

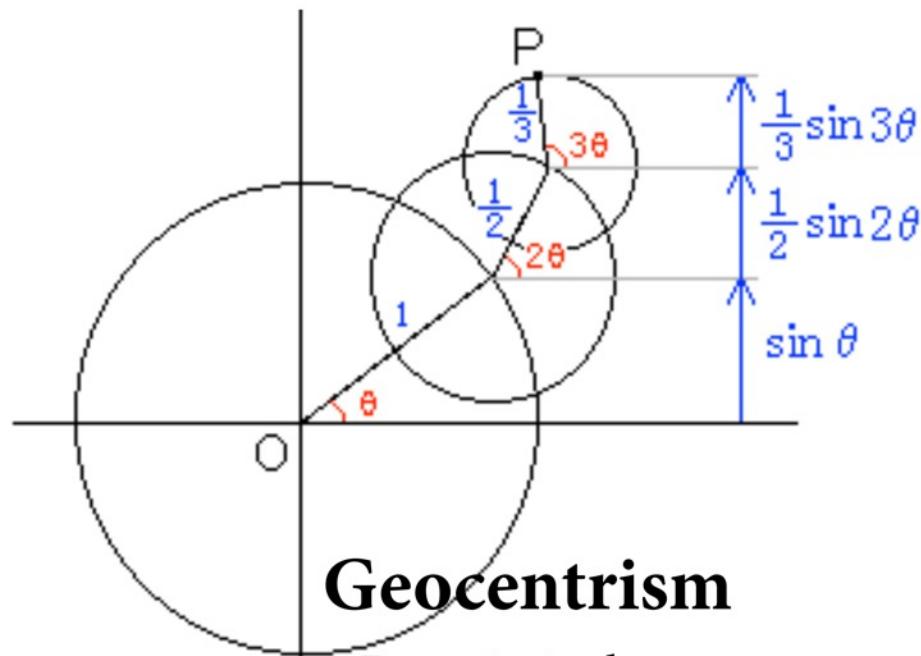
Methods 4 - 3

Chris Mathys



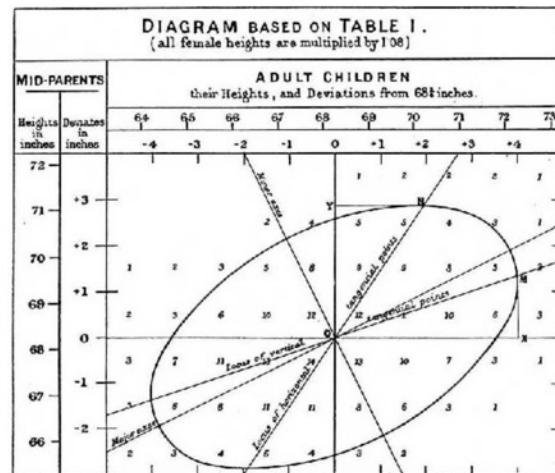
BSc Programme in Cognitive Science

Spring 2022



Geocentrism

- Descriptively accurate
- Mechanistically wrong
- General method of approximation
- Known to be wrong



Linear Regression

- Descriptively accurate
- Mechanistically wrong
- General method of approximation
- Taken too seriously

Linear Regression

Simple statistical golems

Model of **mean** and **variance** of variable

Mean as **weighted sum** of other variables

Many special cases: ANOVA, ANCOVA,
t-test, MANOVA

Can also be generative models



From *Breath of Bones: A Tale of the Golem*

Why Normal?

Two arguments

- (1) Generative: Summed fluctuations tend towards normal distribution
- (2) Statistical: For estimating mean and variance, normal distribution is least informative distribution (maxent)

Variable does not have to be normally distributed for normal model to be useful

Making Normal Models

Goals:

- (1) Language for representing models
- (2) How to calculate bigger posterior distributions
- (3) Constructing & understanding linear models



Language for modeling

Revisit globe tossing model:

$$W \sim \text{Binomial}(N, p)$$

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

Language for modeling

Revisit globe tossing model:

outcome

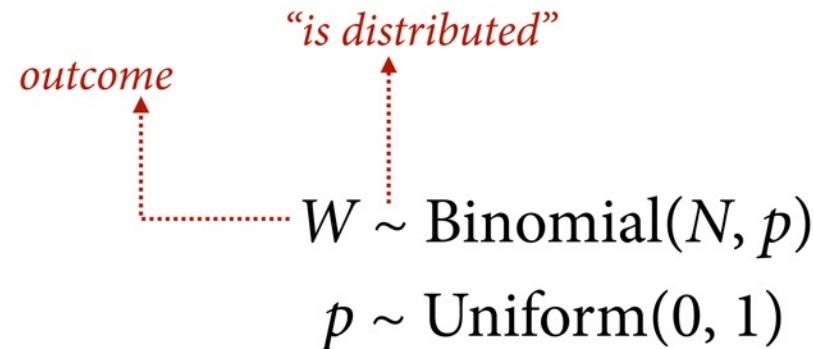


$$W \sim \text{Binomial}(N, p)$$

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

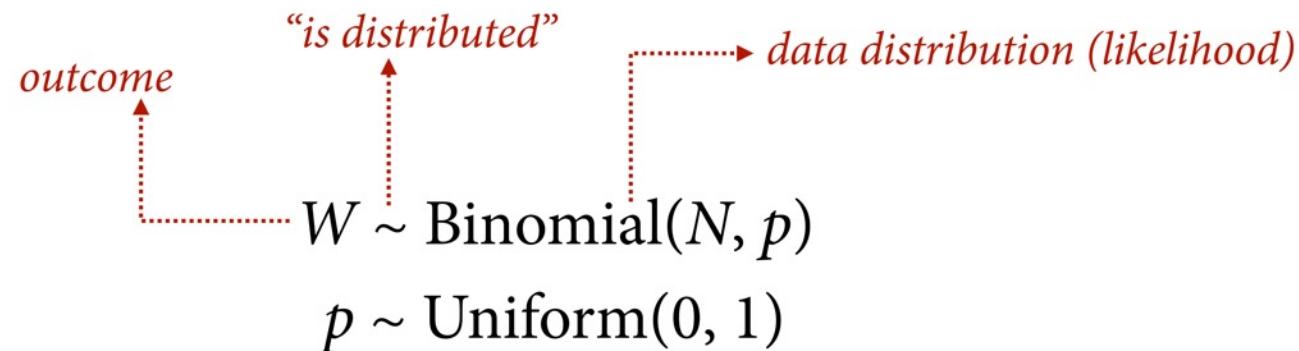
Language for modeling

Revisit globe tossing model:



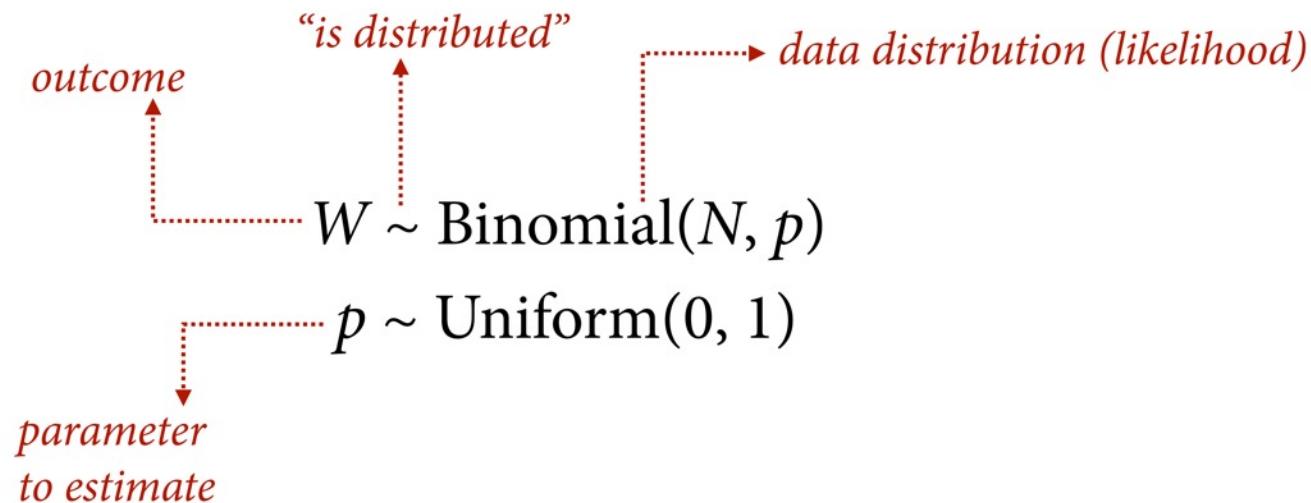
Language for modeling

Revisit globe tossing model:



