

# Statistical Techniques for Data Science & Robotics

Week 10

# Objectives for today

- Bayesian statistics.
  - Applications of Bayesian statistics
- Sampling
  - Accept-Reject Sampling
  - Importance Sampling
- Metropolis-Hastings algorithm
  - Random walk Metropolis

# Bayesian Statistics. Source code Examples

Slides are from the “Statistical Rethinking” book by Richard McElreath

# Linear Regression

```
data(Howell1)
d <- Howell1[Howell1$age>=18,]
```

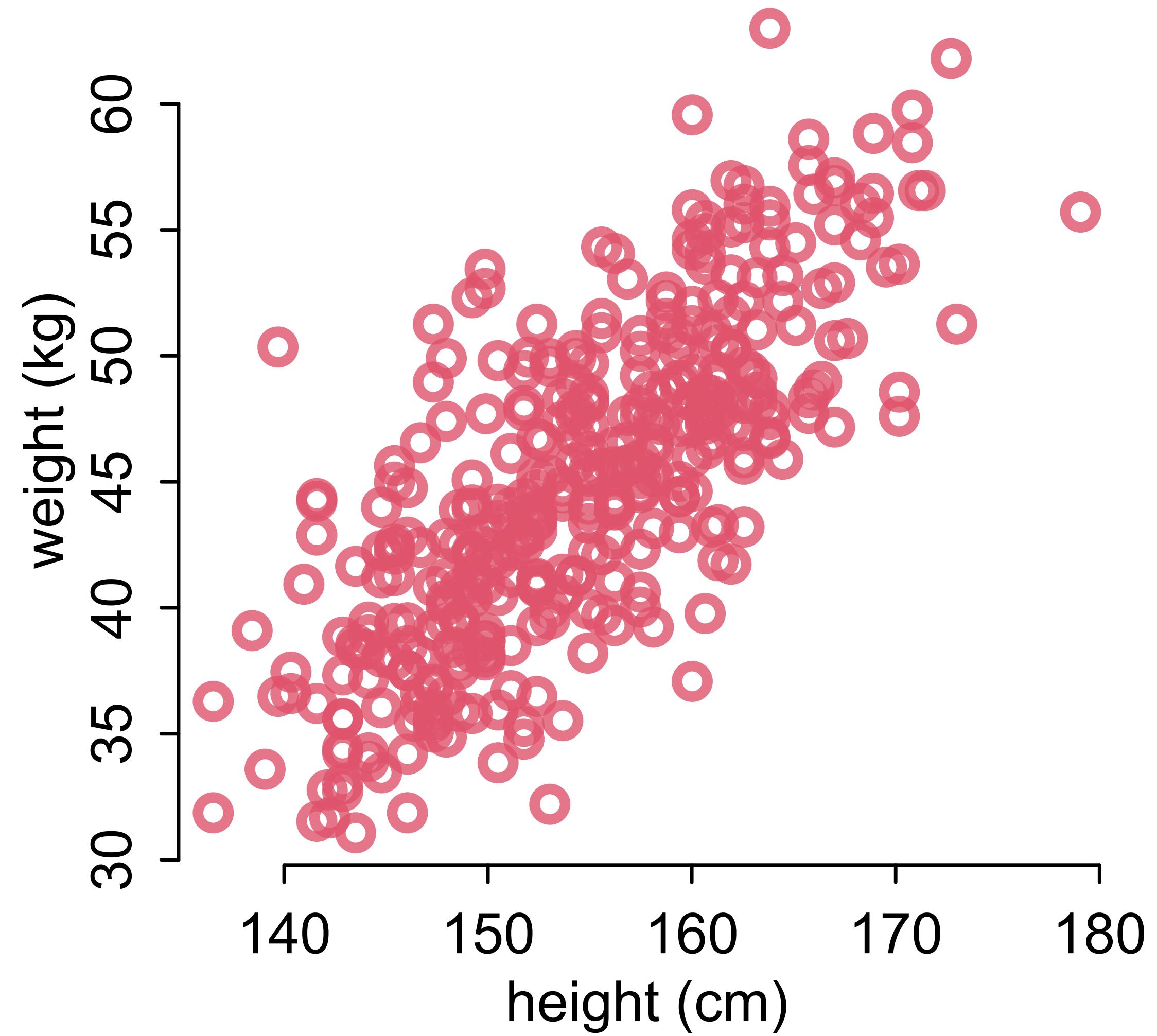
## (2) Scientific model

How does **height** influence  
**weight**?

