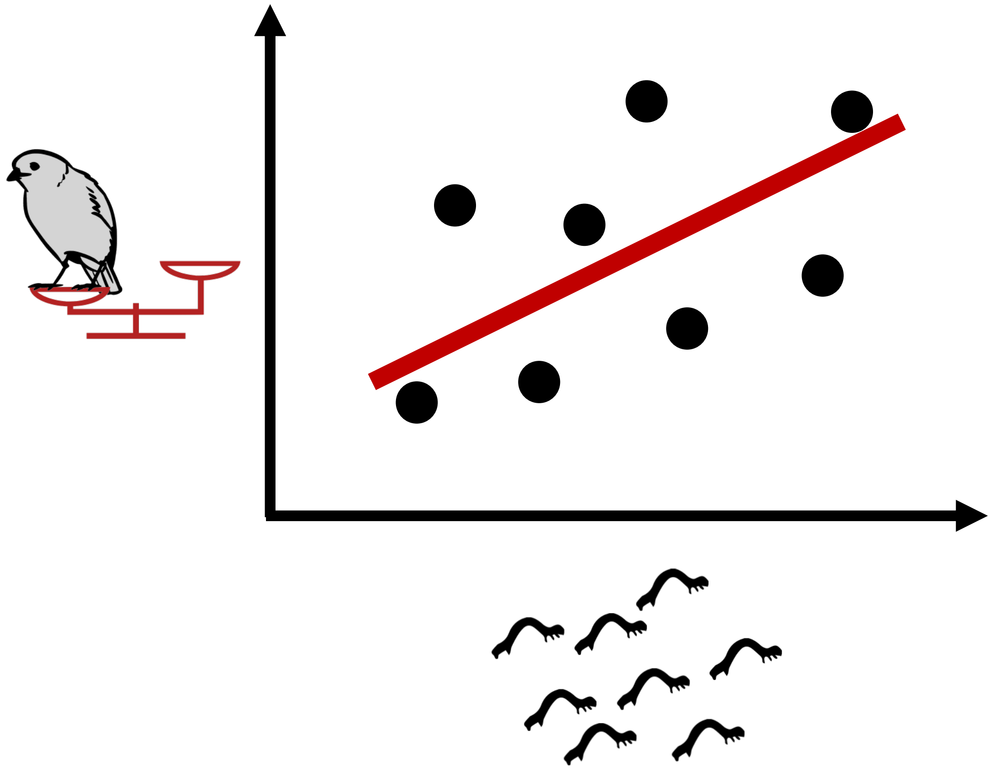
Expectation



Choosing priors

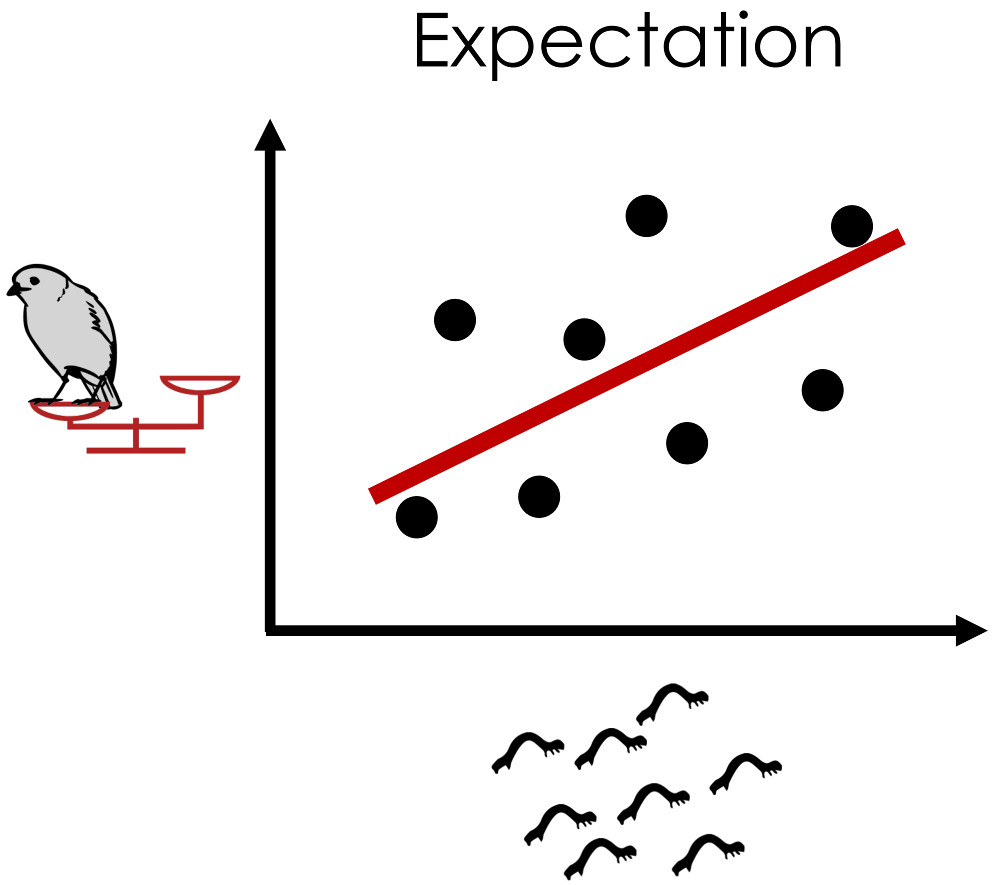
$$y_i \sim Normal(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \times x_i$$