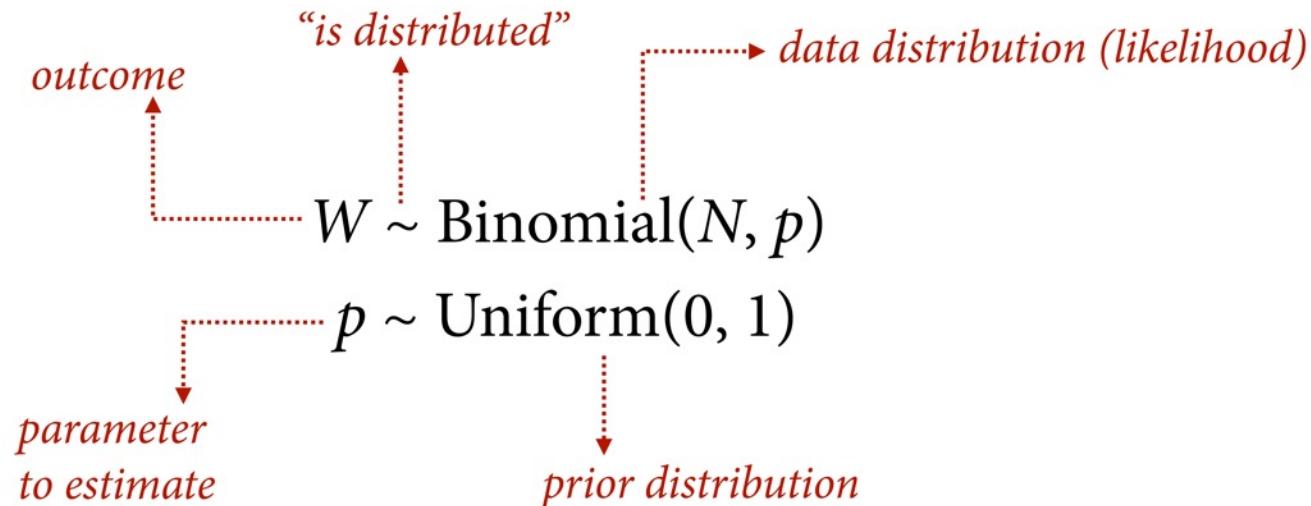
Language for modeling

Revisit globe tossing model:



Language for modeling

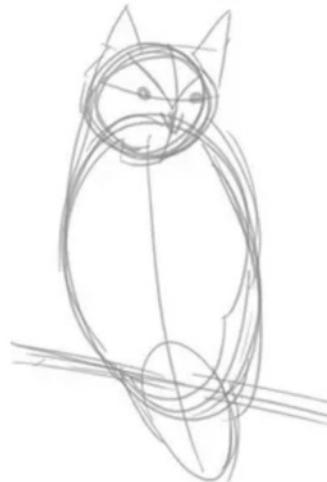
Revisit globe tossing model:



Now make it compute — arrange as probability statements

$$W \sim \text{Binomial}(N, p)$$

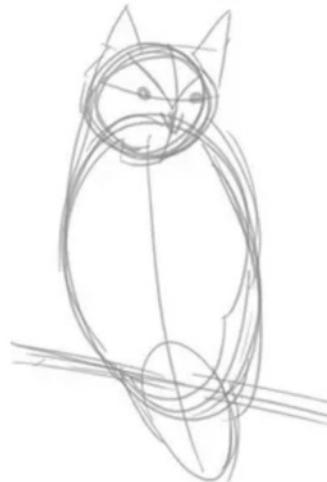
$$p \sim \text{Uniform}(0, 1)$$



Now make it compute — arrange as probability statements

$$\Pr(W|N, p) = \text{Binomial}(W|N, p)$$

$$\Pr(p) = \text{Uniform}(p|0, 1)$$



Now make it compute — arrange as probability statements

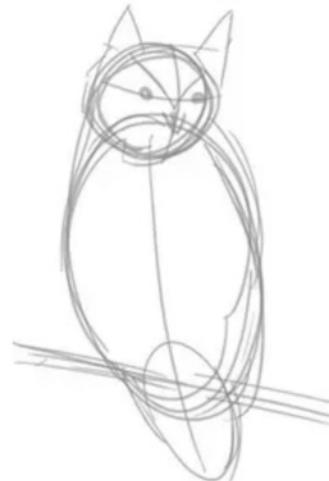
$$\Pr(W|N, p) = \text{Binomial}(W|N, p)$$

$$\Pr(p) = \text{Uniform}(p|0, 1)$$

Posterior distribution

$$\Pr(p|W, N) \propto \text{Binomial}(W|N, p) \text{Uniform}(p|0,1)$$

↓
“*proportional to*”



Now make it compute — using code

$$\Pr(p|W, N) \propto \text{Binomial}(W|N, p) \text{Uniform}(p|0,1)$$

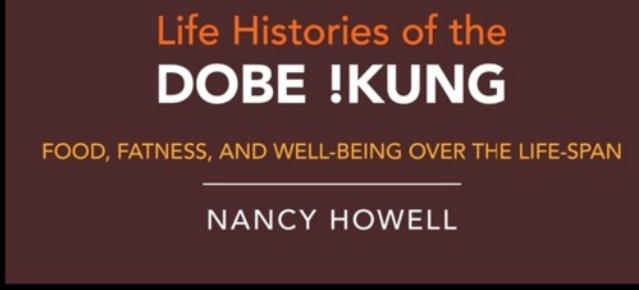
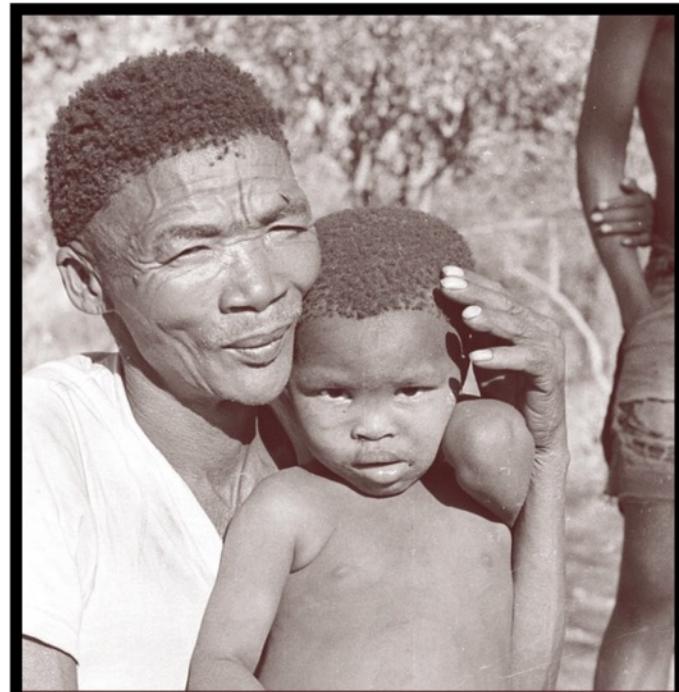
```
W <- 6
N <- 9
p <- seq(from=0,to=1,len=1000)
PrW <- dbinom(W,N,p)
Prp <- dunif(p,0,1)
posterior <- PrW * Prp
```



Linear Regression

Drawing the Owl

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

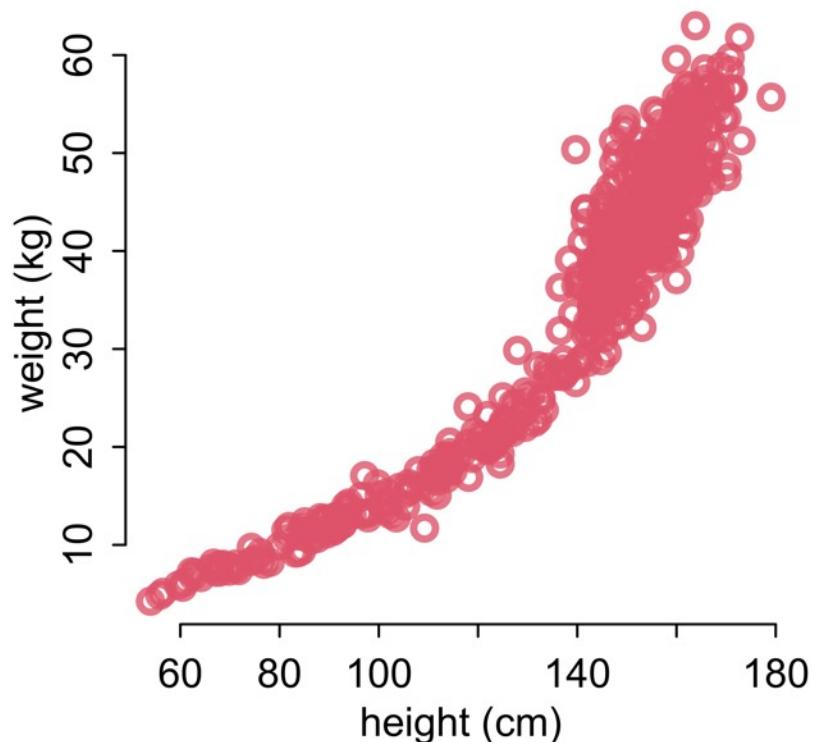


Linear Regression

Drawing the Owl

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

```
library(rethinking)  
data(Howell1)
```



why analyze in last step? How do you know your model is good? By simulating data from the model.