$$H \longrightarrow W$$

$$W = f(H)$$

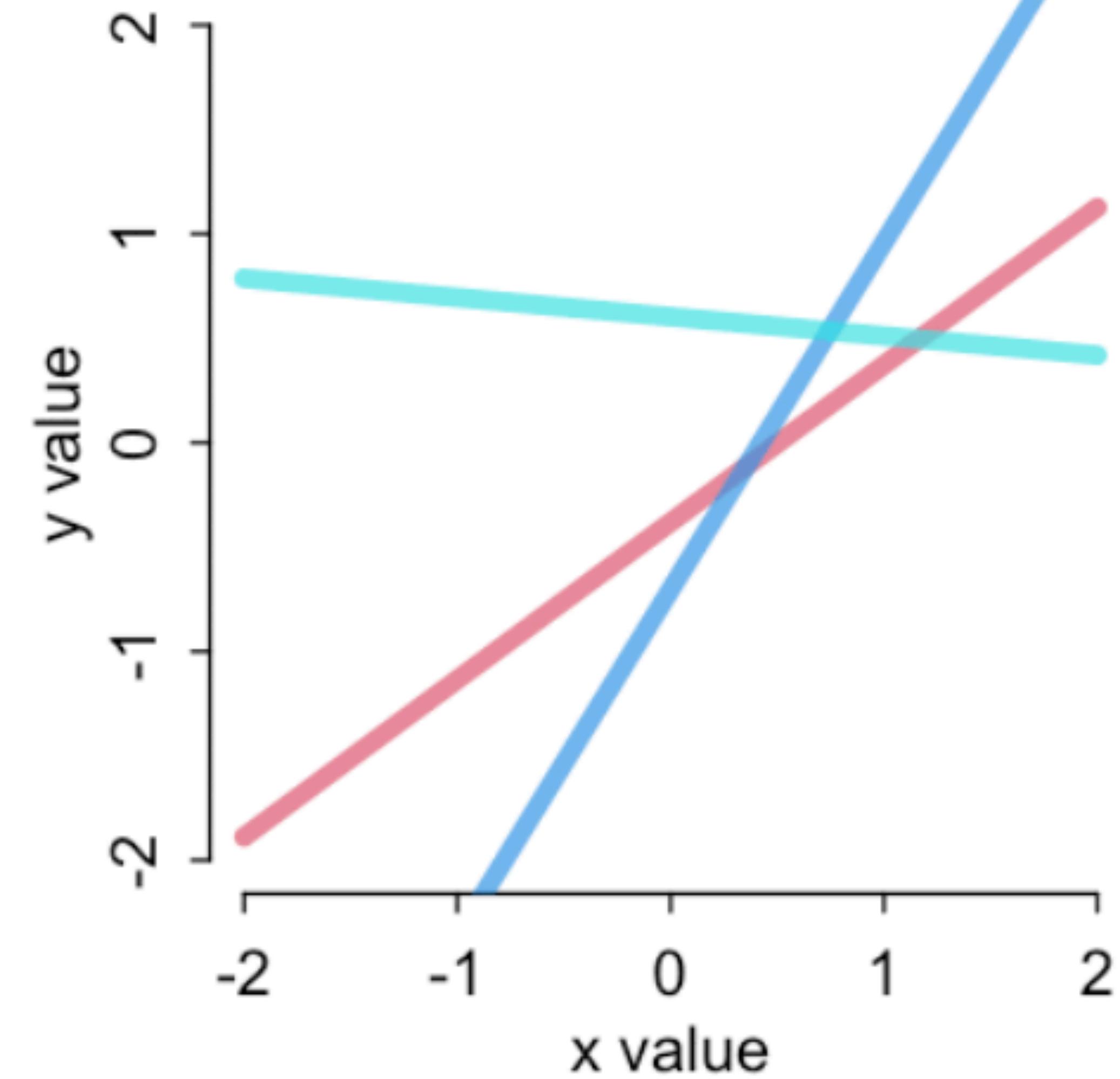
*“Weight is some function of height”*



# Anatomy of a linear model

$$y_i = \alpha + \beta x_i$$

index  
slope  
intercept



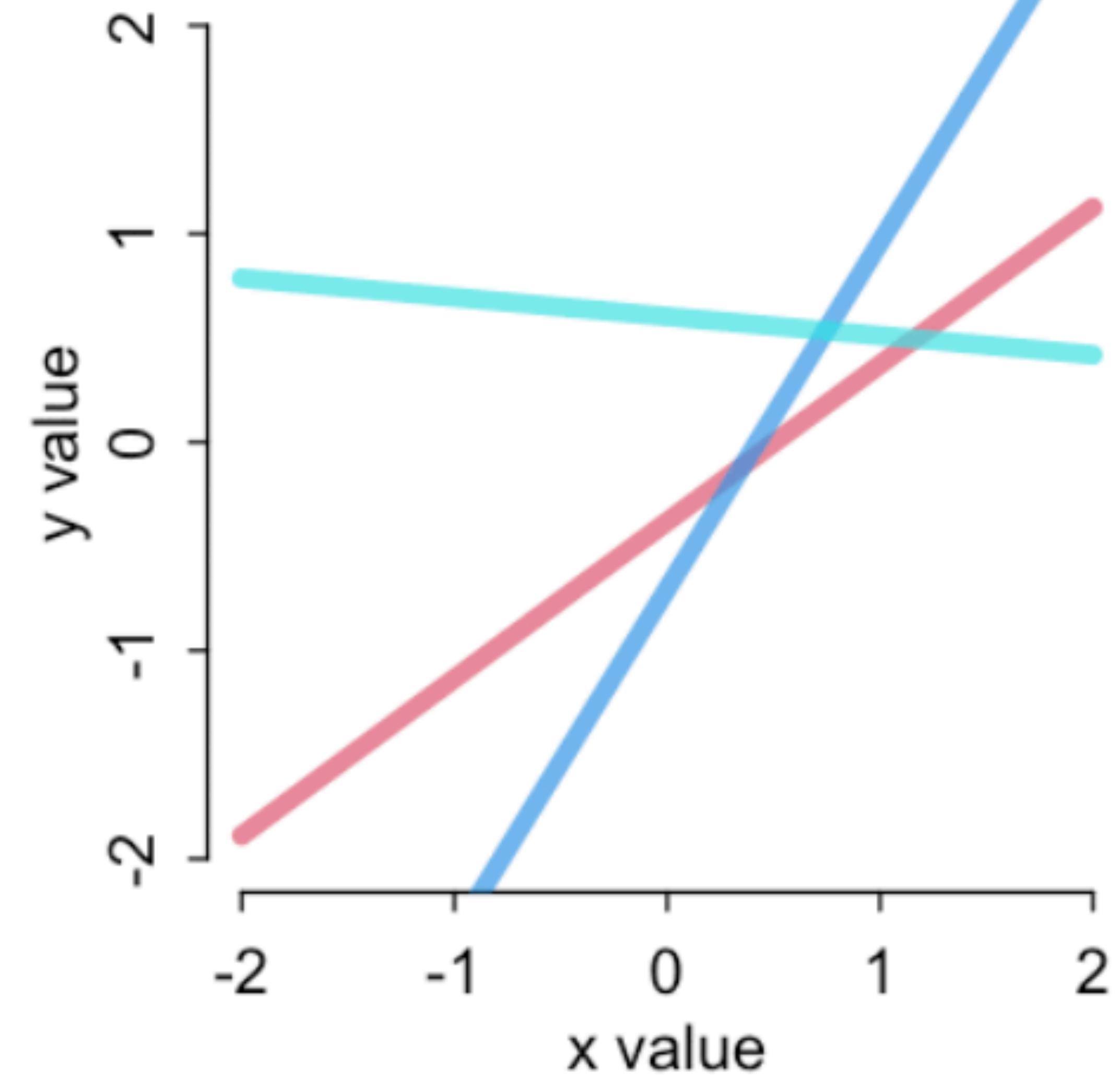
# Anatomy of a linear model

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

*expectation*      *standard deviation*

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

*“Each  $x$  value has a different expectation,  $E(y|x) = \mu$ ”*



# Generative model: $H \rightarrow W$

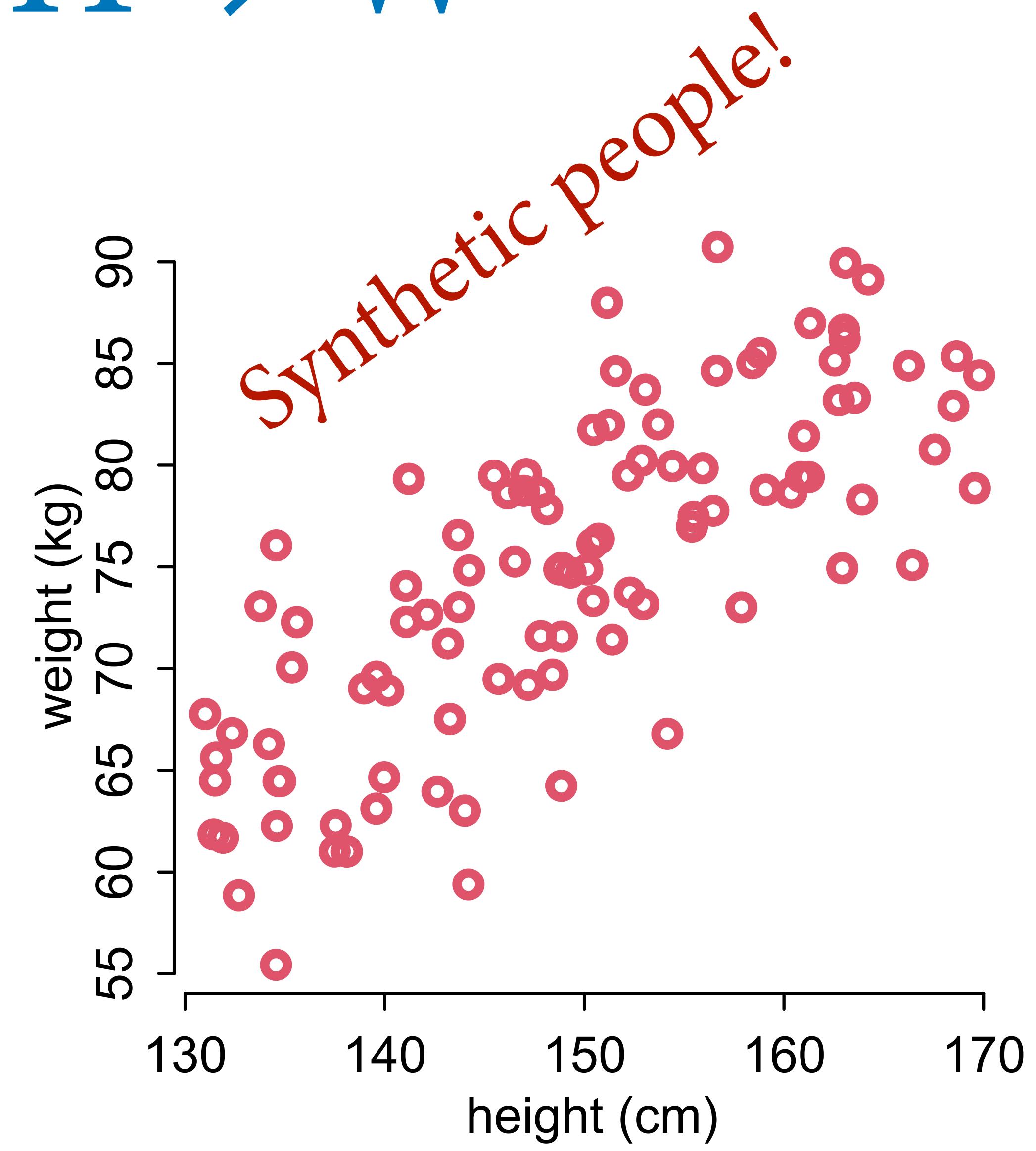
$$W_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta H_i$$

```
alpha <- 0
beta <- 0.5
sigma <- 5
n_individuals <- 100

H <- runif(n_individuals, 130, 170)

mu <- alpha + beta * H
W <- rnorm(n_individuals, mu, sigma)
```

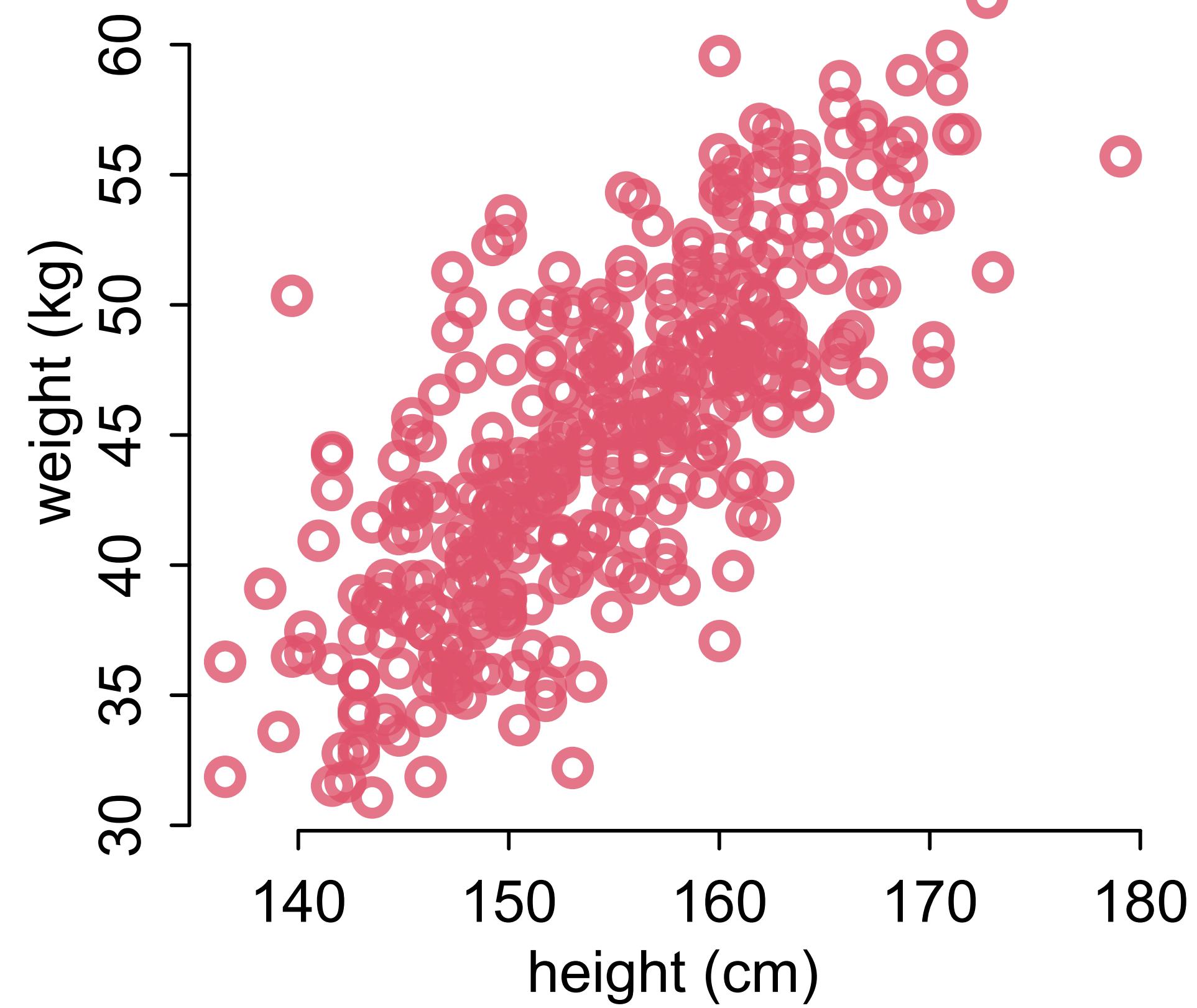


# Linear Regression

```
data(Howell1)
d <- Howell1[Howell1$age>=18,]
```

Drawing the Owl

- (1) Question/goal/estimand
- (2) Scientific model
- (3) Statistical model(s)**
- (4) Validate model
- (5) Analyze data



# Anatomy of a linear model

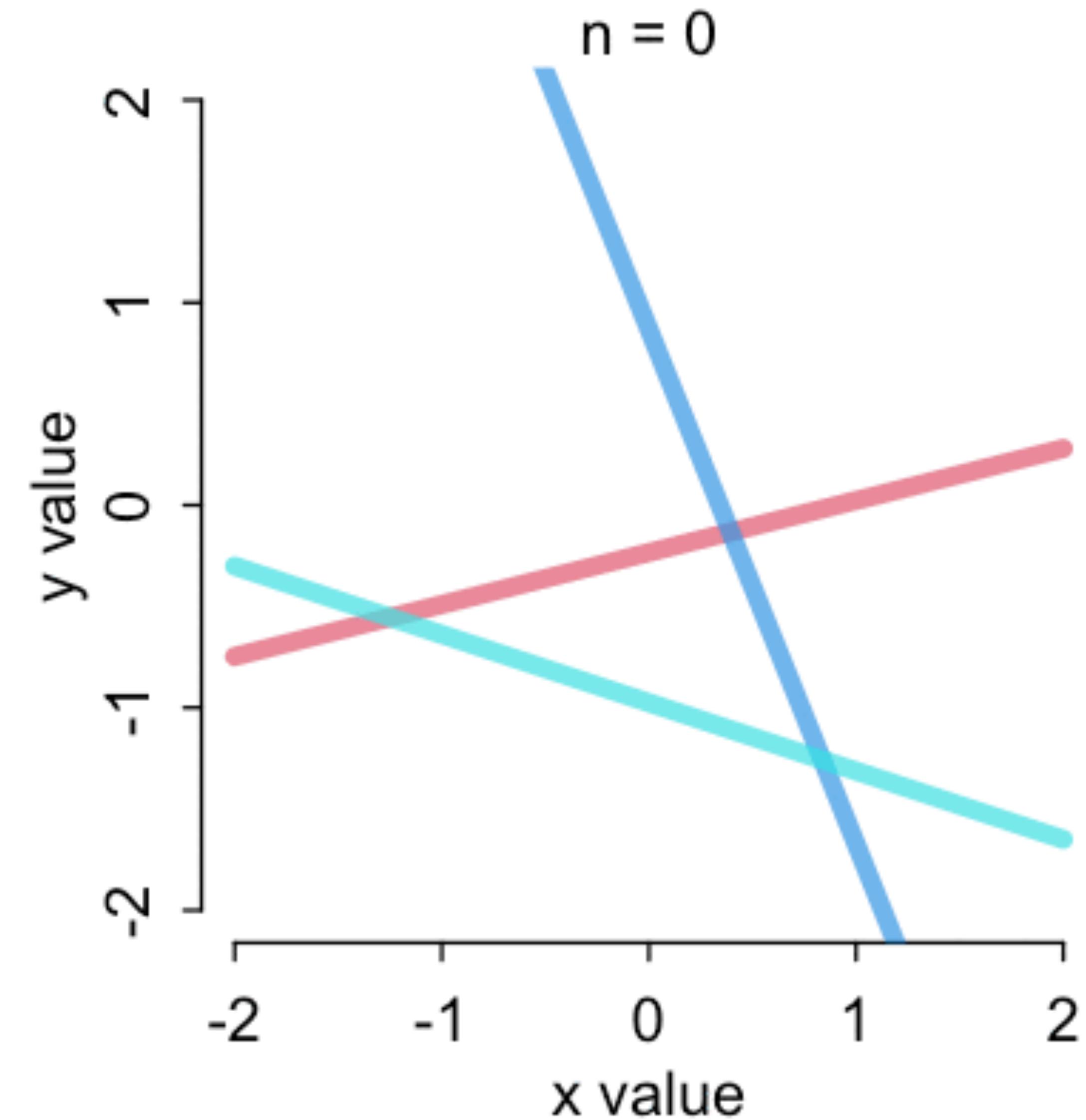
$$y_i \sim \text{Normal}(\mu_i, \sigma)$$