Expectation



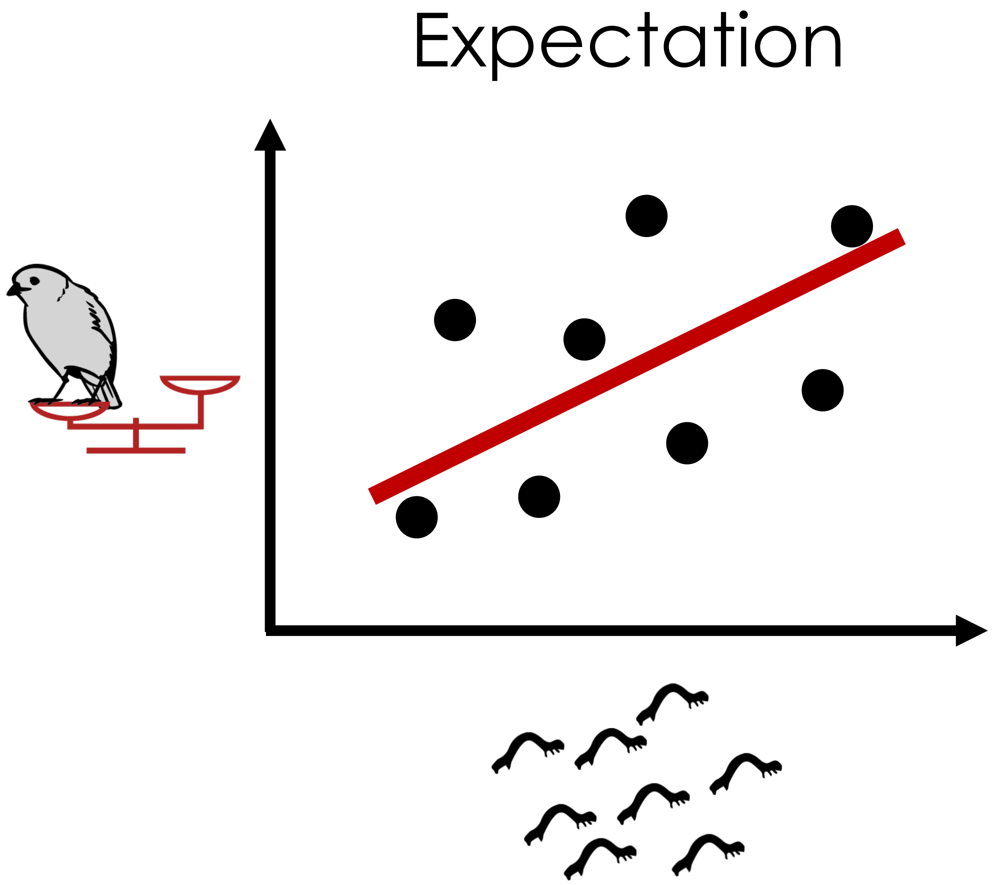
Choosing priors

$$y_i \sim Normal(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \times x_i$$
$$\alpha \sim Normal(?, ?)$$