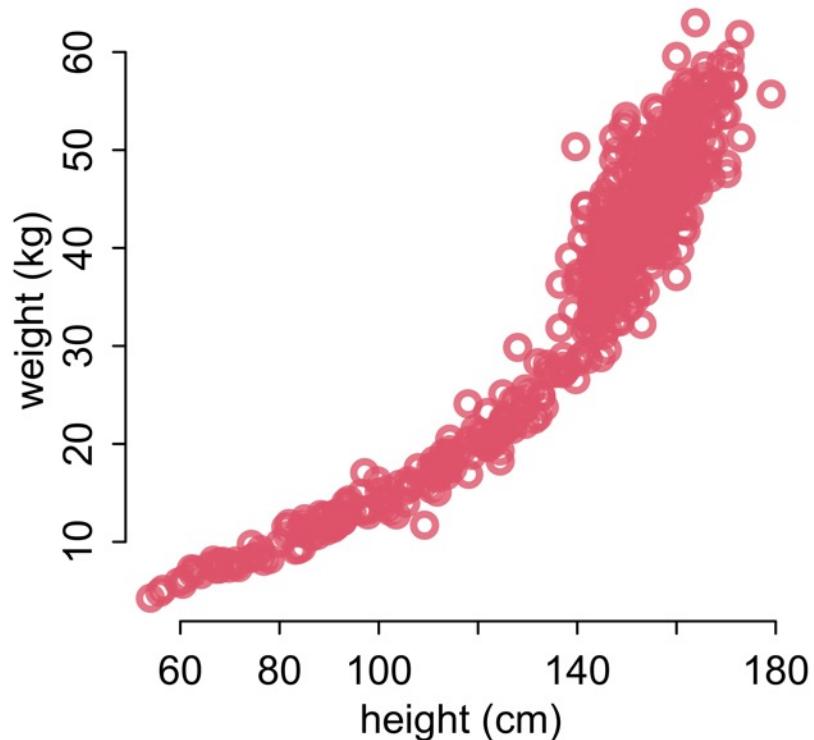
Linear Regression

Drawing the Owl

(1) Question/goal/estimand

Describe association between
weight and **height**

```
library(rethinking)  
data(Howell1)
```



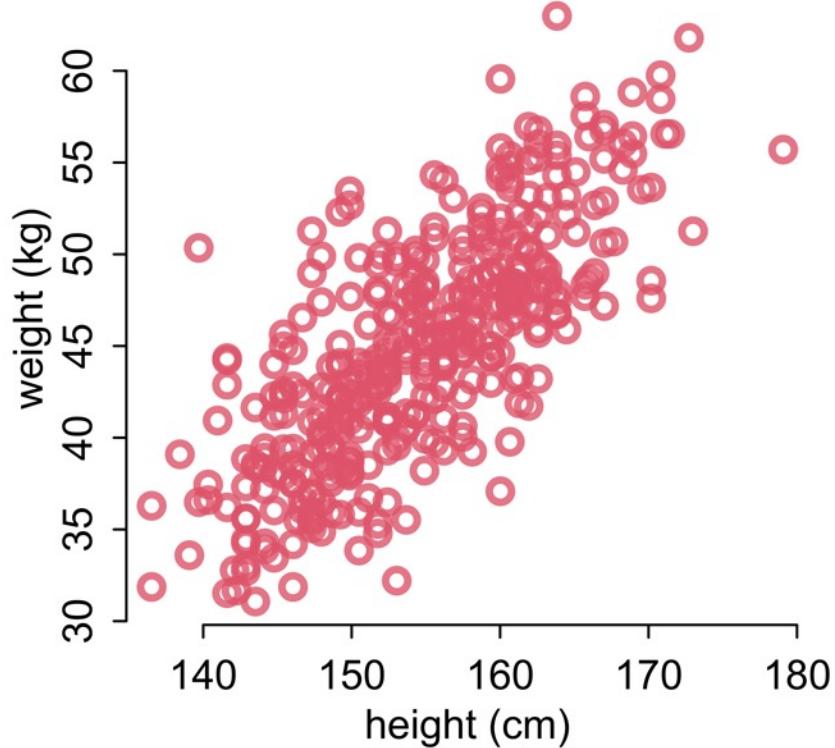
Linear Regression

Drawing the Owl

(1) Question/goal/estimand

Describe association between
ADULT **weight** and **height**

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



Linear Regression

(2) Scientific model

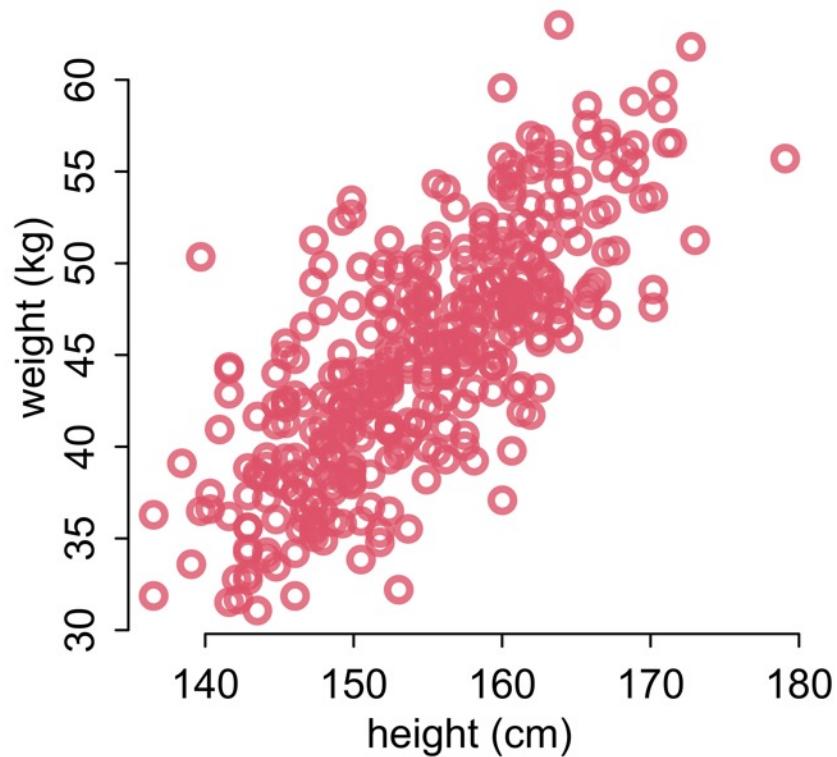
How does **height** influence
weight?

$$H \longrightarrow W$$

$$W = f(H)$$

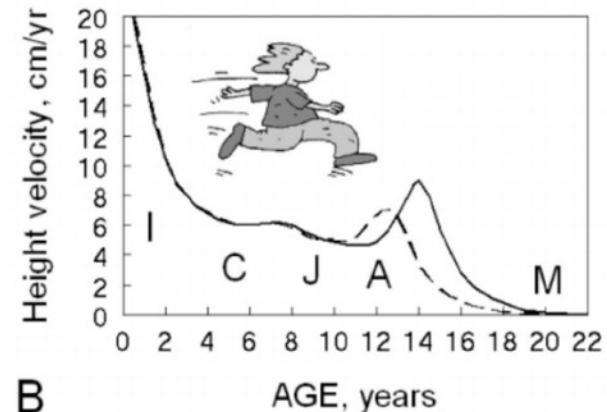
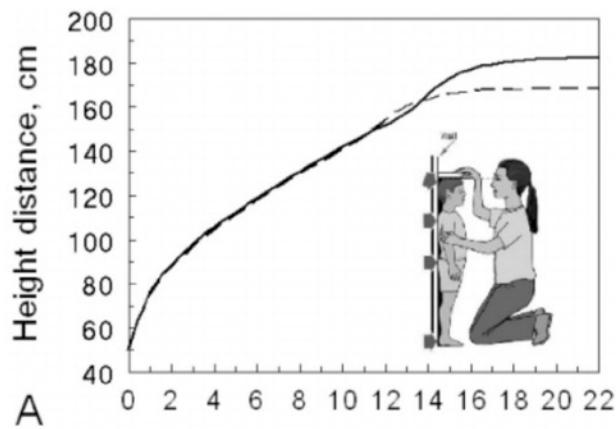
“Weight is some function of height”