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

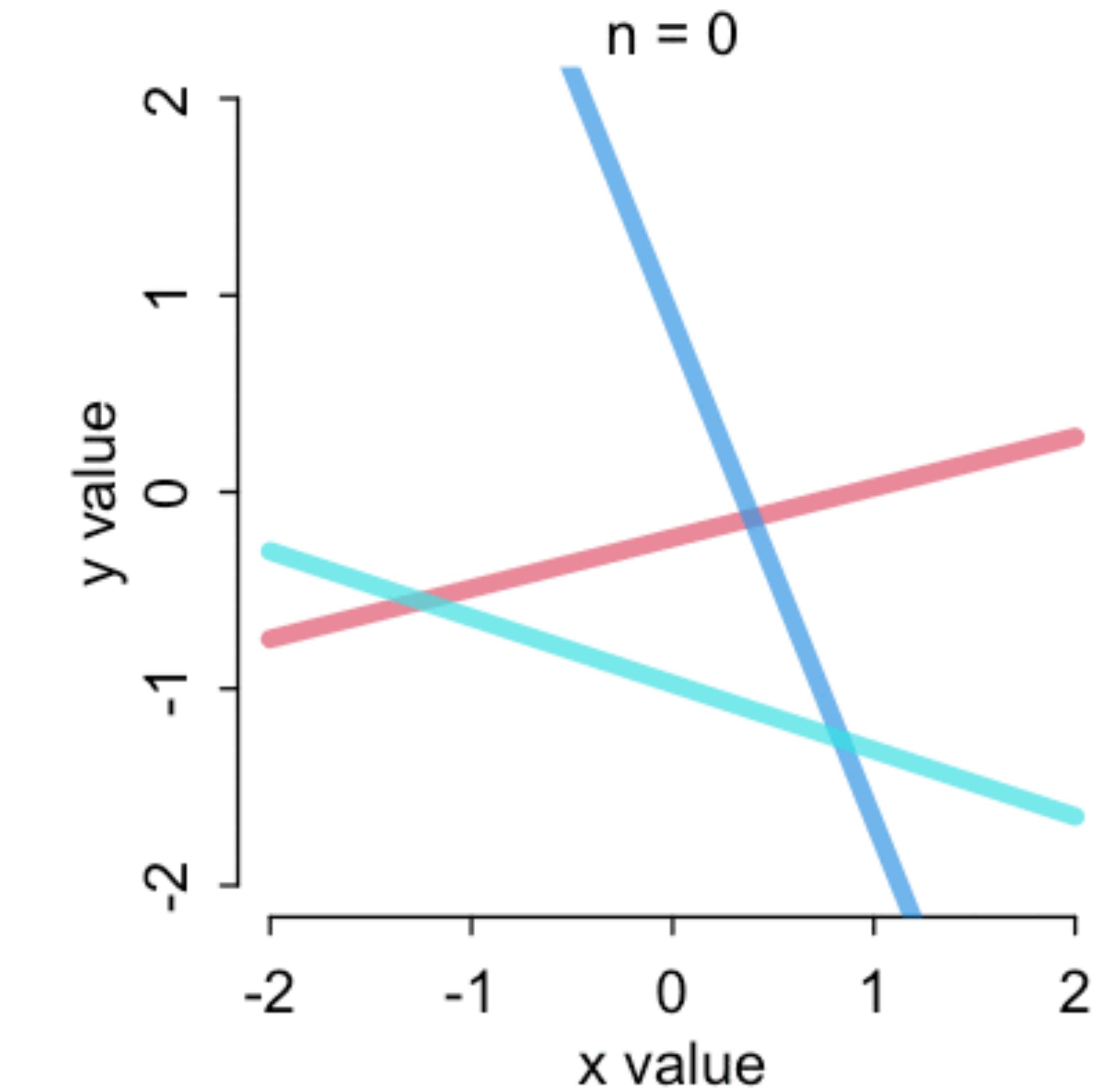
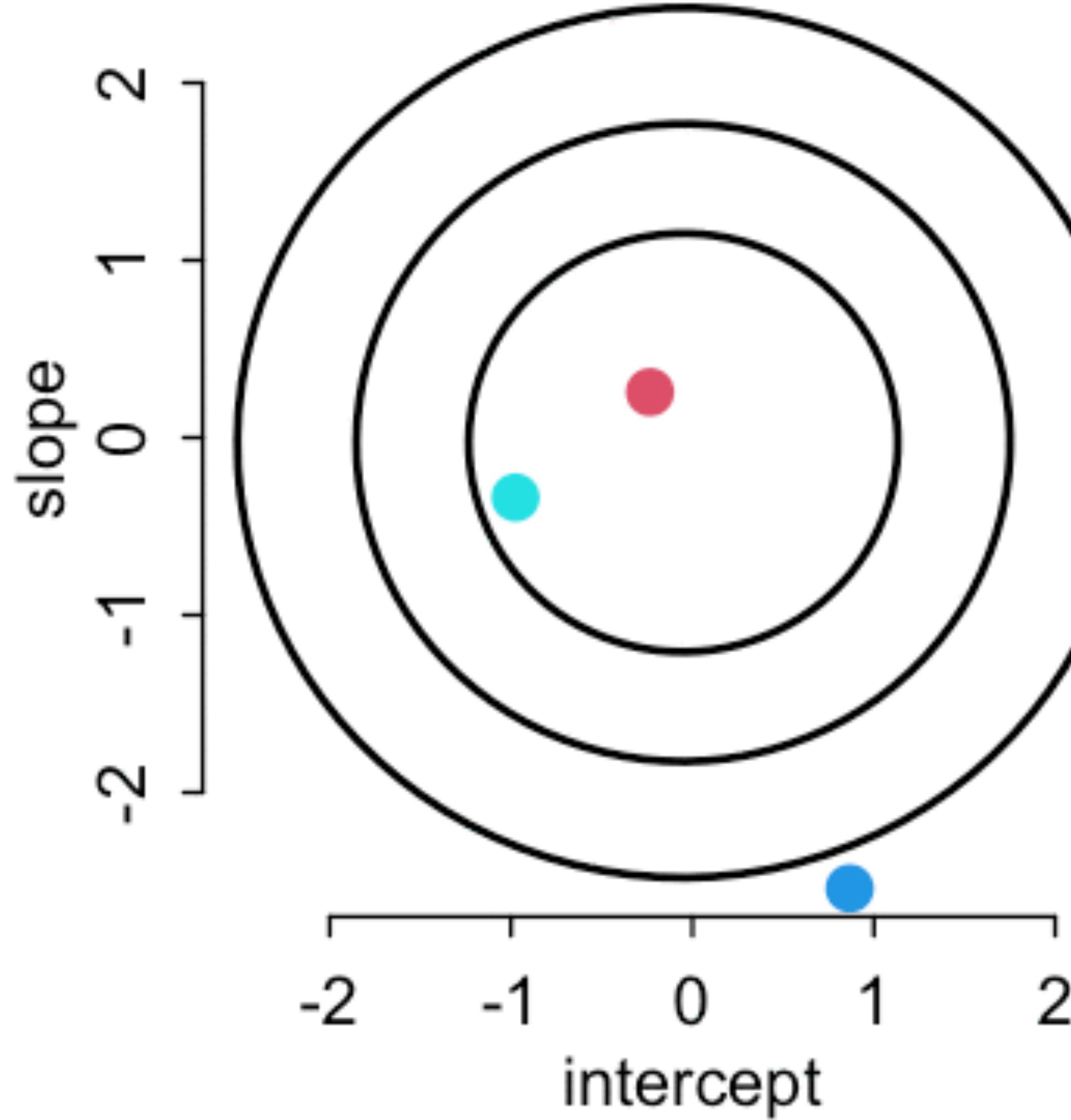
$$\alpha \sim \text{Normal}(0, 1)$$

$$\beta \sim \text{Normal}(0, 1)$$

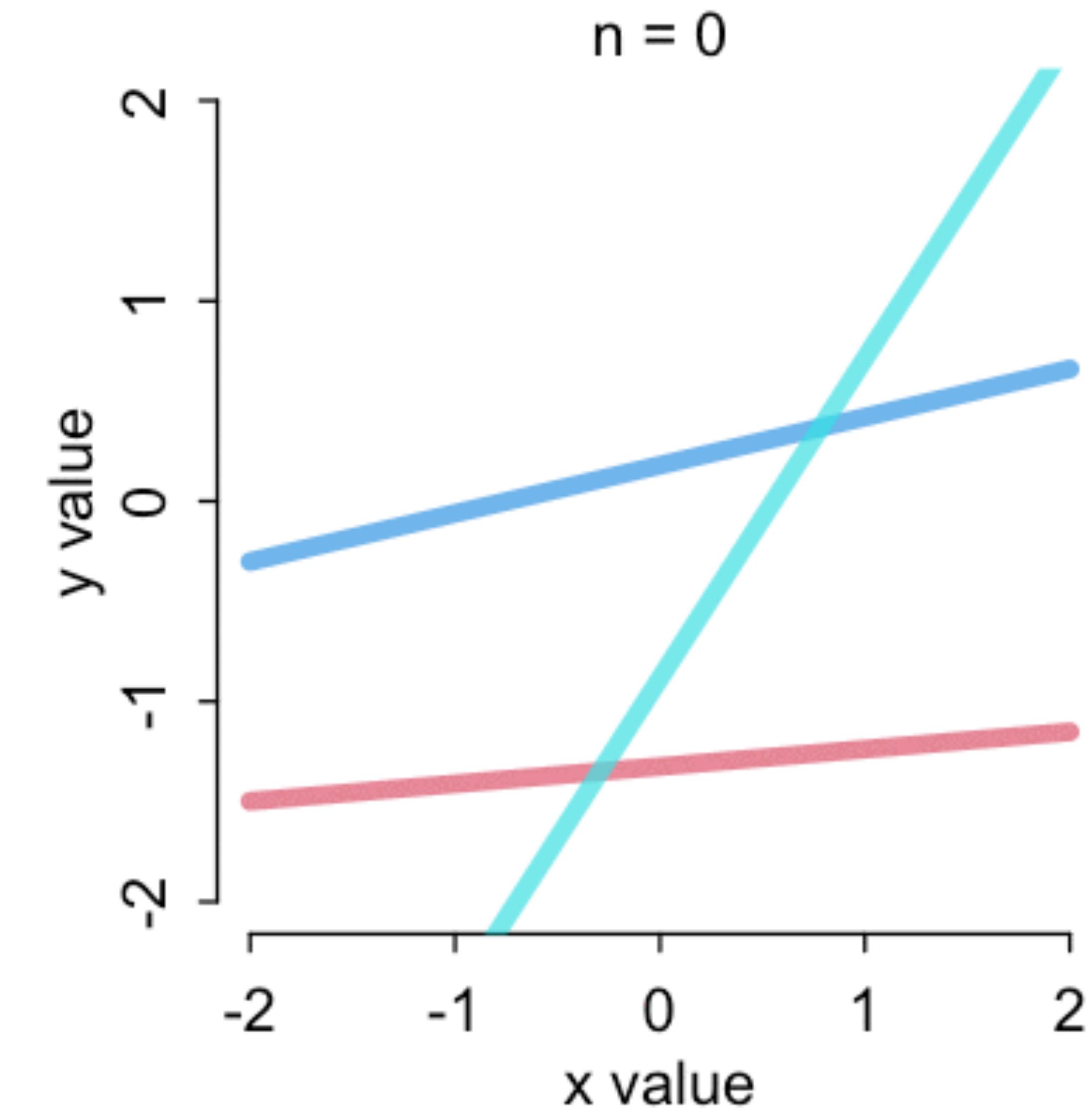
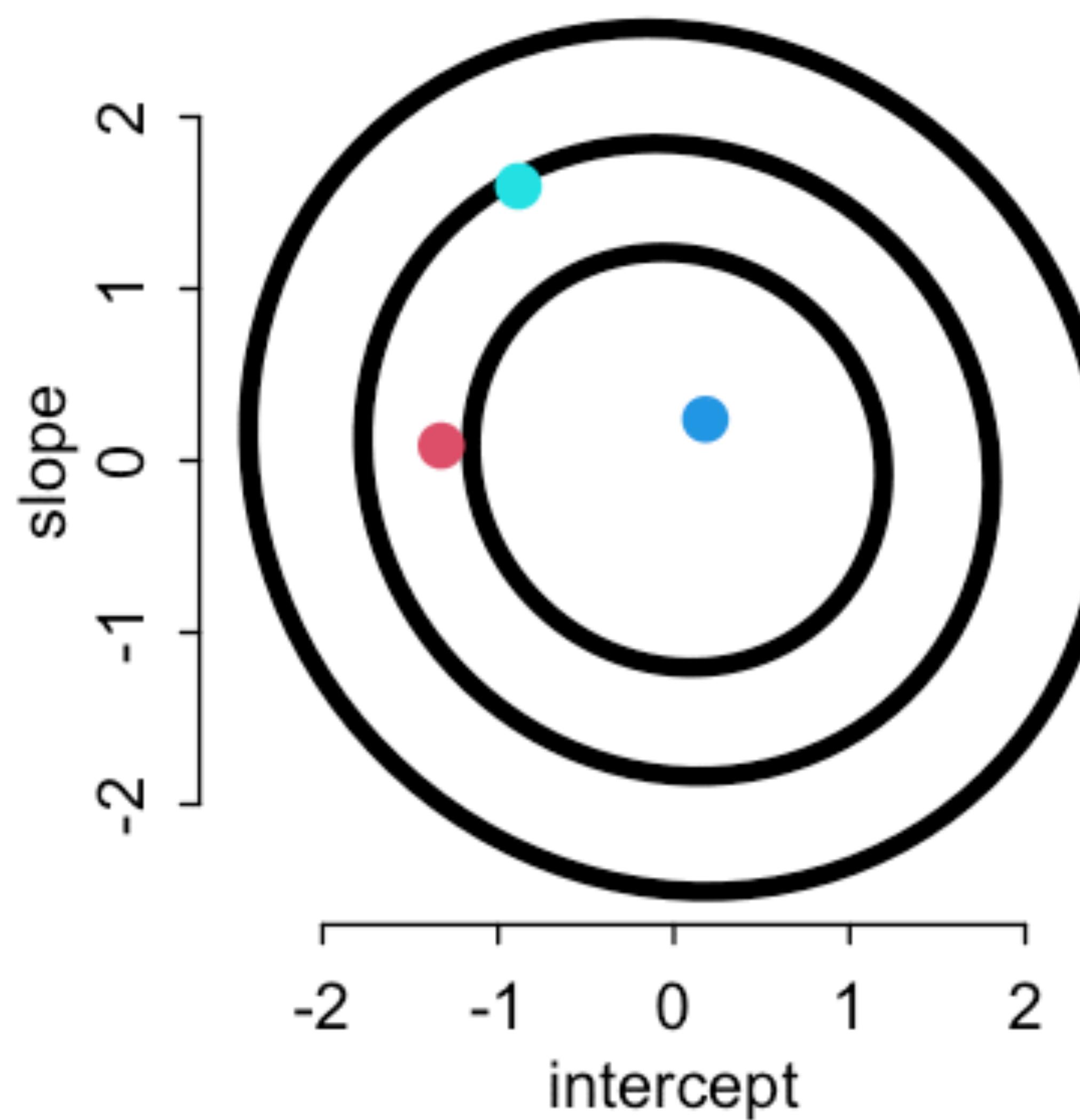
$$\sigma \sim \text{Uniform}(0, 1)$$



# Sampling the prior distribution



# Updating the posterior



# Statistical model for $H \rightarrow W$

Structure of statistical model similar to generative model, BUT

- (1) Useful to re-scale variables
- (2) Must think about priors

These two things go together

$$\begin{aligned}W_i &\sim \text{Normal}(\mu_i, \sigma) \\ \mu_i &= \alpha + \beta(H_i - \bar{H}) \\ \alpha &\sim \text{Normal}(?, ?) \\ \beta &\sim \text{Normal}(?, ?) \\ \sigma &\sim \text{Uniform}(0, ?)\end{aligned}$$

# Statistical model for $H \rightarrow W$

Re-scaling **height** so that the  
**intercept** makes sense

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

$$\mu_i = \alpha + \beta(H_i - \bar{H})$$

*value of  $\mu$  when  
 $H_i - \bar{H} = 0$*

*mean value of  $H_i$*

# Statistical model for $H \rightarrow W$

Now what are scientifically reasonable priors?