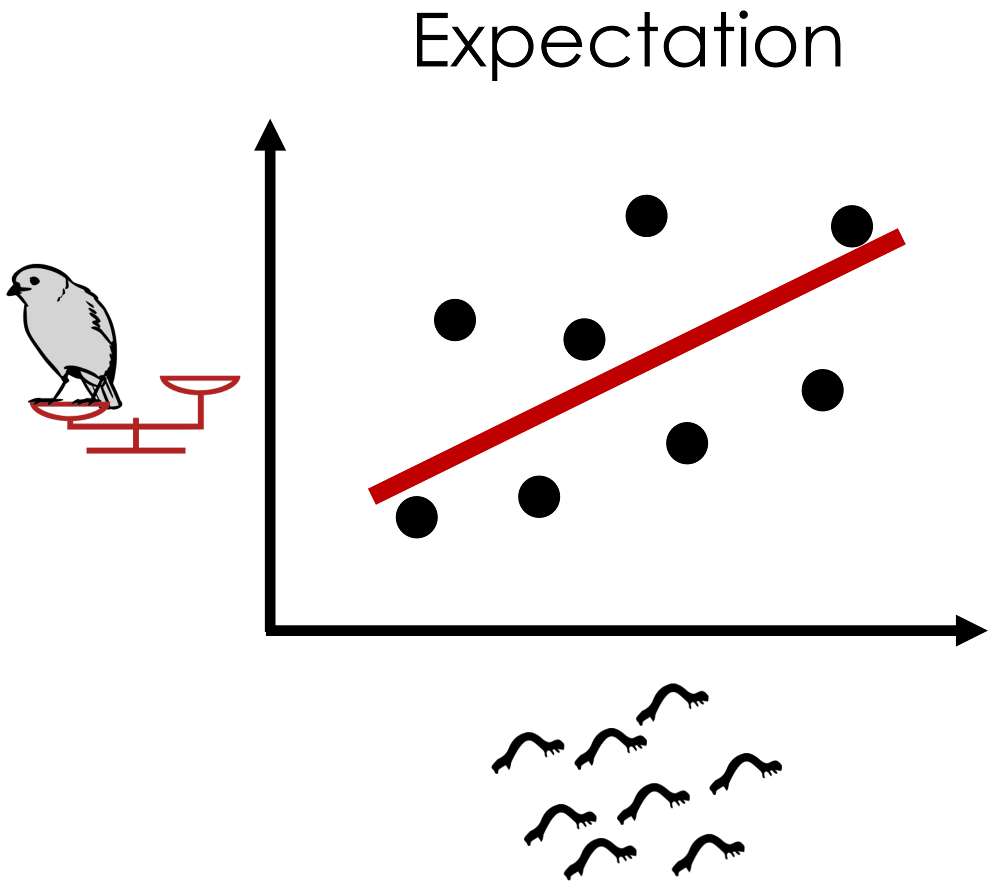
Choosing priors

$$y_i \sim Normal(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \times x_i$$
$$\alpha \sim Normal(?, ?)$$
$$\beta \sim Normal(?, ?)$$