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



Generative models

Options

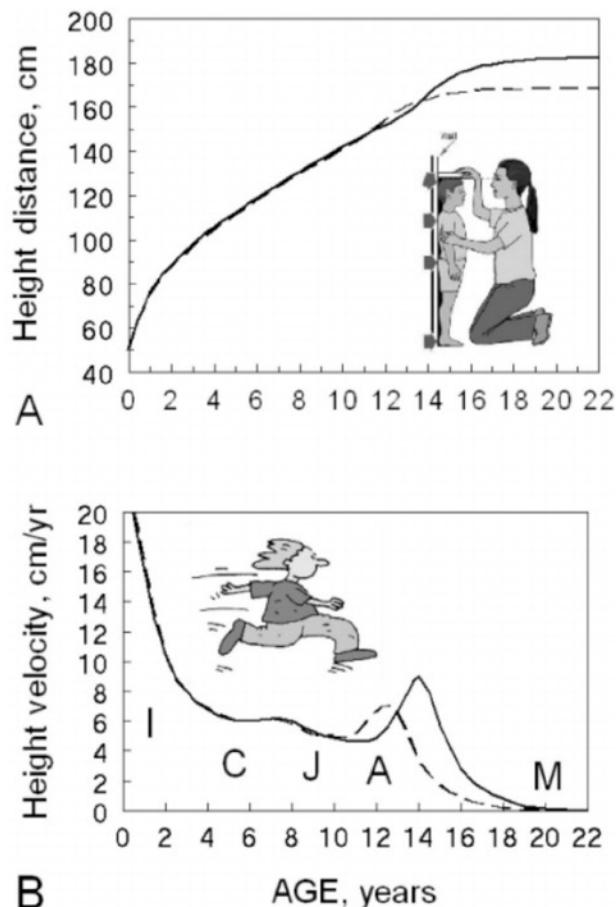


Generative models

Options

(1) **Dynamic**: Incremental growth of organism; both mass and height (length) derive from growth pattern; Gaussian variation result of summed fluctuations

look at how change in one thing induce change in another thing.



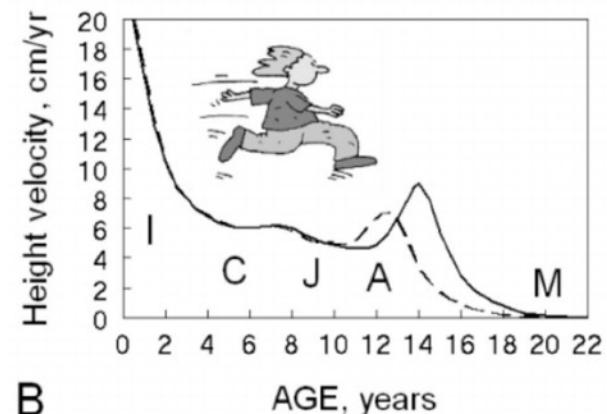
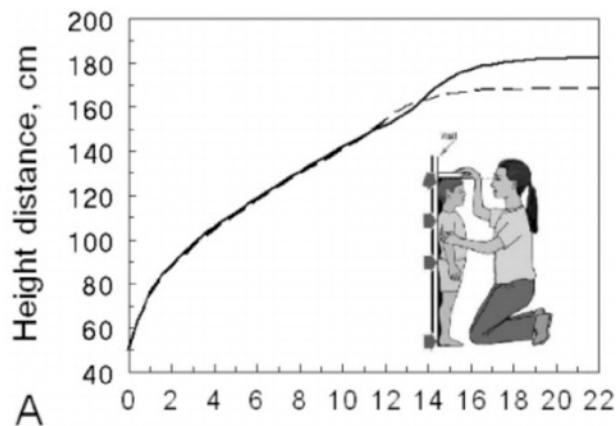
Generative models

Options

(1) **Dynamic**: Incremental growth of organism; both mass and height (length) derive from growth pattern; Gaussian variation result of summed fluctuations

(2) **Static**: Changes in height result in changes in weight, but no mechanism; Gaussian variation result of growth history

Static: more simple way to go.

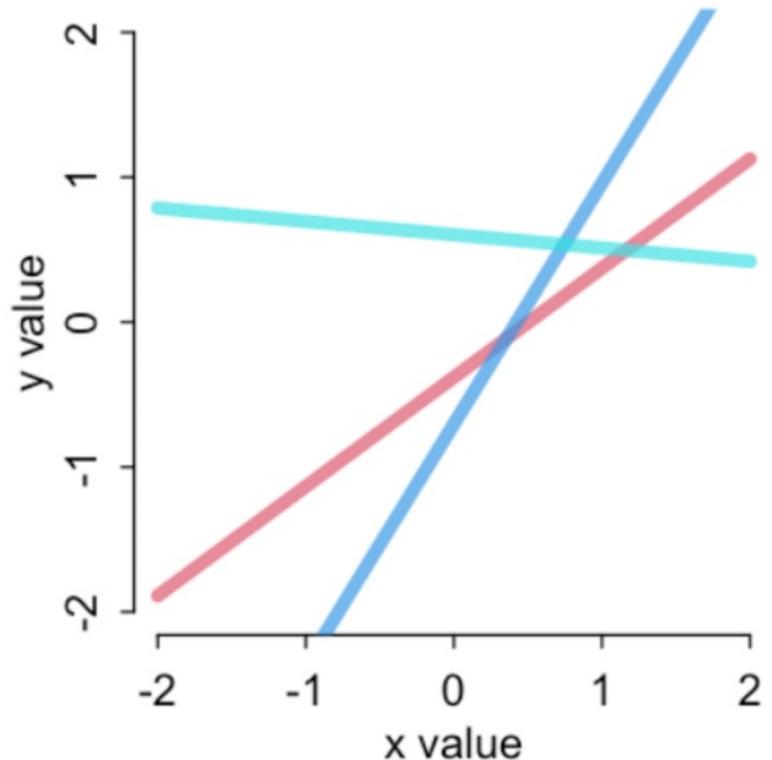


Anatomy of a linear model

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

index ↓ *slope*
 ↓
intercept

Text



Anatomy of a linear model

$$y_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta x_i$$

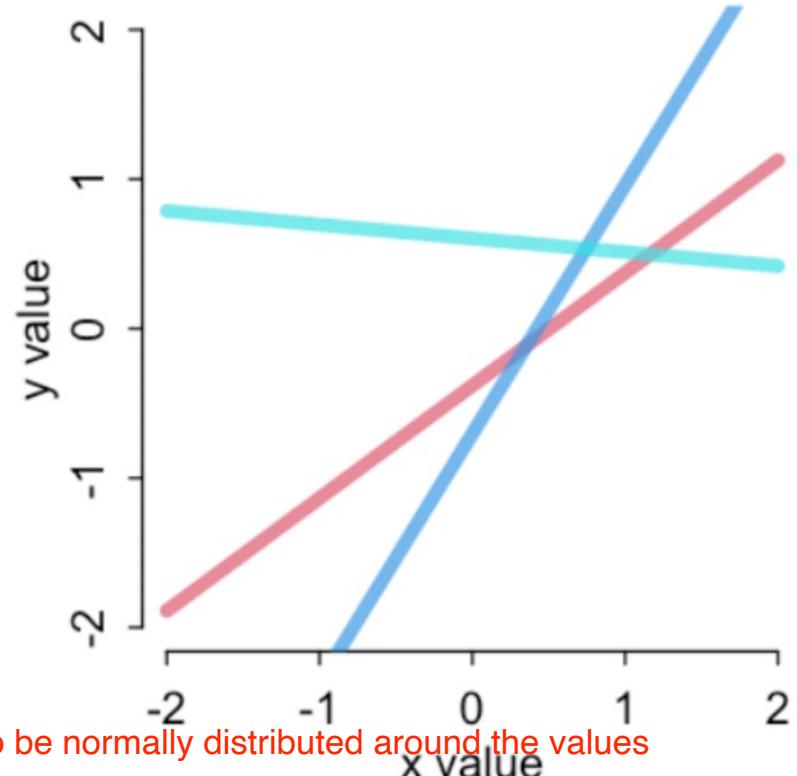
expectation *standard deviation*

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

Assumption: the observations are going to be normally distributed around the values of a linear predictor.

Model entry: one that defines distribution, and one that defines quantities and assign them to mu.

mu is NOT a parameter, but a deterministic function of beta and alpha.