$\alpha$ : average adult weight

$\beta$ : kilograms per centimeter

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

$$\mu_i = \alpha + \beta(H_i - \bar{H})$$

$$\alpha \sim \text{Normal}(60, 10)$$

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

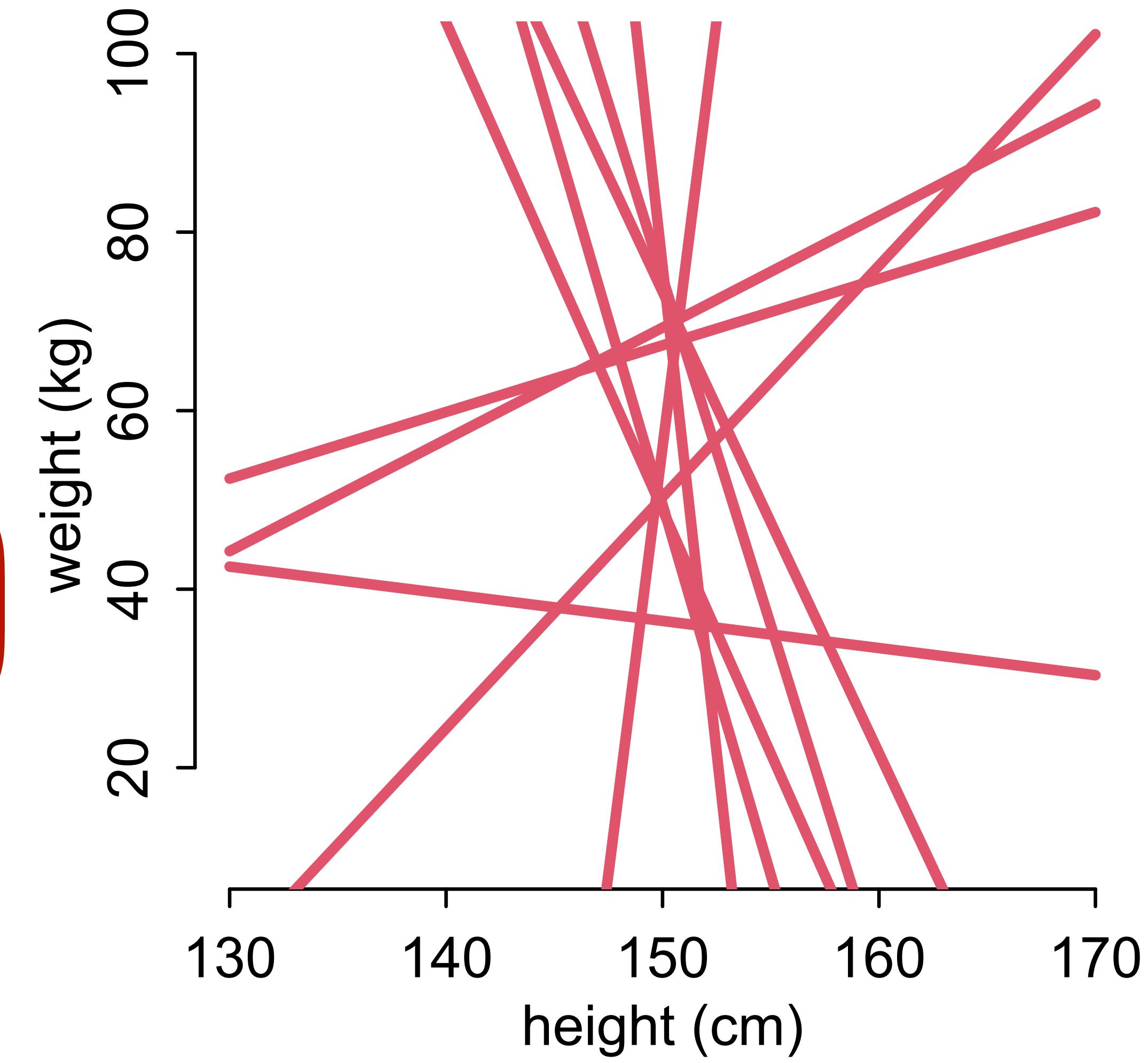
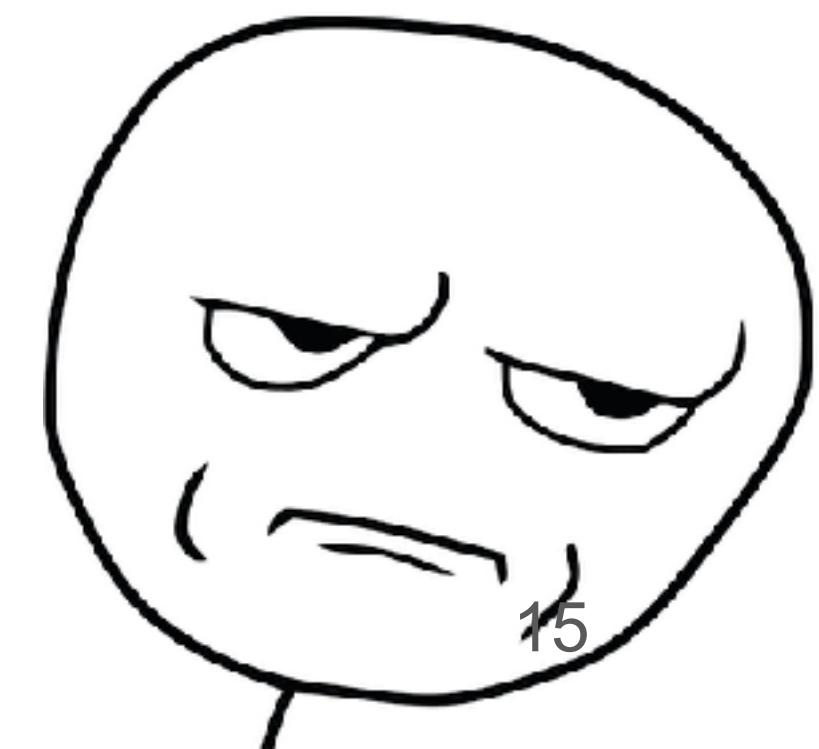
$$\sigma \sim \text{Uniform}(0, 10)$$

Region	Adult population (millions)	Average weight
Africa	535	60.7 kg (133.8 lb)
Asia	2,815	57.7 kg (127.2 lb)

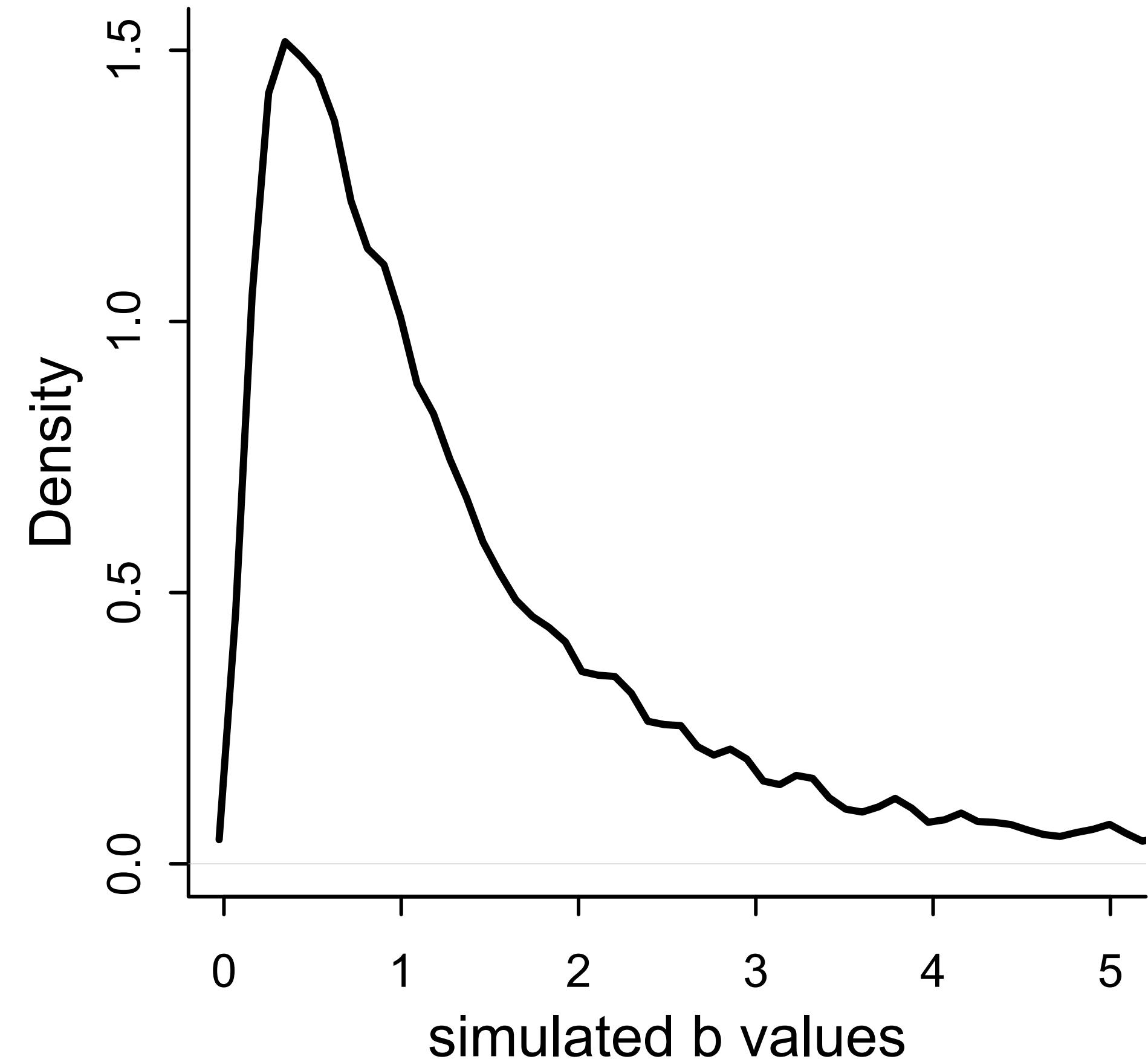
# Sampled regression lines

```
n <- 10
alpha <- rnorm(n,60,10)
beta <- rnorm(n,0,10)

Hbar <- 150
Hseq <- seq(from=130,to=170,len=30)
plot(NULL,xlim=c(130,170),ylim=c(10,100),
     xlab="height (cm)".vlab="weight (kg)")
for ( i in 1:n )
  lines( Hseq , alpha[i] + beta[i]*(Hseq-Hbar) ,
         lwd=3 , col=2 )
```