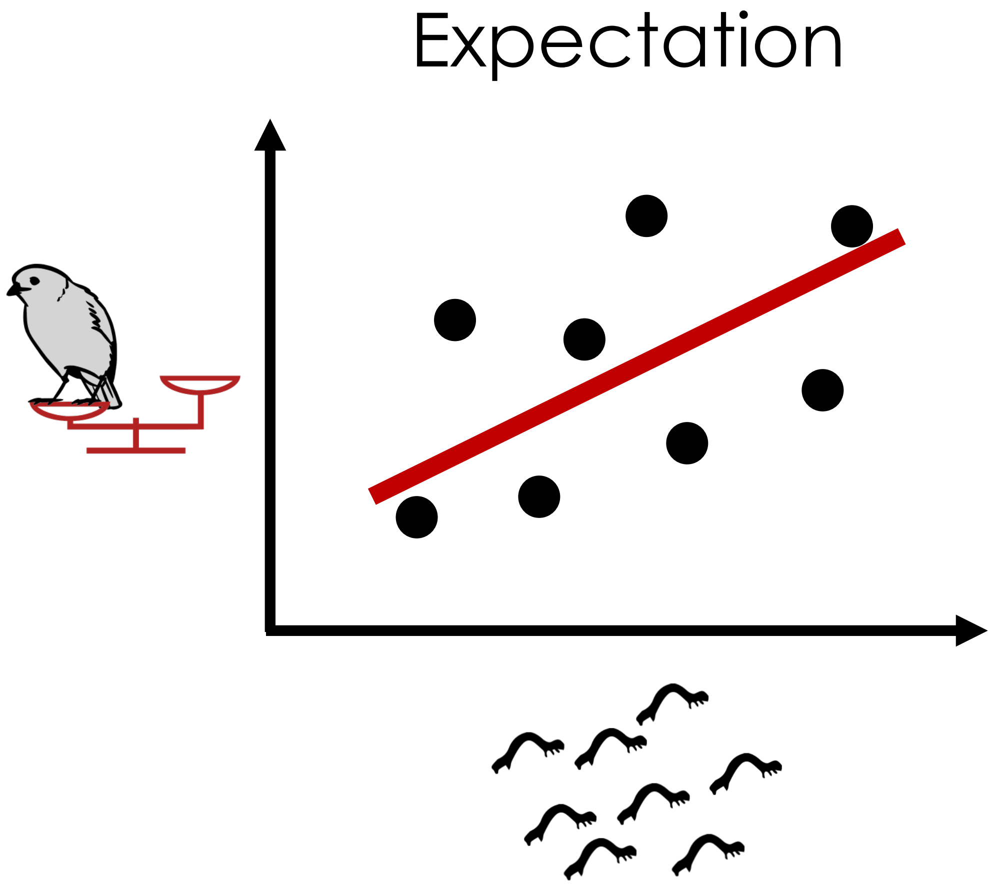
Choosing priors

$$y_i \sim Normal(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \times x_i$$
$$\alpha \sim Normal(?, ?)$$
$$\beta \sim Normal(?, ?)$$
$$\sigma \sim HalfNormal(?, ?)$$