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)
```

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)
```

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)
```

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)
```

Generative model: $H \rightarrow W$

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

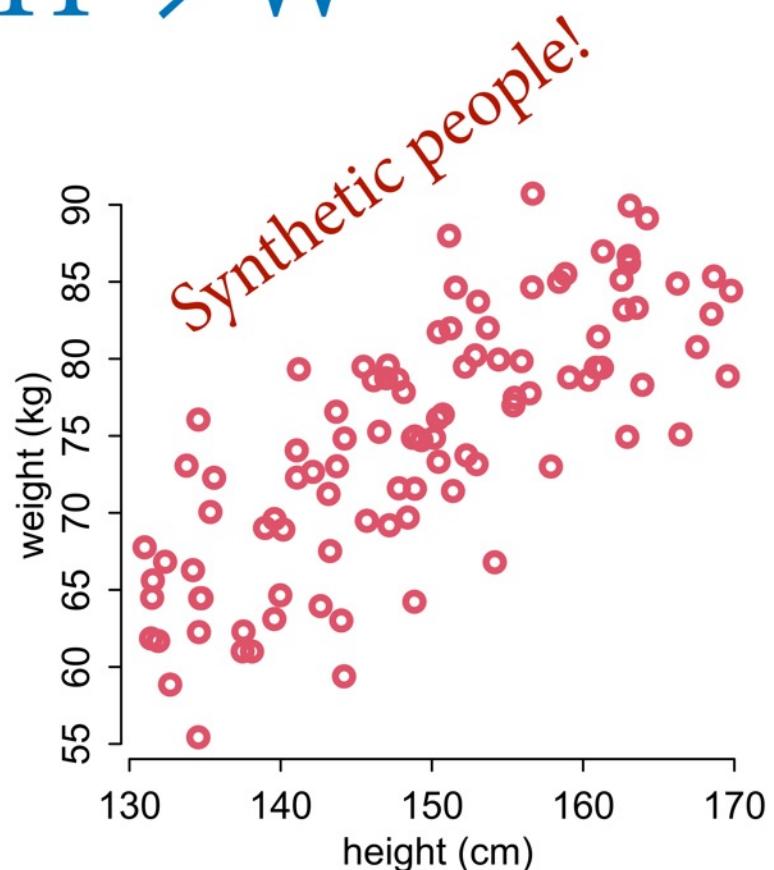
$$\mu_i = \alpha + \beta H_i$$

```
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)
```

Text

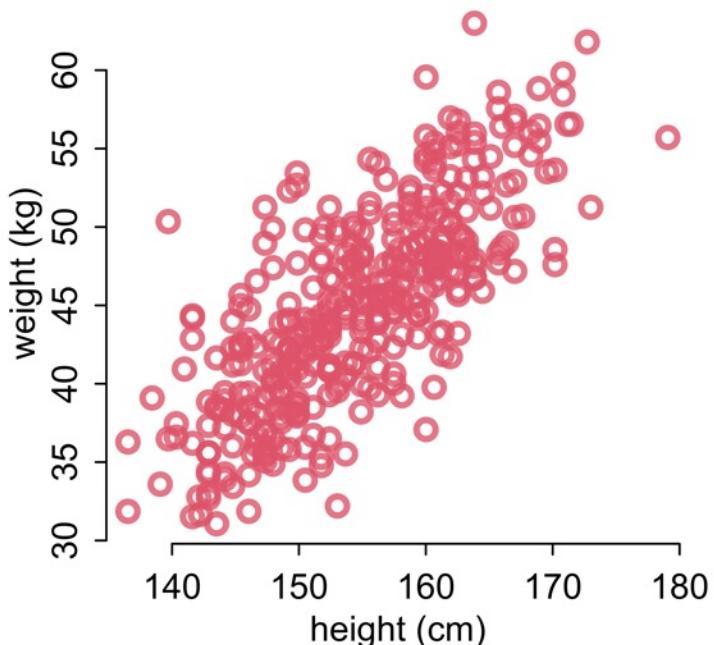


Linear Regression

Drawing the Owl

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

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



Anatomy of a linear model

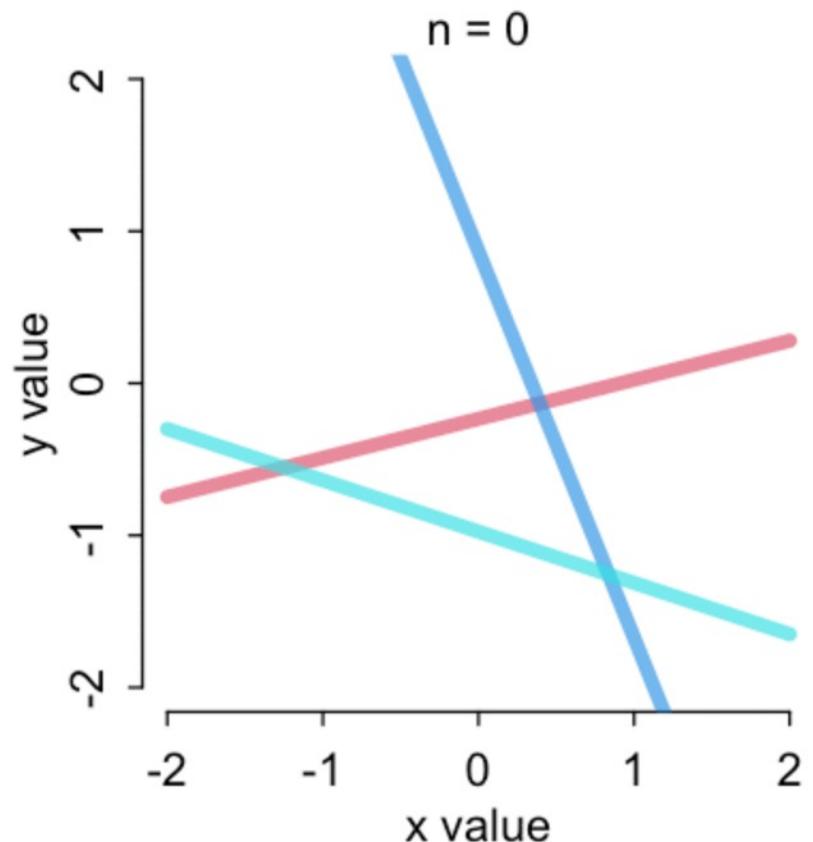
$$y_i \sim \text{Normal}(\mu_i, \sigma)$$

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

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

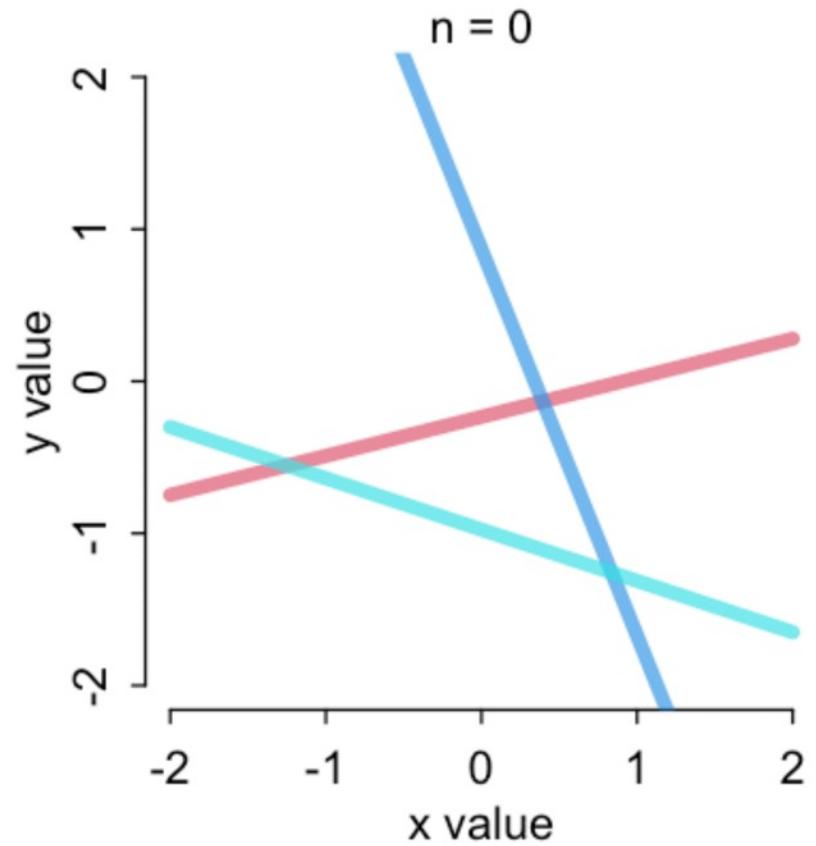
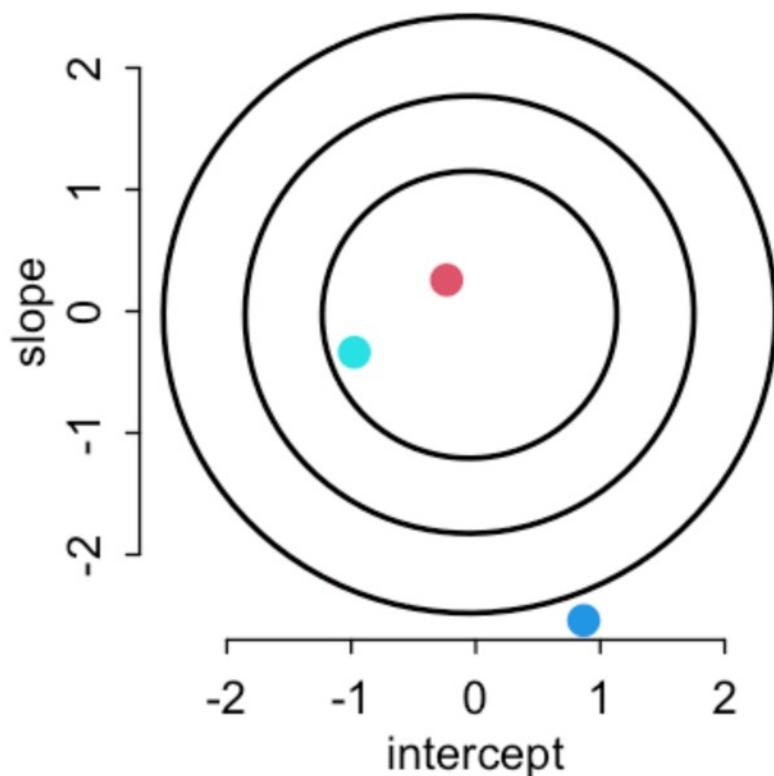
$$\beta \sim \text{Normal}(0, 1)$$

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



These are wide priors; pretty far off an actual distribution.