# Statistical model for $H \rightarrow W$



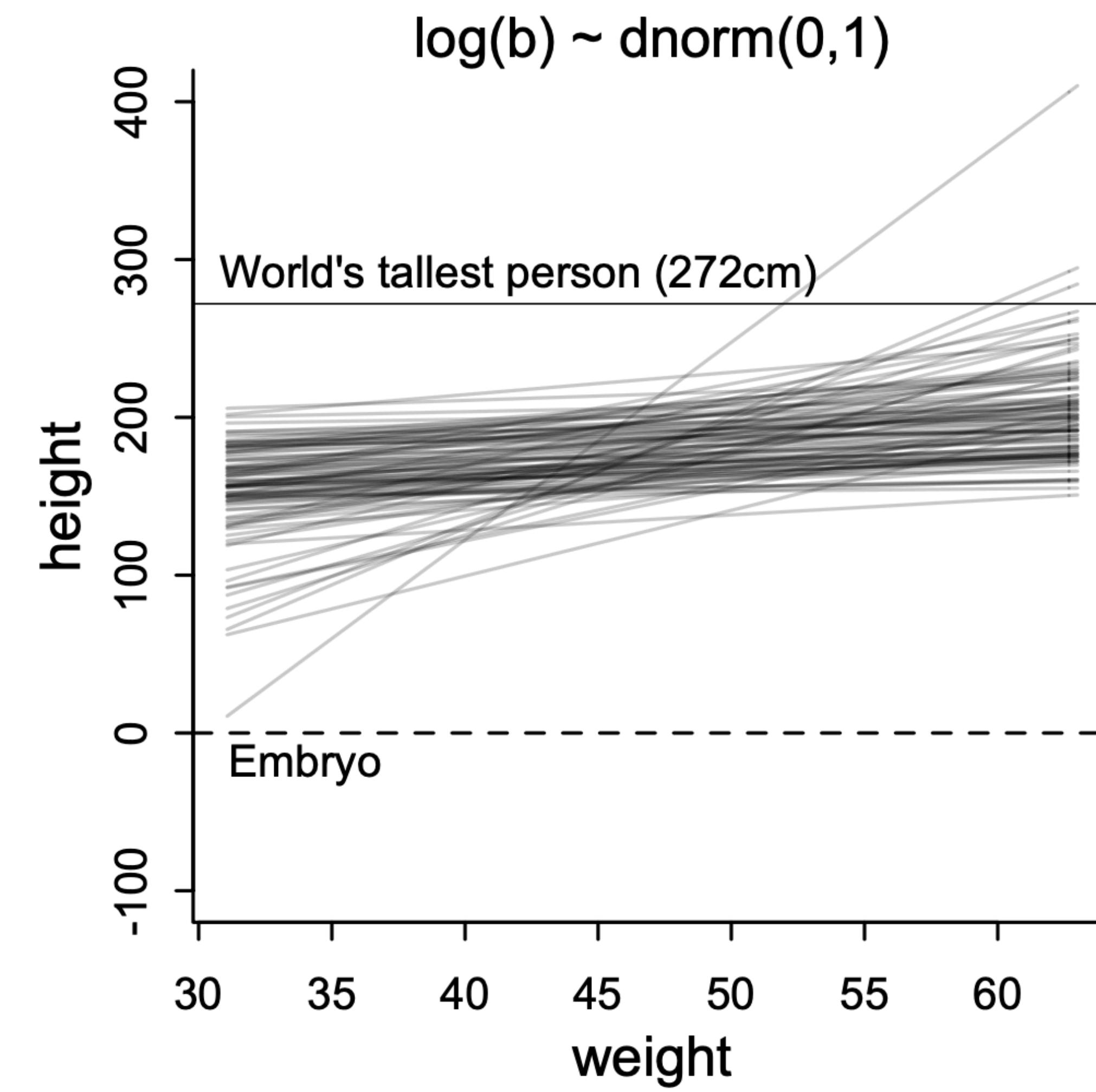
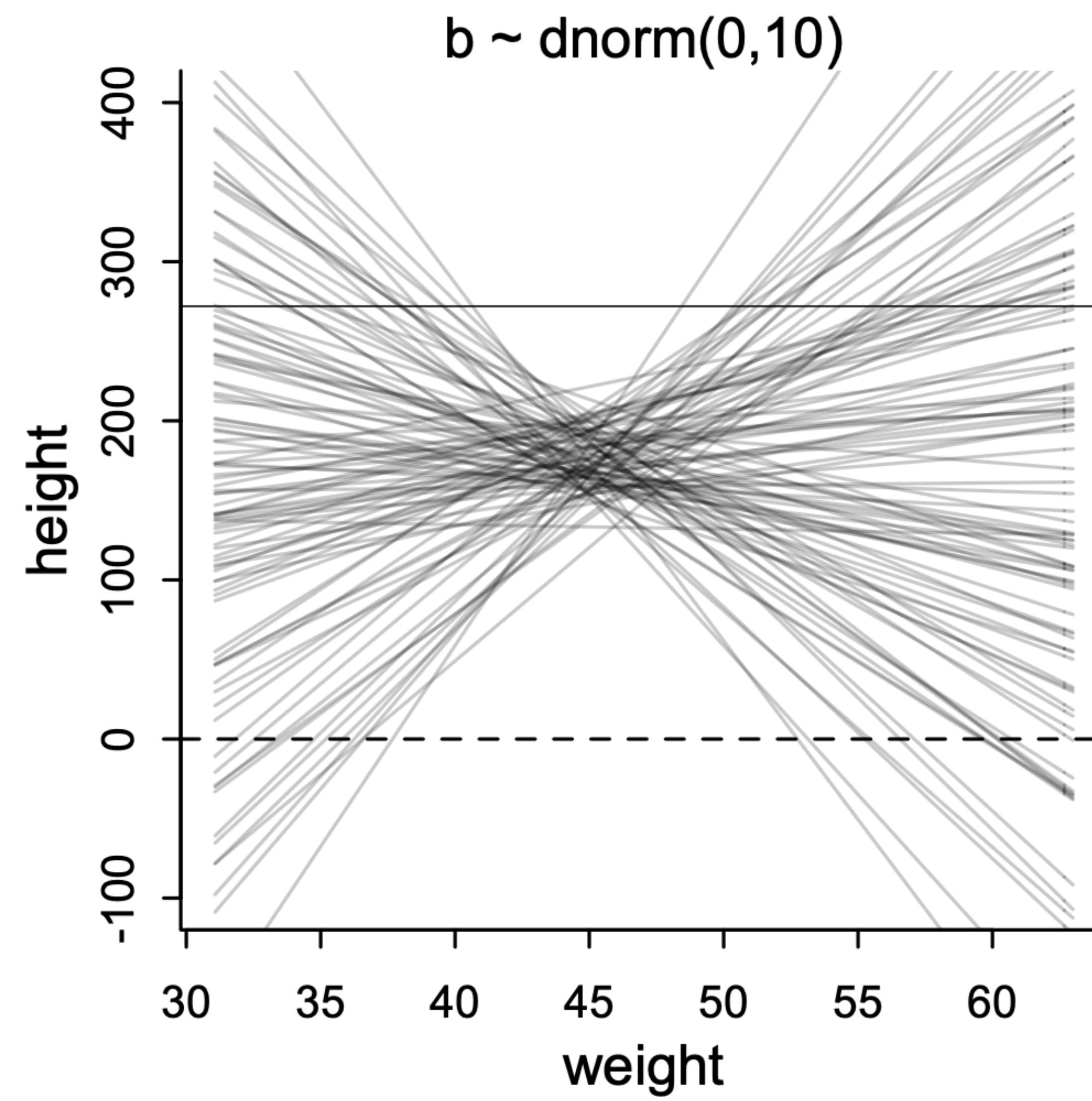
$$W_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta(H_i - \bar{H})$$

$$\alpha \sim \text{Normal}(60, 10)$$

$$\beta \sim \text{LogNormal}(0, 1)$$

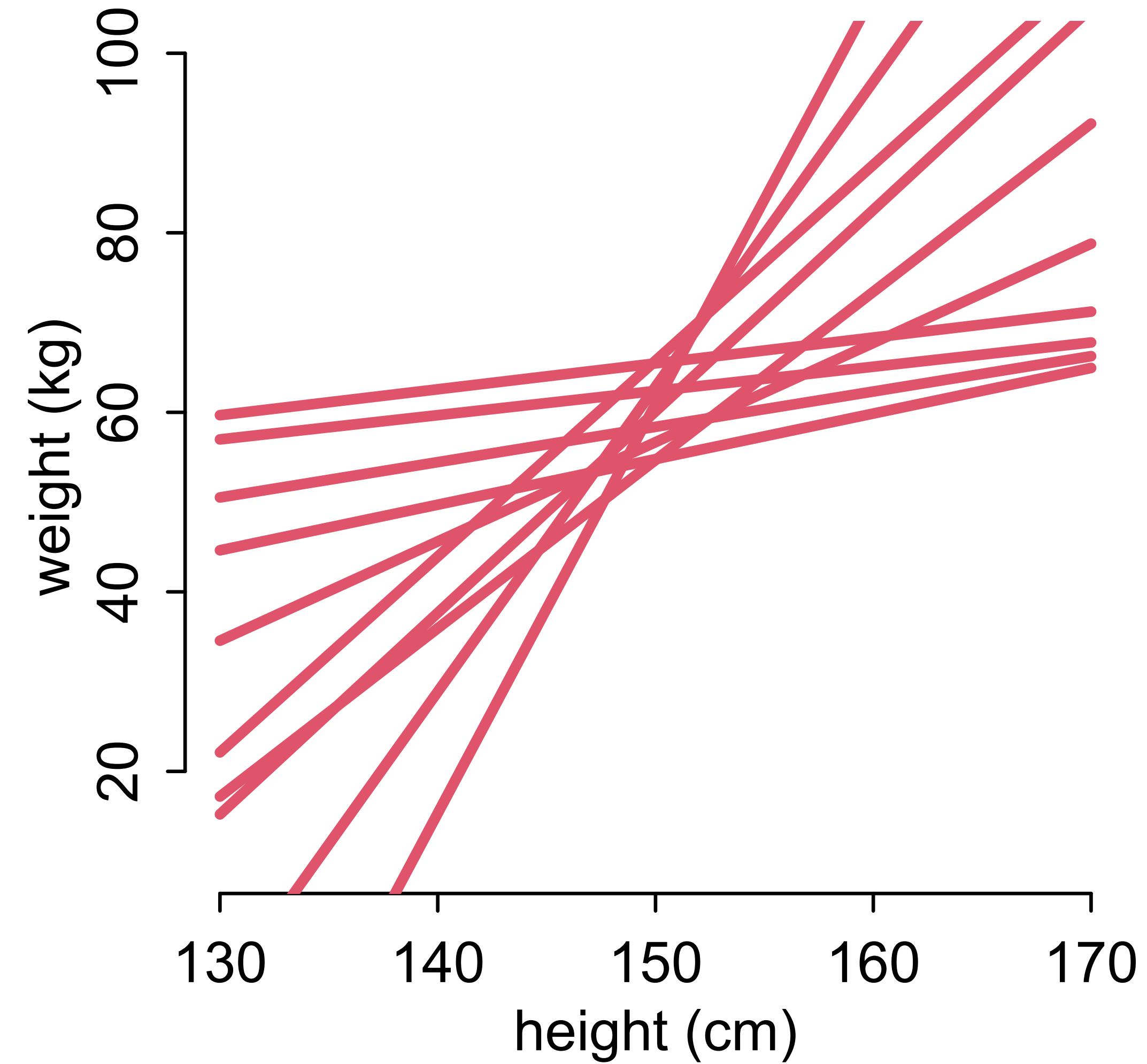
$$\sigma \sim \text{Uniform}(0, 10)$$



# Sampled regression lines

```
n <- 10
alpha <- rnorm(n,60,10)
beta <- rlnorm(n,0,1)

Hbar <- 150
Hseq <- seq(from=130,to=170,len=30)
plot(NULL,xlim=c(130,170),ylim=c(10,100),
  xlab="height (cm)",ylab="weight (kg)")
for ( i in 1:n )
  lines( Hseq , alpha[i] + beta[i]*(Hseq-Hbar) ,
  lwd=3 , col=2 )
```



# Fitting the model

$$\begin{aligned}\Pr(\alpha, \beta, \sigma | W, H) \propto & \text{Normal}(W | \mu, \sigma) \\ & \times \text{Normal}(\alpha | 60, 10) \\ & \times \text{LogNormal}(\beta | 0, 1) \\ & \times \text{Uniform}(\sigma | 0, 10)\end{aligned}$$

Grid approximation expensive:  
100 values of each parameter => 1 million calculations