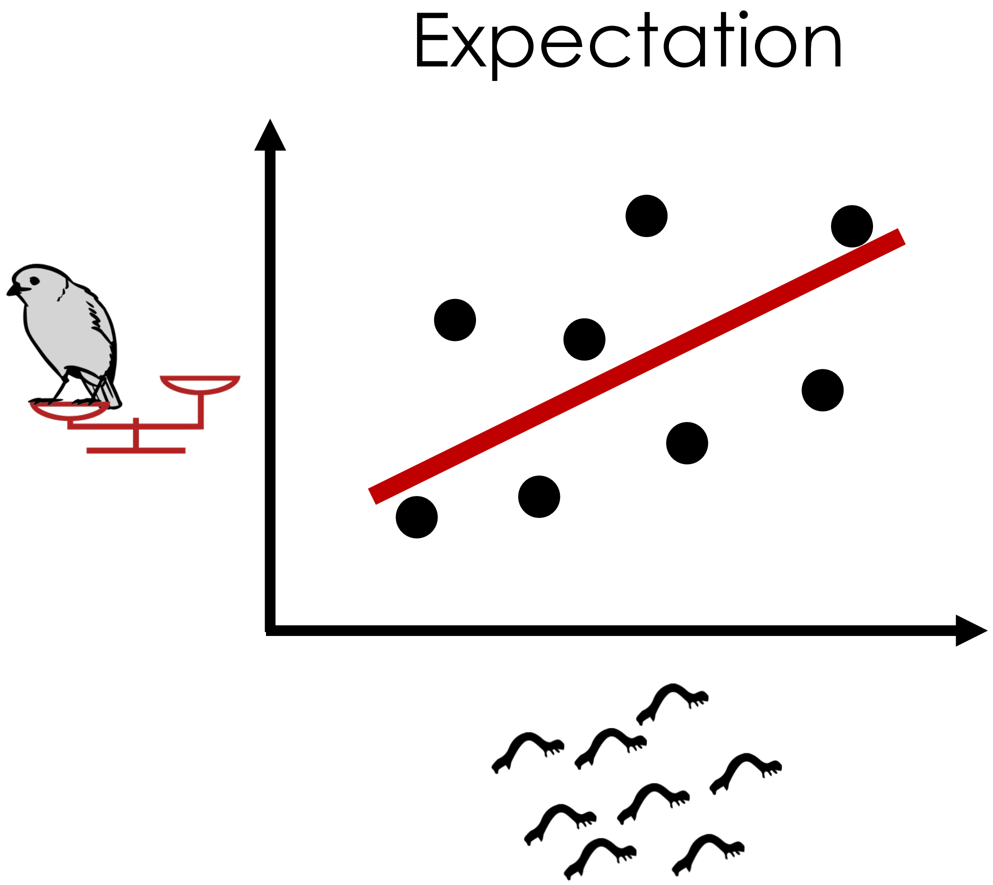
Choosing priors

$$y_i \sim Normal(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \times x_i$$
$$\alpha \sim Normal(50, 15)$$
$$\beta \sim Normal(?, ?)$$
$$\sigma \sim HalfNormal(?, ?)$$



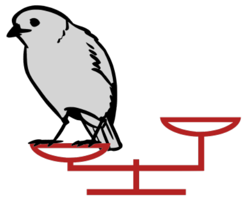
Choosing priors

$$y_i \sim Normal(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \times x_i$$
$$\alpha \sim Normal(50, 15)$$
$$\beta \sim Normal(0, 10)$$
$$\sigma \sim HalfNormal(?, ?)$$