Sampling the prior distribution

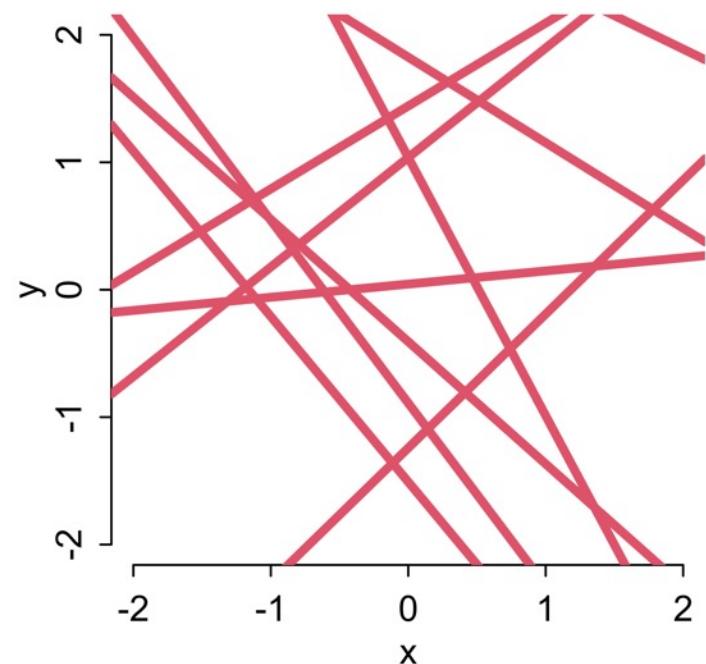


Sampling the prior distribution

```
n_samples <- 10

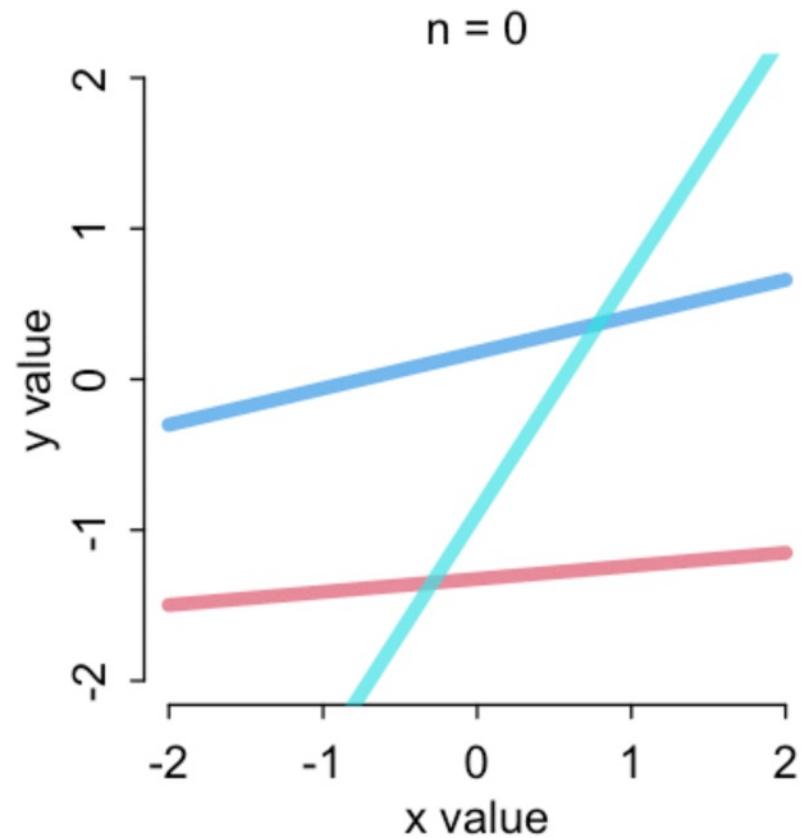
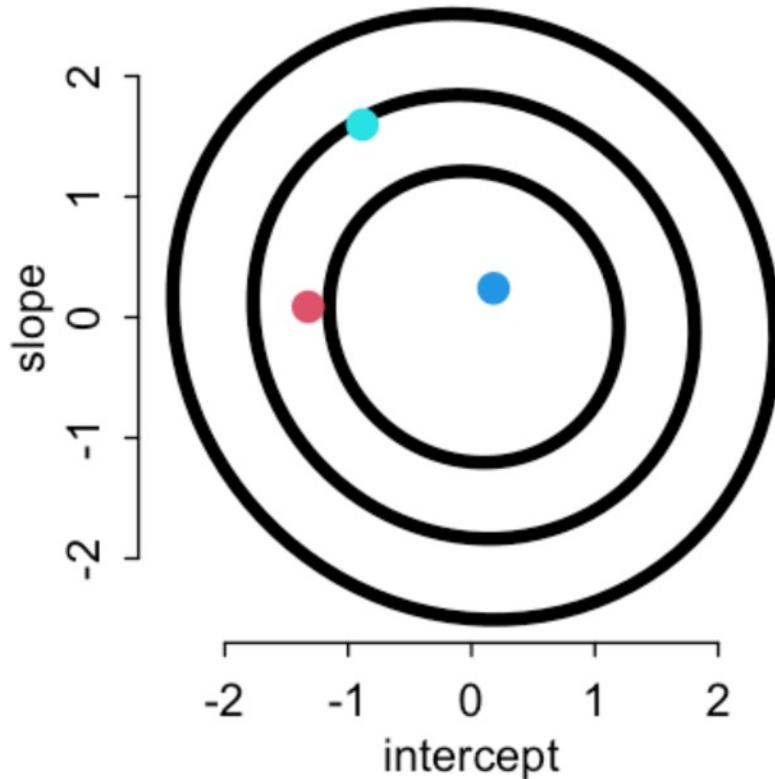
alpha <- rnorm(n_samples,0,1)
beta <- rnorm(n_samples,0,1)

plot(NULL,xlim=c(-2,2),ylim=c(-2,2),
     xlab="x",ylab="y")
for ( i in 1:n_samples )
  abline(alpha[i],beta[i],lwd=4,col=2)
```



the priors are not constrained enough at this point

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 μ when
 $H_i - \bar{H} = 0$*

mean value of H_i

standardizing predictors makes model more interpretable

Statistical model for $H \rightarrow W$

Now what are scientifically reasonable priors?

α : average adult weight

β : kilograms per centimeter

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

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

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

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

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

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

Now we have informed priors in red.

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)",ylab="weight (kg)")
for ( i in 1:n )
  lines( Hseq , alpha[i] + beta[i]*(Hseq-Hbar) ,
  lwd=3 , col=2 )
```

Text

Sampled regression lines

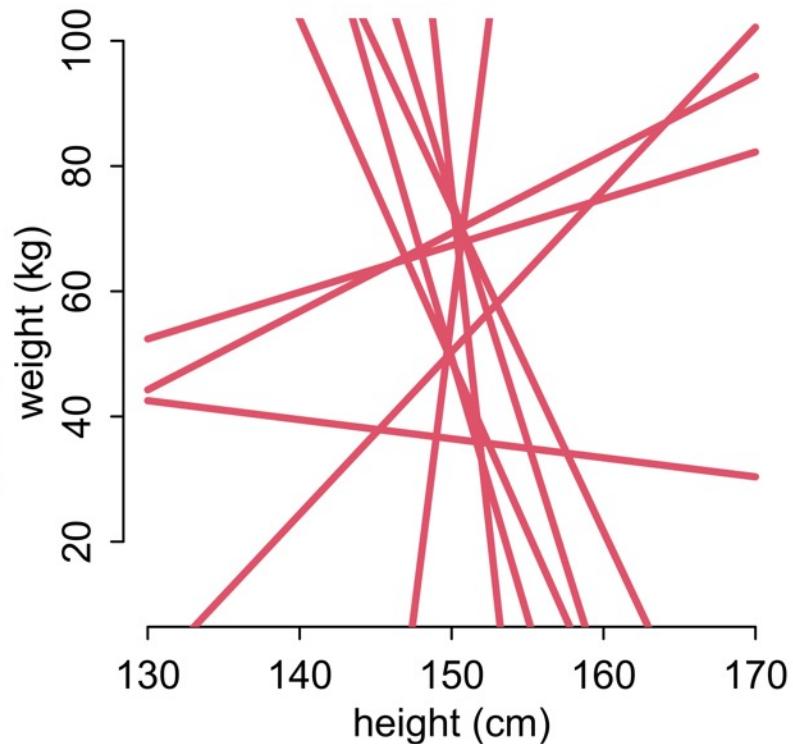
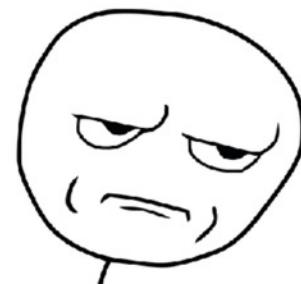
```
n <- 10
alpha <- rnorm(n,60,10)
beta <- rnorm(n,0,10)