# Linear Regression

```
data(Howell1)
d <- Howell1[Howell1$age>=18,]
```

Drawing the Owl

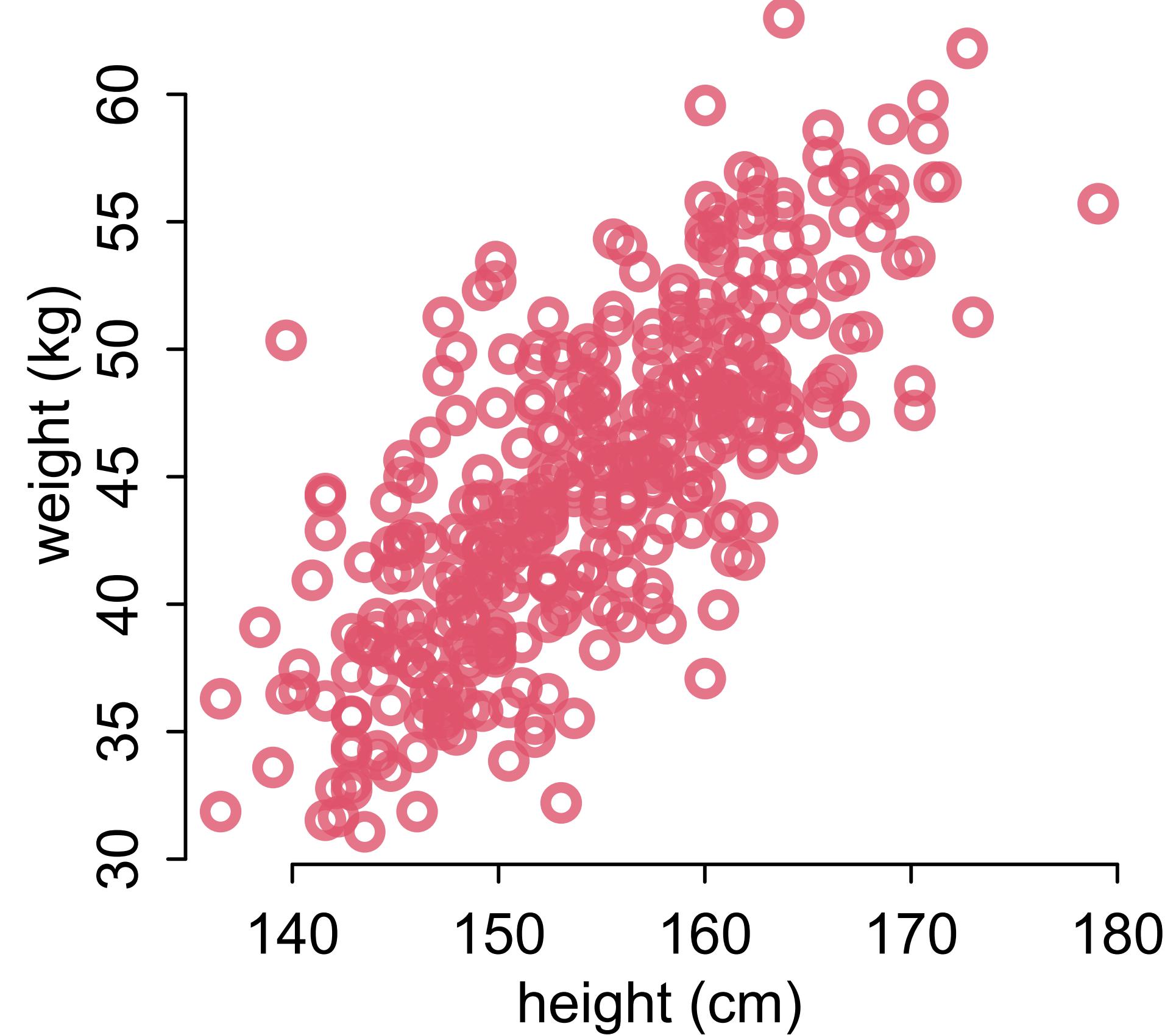
(1) Question/goal/estimand

(2) Scientific model

(3) Statistical model(s)

**(4) Validate model**

**(5) Analyze data**



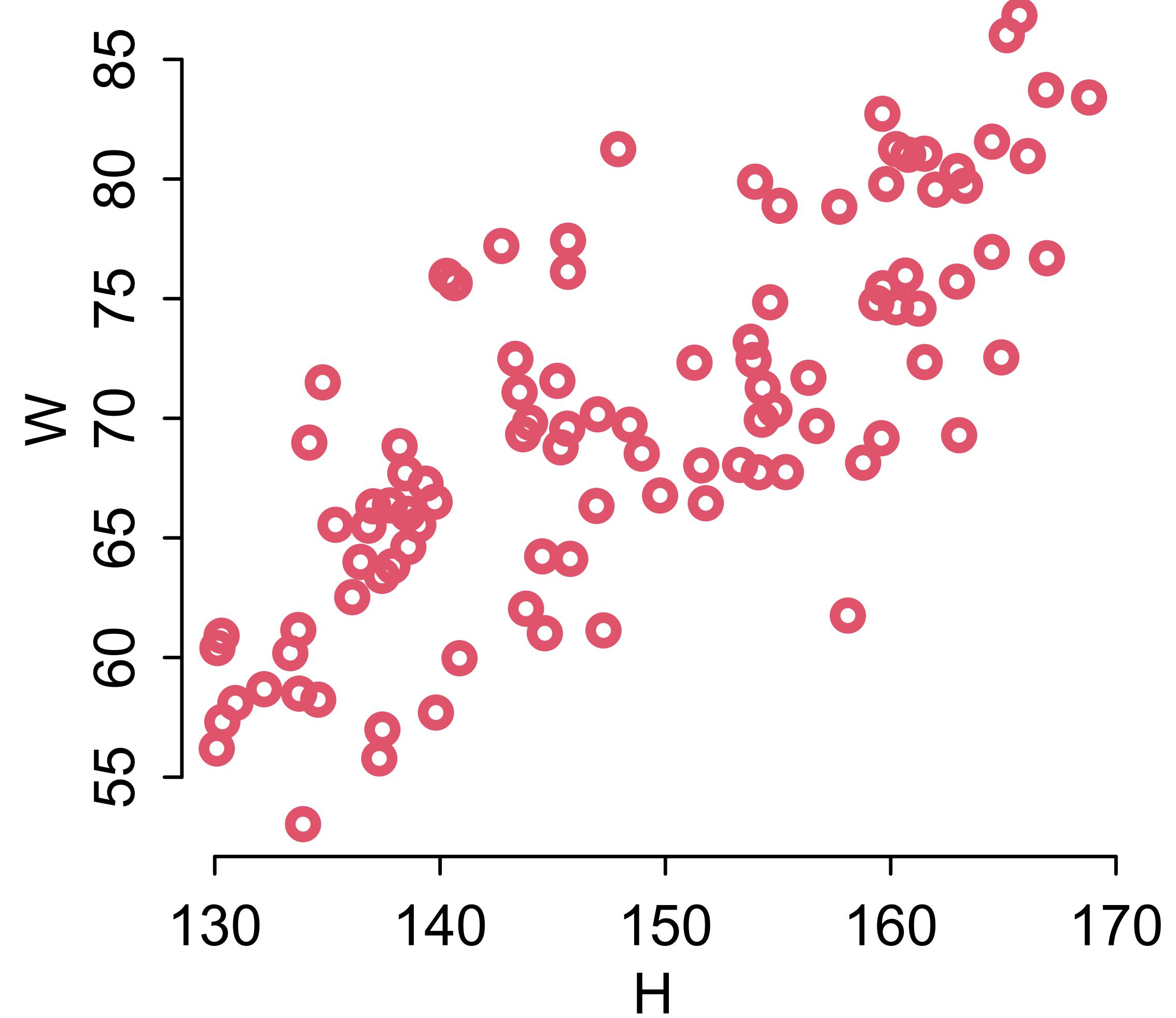
# First validate with simulation

```
alpha <- 70
beta <- 0.5
sigma <- 5
n_individuals <-
H <- runif(n_individuals,130,170)
mu <- alpha + beta*(H-mean(H))
W <- rnorm(n_individuals,mu,sigma)

dat <- list( H=H , W=W , Hbar=mean(H) )

m_validate <- quap(
  alist(
    W ~ dnorm(mu,sigma),
    mu <- a + b*(H-Hbar),
    a ~ dnorm(60,10),
    b ~ dlnorm(0,1),
    sigma ~ dunif(0,10)
  ), data=dat )
```

```
> precis(m_validate)
      mean   sd  5.5% 94.5%
a     69.57 0.45 68.85 70.28
b     0.49 0.04  0.43  0.56
sigma 4.48 0.32  3.97  4.98
```



# Now with the real data

```
data(Howell1)
d <- Howell1
d <- d[ d$age>=18 , ]

dat <- list(
  W = d$weight,
  H = d$height,
  Hbar = mean(d$height) )

m_adults <- quap(
  alist(
    W ~ dnorm(mu,sigma),
    mu <- a + b*(H-Hbar),
    a ~ dnorm(60,10),
    b ~ dlnorm(0,1),
    sigma ~ dunif(0,10)
  ), data=dat )
```

$$W_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta(H_i - \bar{H})$$
$$\alpha \sim \text{Normal}(60, 10)$$
$$\beta \sim \text{LogNormal}(0, 1)$$
$$\sigma \sim \text{Uniform}(0, 10)$$

# Obey The Law

First Law of Statistical Interpretation:

The **parameters are not independent** of one another and cannot always be independently interpreted

Instead:

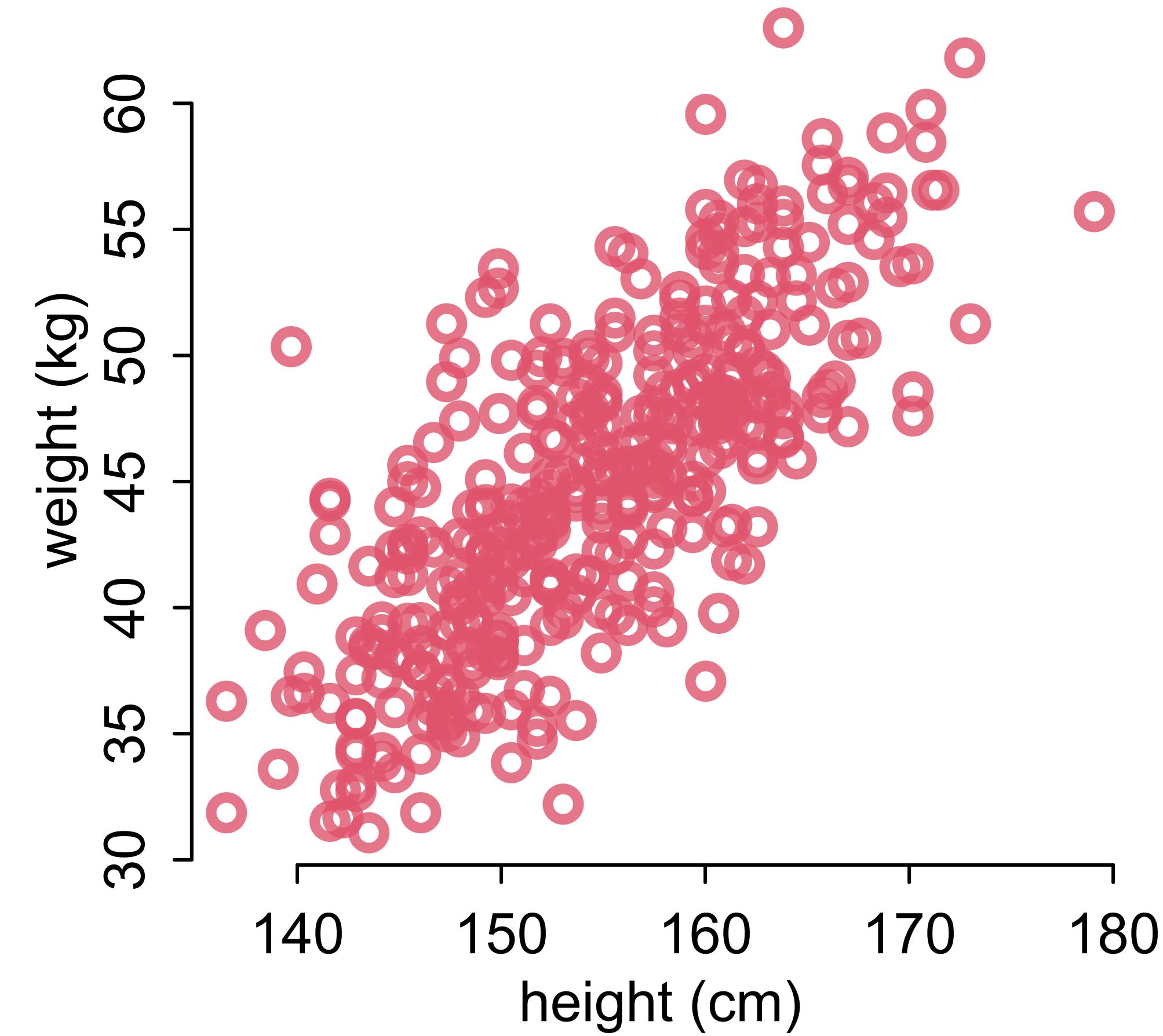
Push out **posterior predictions** and describe/interpret those

```
> precis(m_adults)
    mean   sd  5.5% 94.5%
a     45.00 0.23 44.64 45.36
b      0.63 0.03  0.58  0.68
sigma 4.23 0.16  3.97  4.48
>
```

```
> post <- extract.samples(m_adults)
> head(post)
            a      b      sigma
1 45.14733 0.7045790 4.380254
2 44.97759 0.6461353 4.372925
3 44.94856 0.6537192 4.111149
4 44.85016 0.6597310 4.379347
5 44.75898 0.6532690 4.200026
6 44.91711 0.6090434 4.105432
>
```