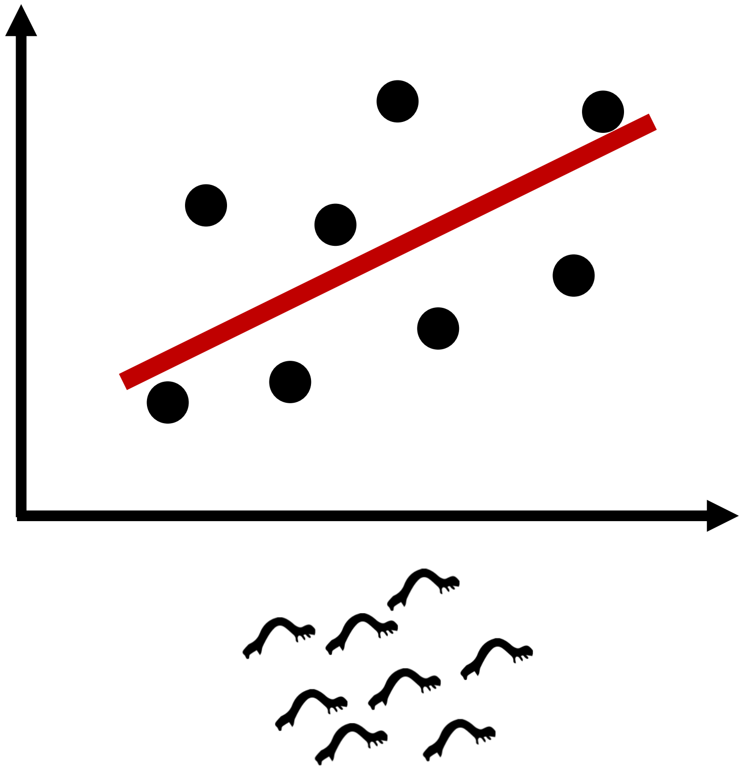


Choosing priors

$$y_i \sim Normal(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \times x_i$$
$$\alpha \sim Normal(50, 15)$$
$$\beta \sim Normal(0, 10)$$
$$\sigma \sim HalfNormal(0, 10)$$



Expectation



$$y_i \sim Normal(\mu_i, \sigma) \text{ - - - - } Pr(y, x | \alpha, \beta, \sigma)$$

$$\mu_i = \alpha + \beta \times x_i$$

$$\alpha \sim Normal(50, 15) \text{ - - - - - - - } Pr(\alpha)$$

$$\beta \sim Normal(0, 10) \text{ - - - - - - - - } Pr(\beta)$$

$$\sigma \sim HalfNormal(0, 10) \text{ - - - - - } Pr(\sigma)$$

$$
\begin{aligned}
y_i &\sim Normal(\mu_i, \sigma) \;\text{- - - -}\; \overbrace{Pr(y, x | \alpha, \beta, \sigma)}^{\text{Likelihood}} \\
\mu_i &= \alpha + \beta \times x_i
\end{aligned}
$$

$$
\left.
\begin{aligned}
\alpha &\sim Normal(50, 15) \;\text{- - - - - - -}\; Pr(\alpha) \\
\beta &\sim Normal(0, 10) \;\text{- - - - - - - - -}\; Pr(\beta) \\
\sigma &\sim HalfNormal(0, 10) \;\text{- - - - -}\; Pr(\sigma)
\end{aligned}
\right\} \text{Priors}
$$

$$\overbrace{Pr(\alpha, \beta, \sigma | y, x)}^{\text{Posterior}}$$

$$\underbrace{Pr(\alpha, \beta, \sigma | y, x)}_{\text{Posterior}} \propto \underbrace{Normal(y, x | \alpha, \beta, \sigma)}_{\text{Likelihood}} \times$$

$$\overbrace{Pr(\alpha, \beta, \sigma | y, x)}^{\text{Posterior}} \propto \overbrace{Normal(y, x | \alpha, \beta, \sigma)}^{\text{Likelihood}} \times$$

$$\text{Priors} \left\{ \begin{array}{l} Normal(\alpha | 50, 15) \times \\ Normal(\beta | 0, 10) \times \\ HalfNormal(\sigma | 0, 10) \end{array} \right.$$

MCMC to the rescue!