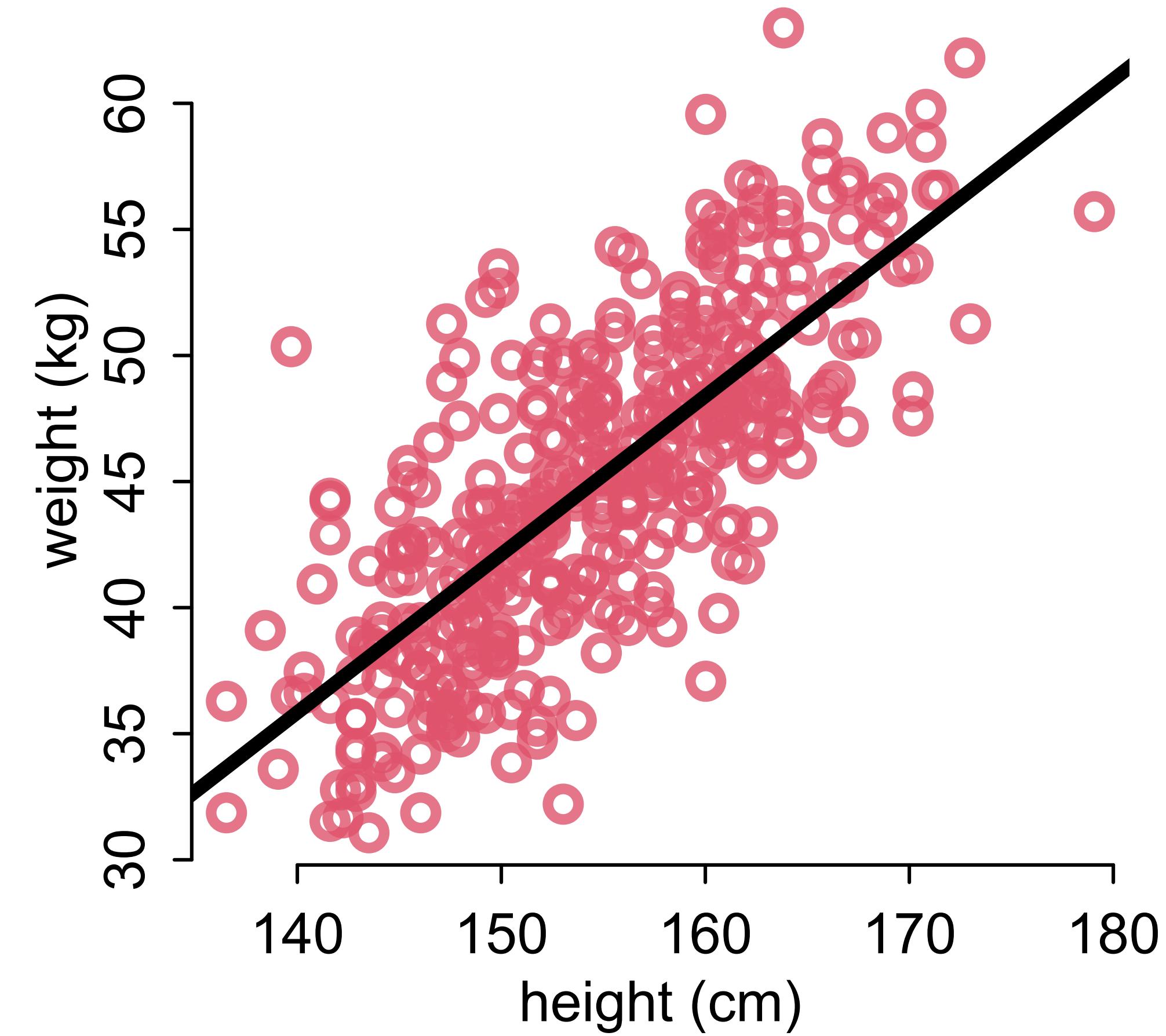
# Posterior predictive distribution

- (1) Plot the sample
- (2) Plot the posterior mean
- (3) Plot uncertainty of the mean
- (4) Plot uncertainty of predictions



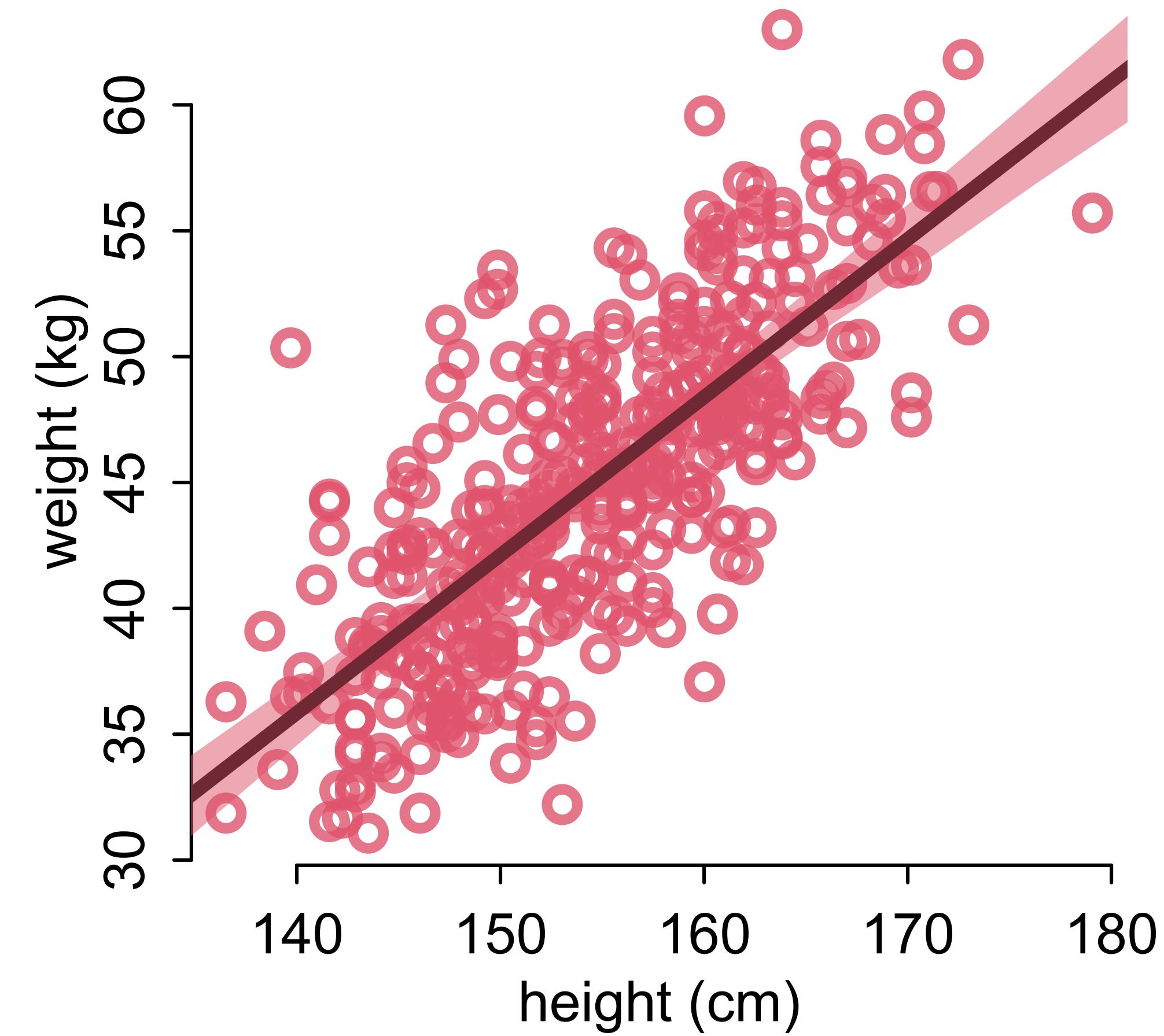
# Posterior predictive distribution

- (1) Plot the sample
- (2) Plot the posterior mean**
- (3) Plot uncertainty of the mean
- (4) Plot uncertainty of predictions



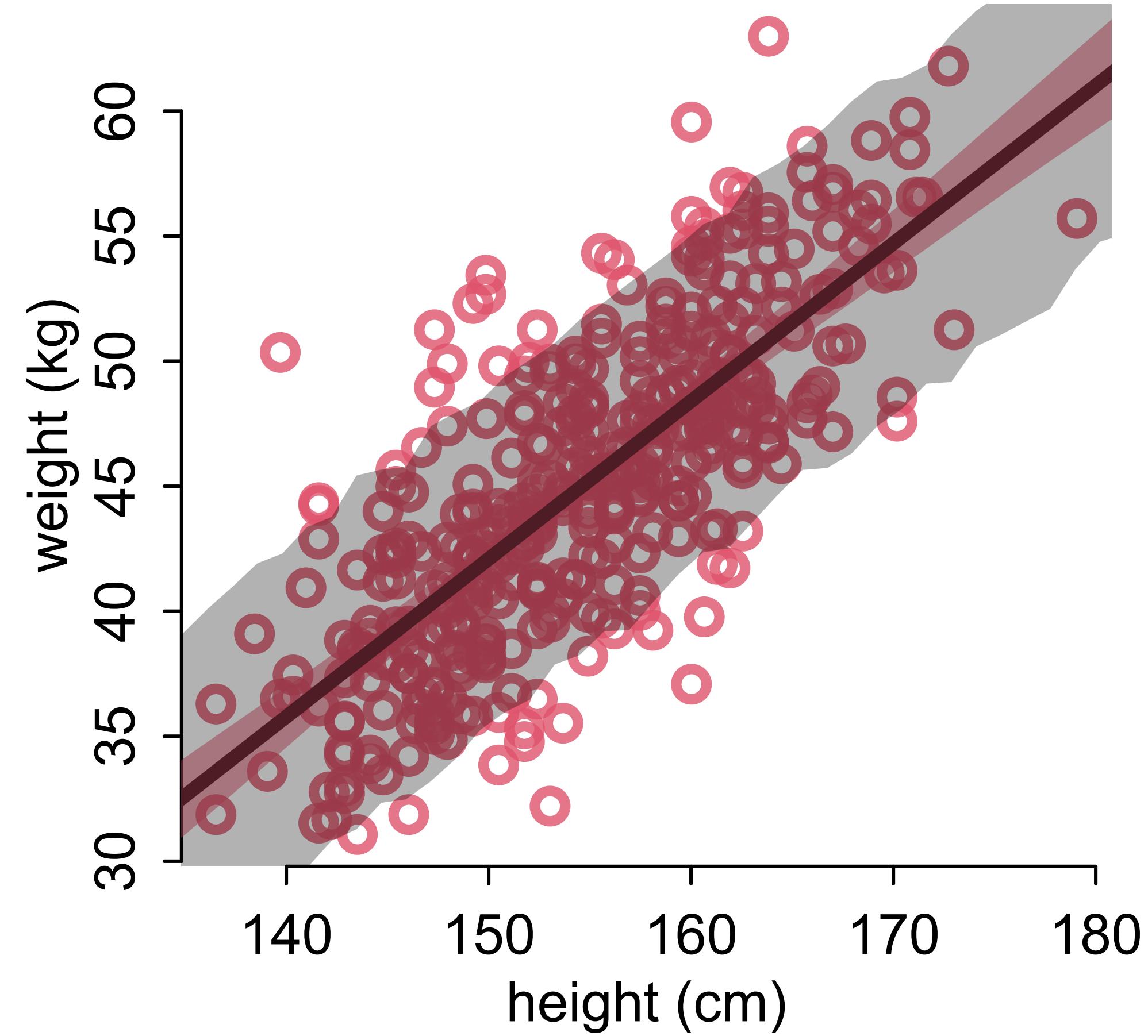
# Posterior predictive distribution

- (1) Plot the sample
- (2) Plot the posterior mean
- (3) Plot uncertainty of the mean**
- (4) Plot uncertainty of predictions



# Posterior predictive distribution

- (1) Plot the sample
- (2) Plot the posterior mean
- (3) Plot uncertainty of the mean
- (4) Plot uncertainty of predictions**

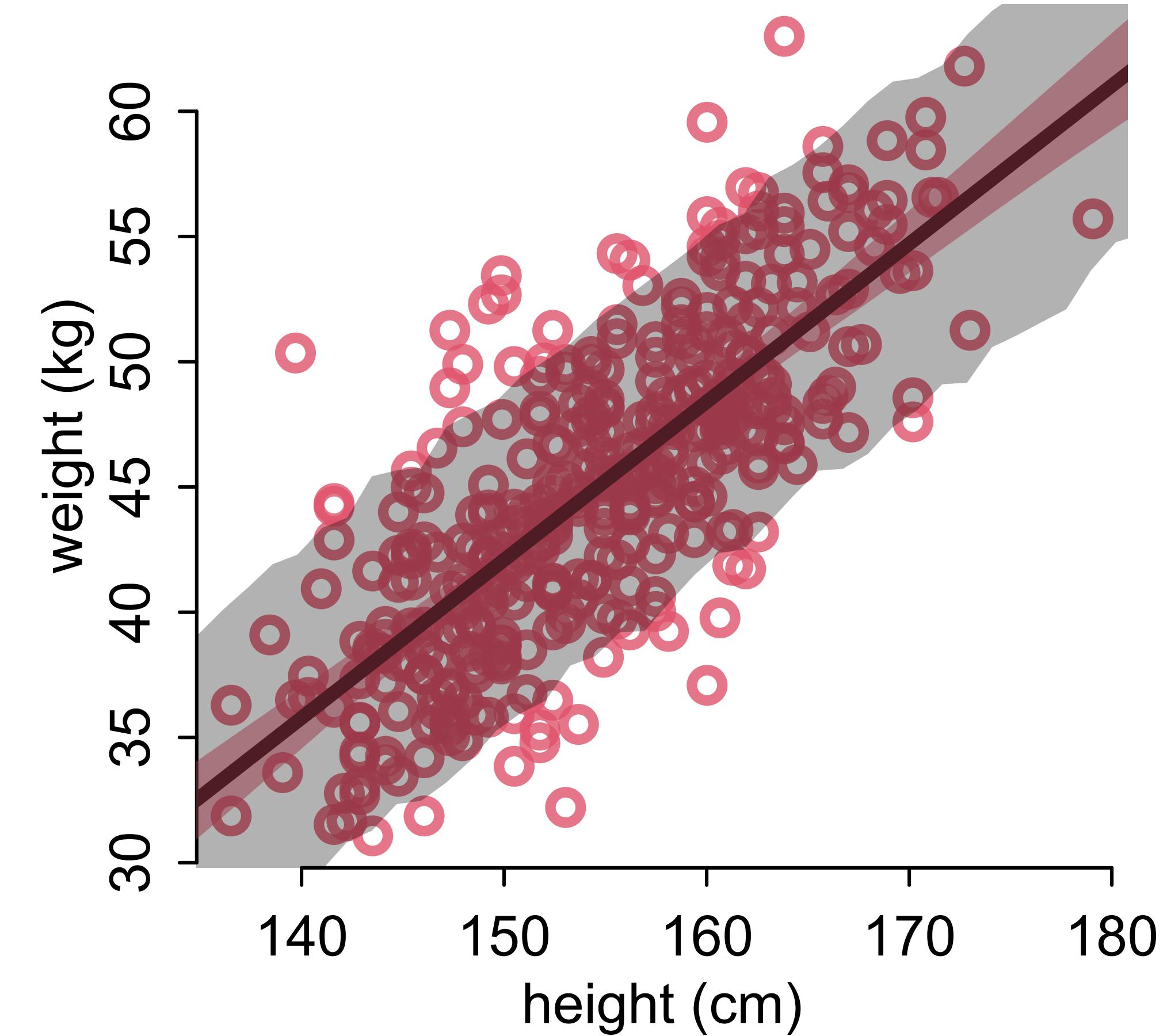


# Posterior predictive distribution

```
# plot sample
col2 <- col.alpha(2,0.8)
plot( d$height , d$weight , col=col2 , lwd=3 ,
      cex=1.2 , xlab="height (cm)" , ylab="weight (kg)" )

# expectation with 99% compatibility interval
xseq <- seq(from=130,to=190,len=50)
mu <- link(m0,data=list(H=xseq,Hbar=mean(d$height)))
lines( xseq , apply(mu,2,mean) , lwd=4 )
shade( apply(mu,2,PI,prob=0.99) , xseq ,
      col=col.alpha(2,0.5) )

# 89% prediction interval
w_sim <- sim(m0,data=list(H=xseq,Hbar=mean(d$height)))
shade( apply(w_sim,2,PI,prob=0.89) , xseq ,
      col=col.alpha(1,0.3) )
```



# A Question

- How to evaluate expectation of a function  $h(x)$
- Given that  $X$  has p.d.f.  $f(x)$  ?

# Answer

- How to evaluate expectation of a function  $h(x)$
- Given that  $X$  has p.d.f.  $p(x)$  ?
- Sample from  $p(x)$  a random sample of size  $n$
- $(x_1, x_2, \dots, x_n)$ 
$$E_p[h(x)] \equiv \int_X h(x)p(x)dx \approx \frac{1}{n} \sum_{i=1}^n h(x_i)$$
- and Evaluate (approximately):

**So, how do we sample?**

# Outline

- Inverse CDF Sampling
- Accept-Reject Sampling
- MCMC
- Metropolis-Hastings Sampling
- Gibbs Sampling

# Inverse CDF

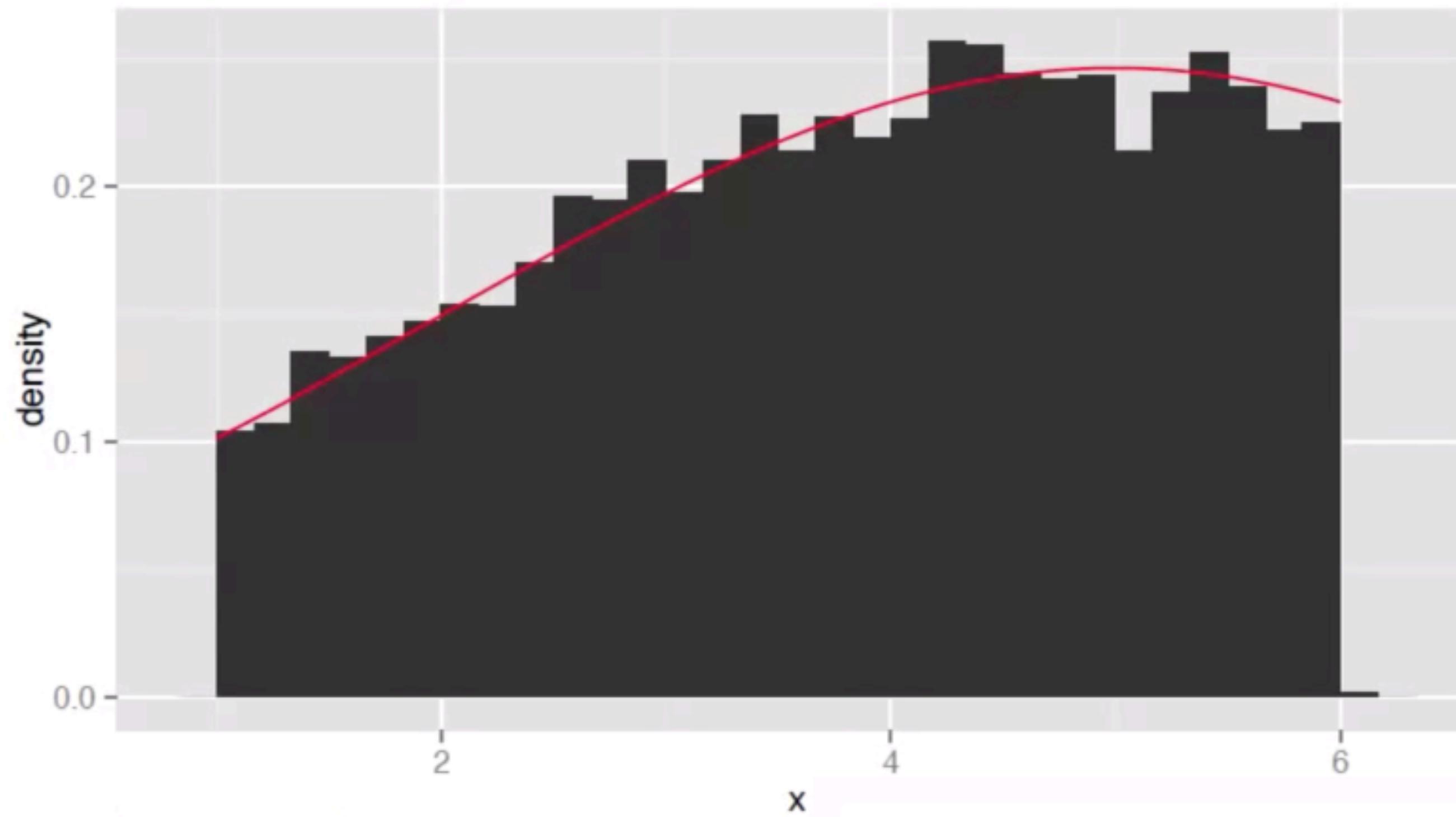
- Let us
  - know the p.d.f.:  $f(x)$
  - know the CDF:  $F(x)$
- If we can easily find the Inverse CDF:  $F^{-1}(x)$
- Then we use the Inverse CDF for sampling

**Hint:**  $X = F^{-1}(Uniform(0,1))$

# Example

- Sample from a truncated Normal Distr.  $X \sim N(\mu, \sigma^2)I(a < X < b)$
  - We know Inverse CDF  $\Phi^{-1}$ 
$$p_a = \Phi((a - \mu)/\sigma) \quad p_b = \Phi((b - \mu)/\sigma)$$
  - Calculate endpoints  $u \sim U = \text{Uniform}(p_a, p_b)$ 
$$X = \sigma\Phi^{-1}(u) + \mu$$
  - Sample from
- $$X \sim N(5,9)I(1 < X < 6)$$
- Derive a sample you need

# Visualization



# Accept-Reject Sampling

# Accept-Reject Sampling

- If you cannot derive Inverse CDF, or calculation of it is expensive
- But you still have the p.d.f., or at least you know  $f(x)$  s.t.  $p(x) = \frac{f(x)}{C}$
- You can use Accept-Reject Sampling

# Accept-Reject: Idea

- Propose a sample  $x(j)$  and decide (using  $f(x)$ ) whether to:
  - Accept or Reject  $x(j)$
  - If Rejected, propose another sample
- Thus, you need some proposal distribution  $q(x)$ ... Why?

# Accept-Reject: Method

- Sample from  $X^* \sim q(x)$ ; and  $U \sim Uniform(0,1)$
- Accept  $X^*$  if  $U \leq \frac{f(X^*)}{Mq(X^*)} = \frac{1}{M} \frac{f(X^*)}{q(X^*)}$
- where  $M$  is selected s.t.  $\forall x : Mq(x) \geq f(x)$
- Otherwise, Reject  $X^*$
- Optimal  $M = \sup_x [f(x)/q(x)]$

# Accept-Reject: Example

- We need to sample from  $X \sim N(0,1)I(X > 5)$

$$f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}(1 - \Phi(5))}$$

- We use exponential distribution as a proposal

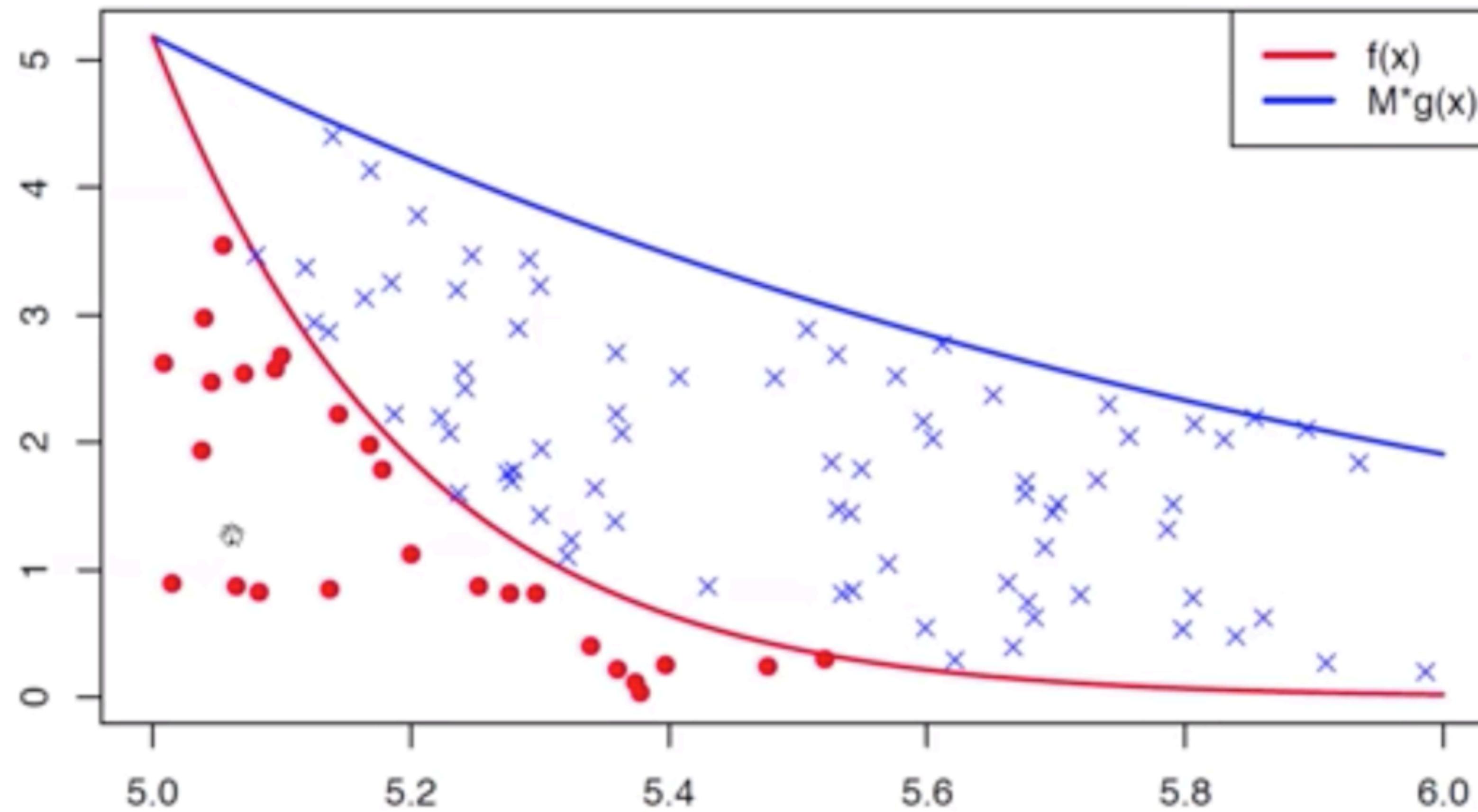
$$Y \sim Exp(1) \quad X^* = Y + 5 \quad q(x) = exp(- (x - 5))$$

- Calculate M

$$M = \sup_x [f(x)/q(x)] = \frac{1}{\sqrt{2\pi}(1 - \Phi(5))} \sup_x \frac{e^{-x^2/2}}{e^{-(x-5)}} \approx 5.19$$

- Apply Accept-Reject Method

# Visualization



# Discussion

- What about efficiency?
- what about independence of samples?