Stan

Stan in RStudio®

$$y_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta \times x_i$$

$$\alpha \sim Normal(50, 15)$$

$$\beta \sim Normal(0, 10)$$

$$\sigma \sim HalfNormal(0, 10)$$

4-fit-linear-bird-model.stan

Fitting the model

`4-fit-linear-bird-model.R`

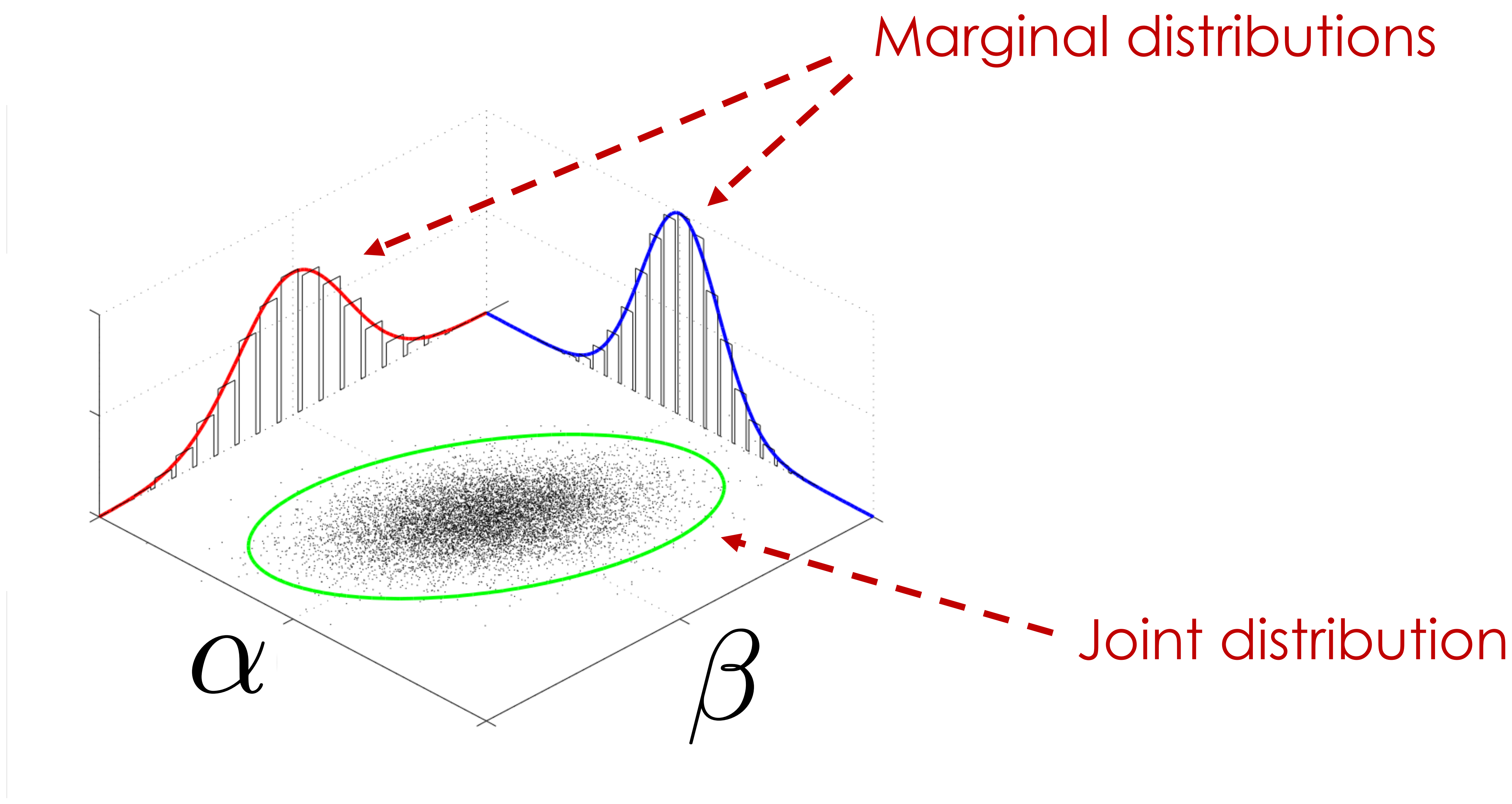
Diagnostic	Description	What you're looking for
$\hat{R}$	How well the chains 'agree' about the posterior density	<= 1.01
n.eff	Number of effective samples from posterior	> 400 for 4 chains
Posterior predictive check	Realizations of response variable generated from posterior	Simulated data looks like observed data
Divergences (Hamiltonian Monte Carlo specific)	Sampling anomaly	0 divergences

Assessing the model

4-fit-linear-bird-model.R

Summarizing the model output

`4-fit-linear-bird-model.R`


Marginal distributions
α
β
Joint distribution

Plot results

`4-fit-linear-bird-model.R`