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 )
```

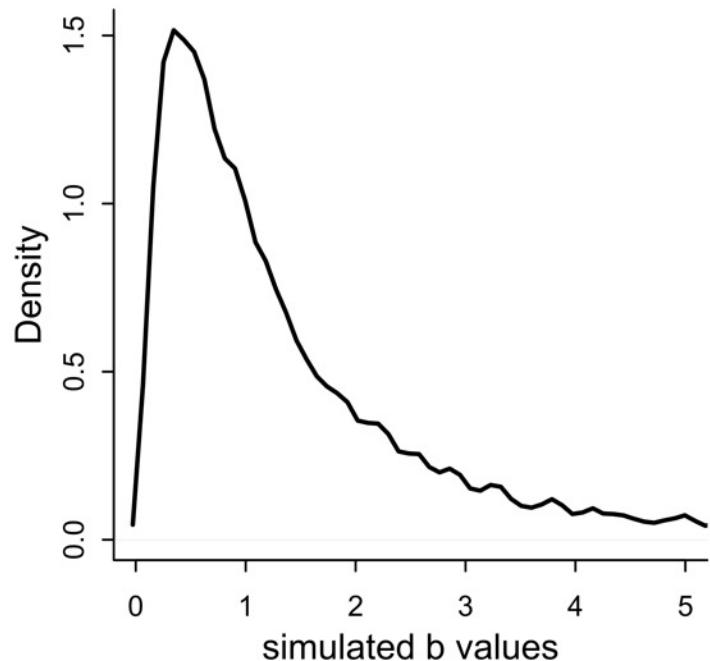
Sampled regression lines

```
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 )
```



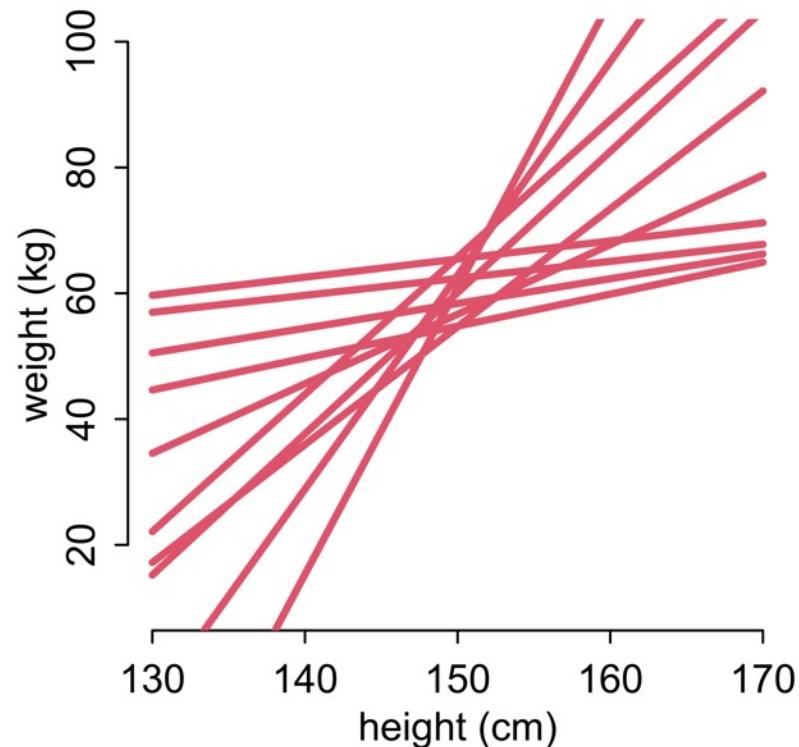
Statistical model for $H \rightarrow W$


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

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 )
```



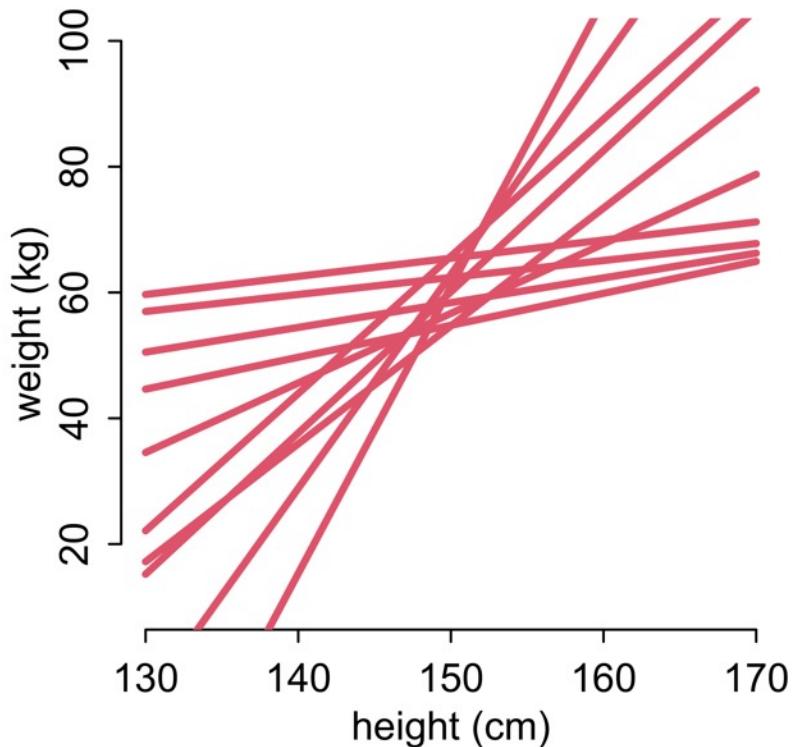
Sermon on Priors

There are no correct priors, only scientifically justifiable priors

Justify with information outside the data — **like rest of model**

Priors not so important in simple linear models

But need to practice now: simulate, understand, expand



Fitting the model

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

$\Pr(W_i | \mu_i, \sigma)$

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

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

$\Pr(\alpha)$

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

$\Pr(\beta)$

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

$\Pr(\sigma)$

Posterior is $\Pr(\alpha, \beta, \sigma | W, H)$

Fitting the model

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

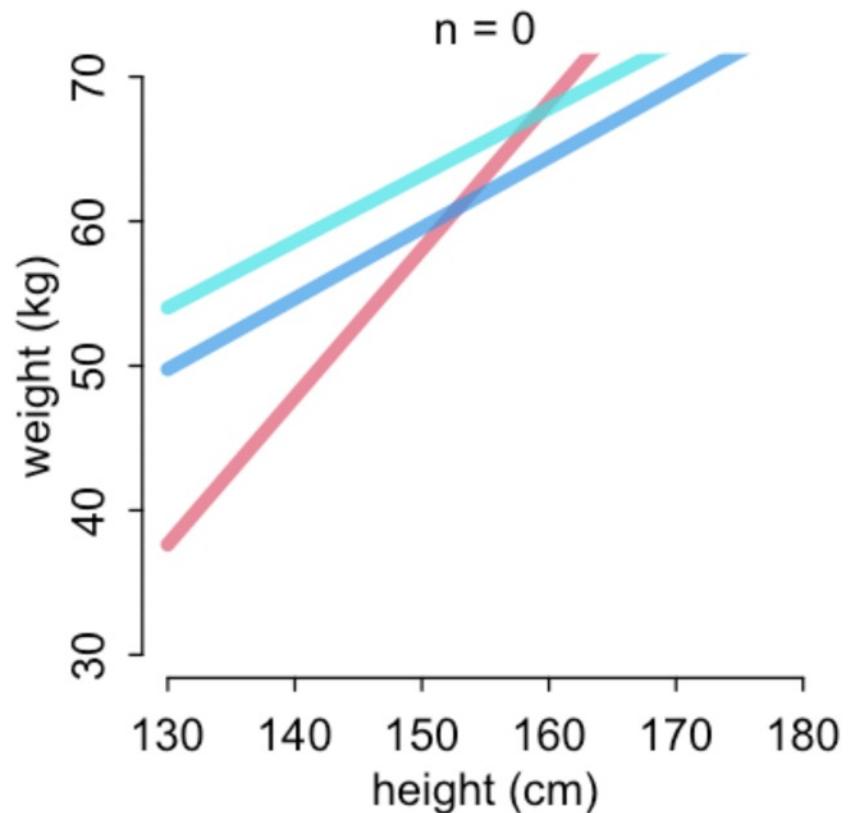
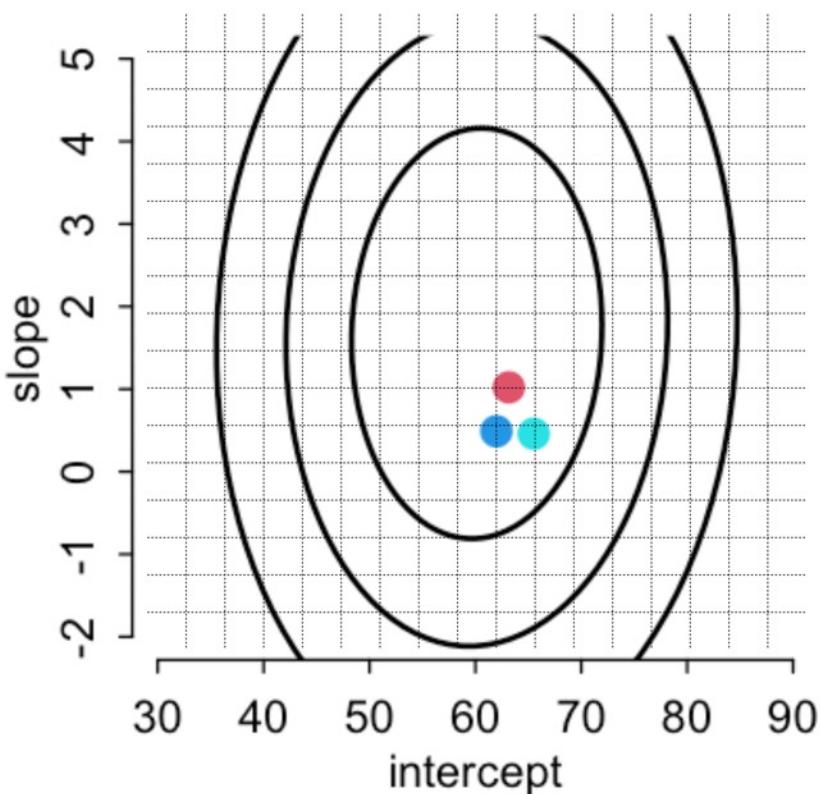
Grid approximation expensive:

100 values of each parameter => 1 million calculations

See page 85 in book for coded example

grid approximation: too expensive as soon as things get complicated.

30 adults from Howell1



Approximate posterior

Many posterior distributions are approximately Gaussian

Instead of grid approximation, Gaussian approximation

Sometimes called **quadratic** or **Laplace approximation**

See page 41 in book for more detail

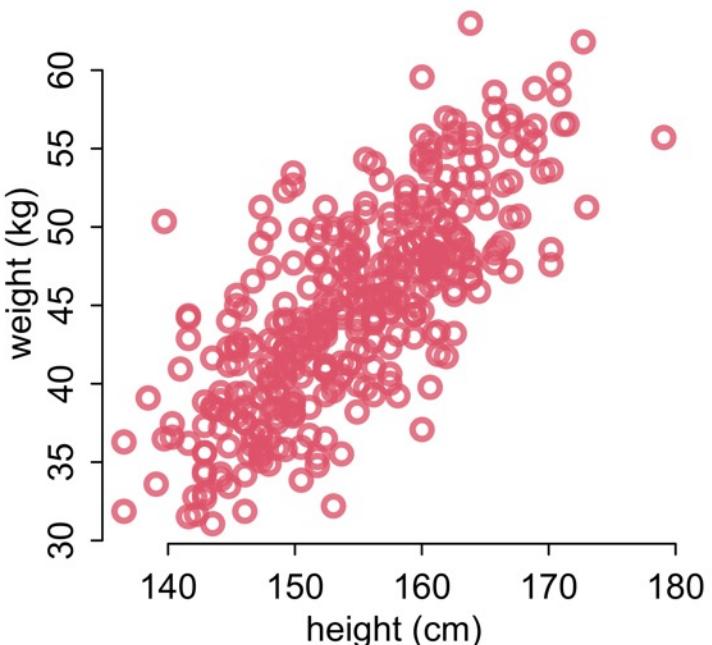


Linear Regression

Drawing the Owl

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

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



Simulation-Based Validation

Bare minimum: Test statistical model
with simulated observations from
scientific model

Golem might be broken

Even working golems might not deliver
what you hoped

Strong test: **Simulation-Based Calibration**



Fahrvergnügen

Model formula

```
W ~ dnorm(mu,sigma),  
mu <- a + b*(H-Hbar),  
a ~ dnorm(60,10),  
b ~ dlnorm(0,1),  
sigma ~ dunif(0,10)
```

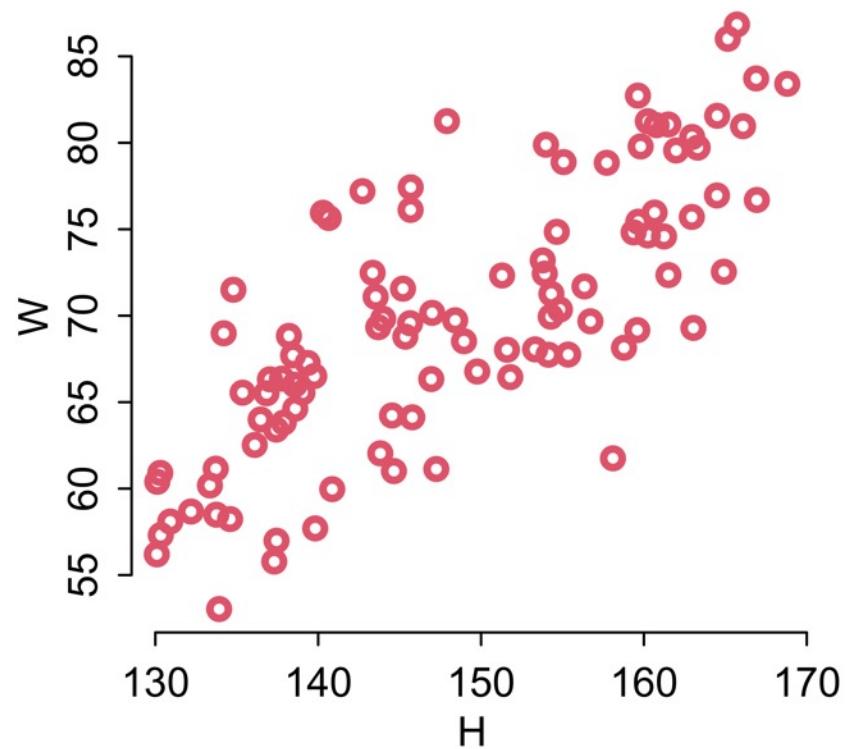
$$W_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta(H_i - \bar{H})$$
$$\alpha \sim \text{Normal}(60, 10)$$
$$\beta \sim \text{LogNormal}(0, 1)$$
$$\sigma \sim \text{Uniform}(0, 10)$$

First validate with simulation

```
alpha <- 70
beta <- 0.5
sigma <- 5
n_individuals <- 100
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 )
```

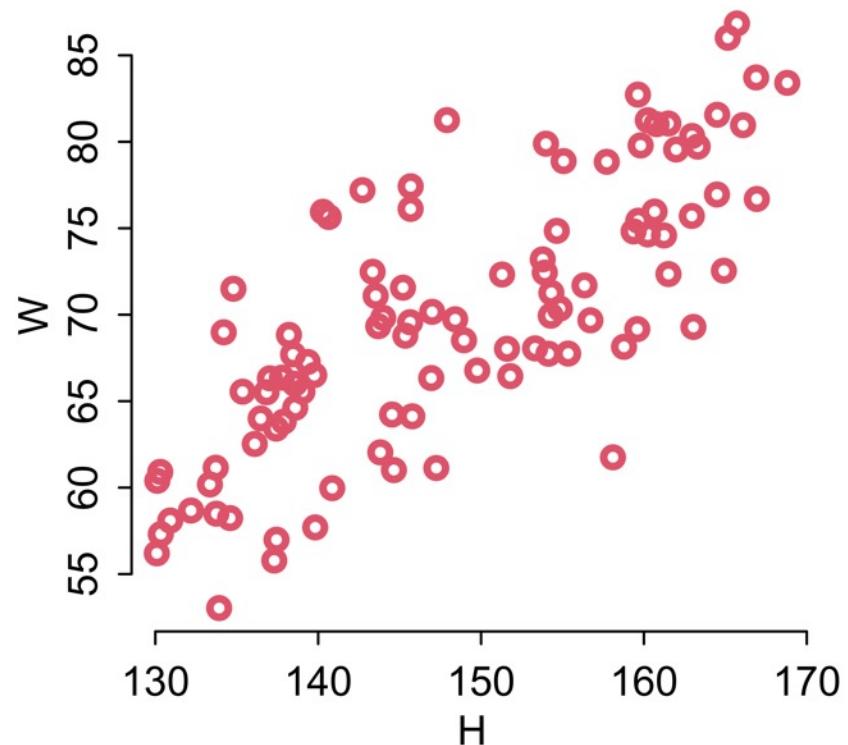


First validate with simulation

```
alpha <- 70
beta <- 0.5
sigma <- 5
n_individuals <- 100
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 )
```

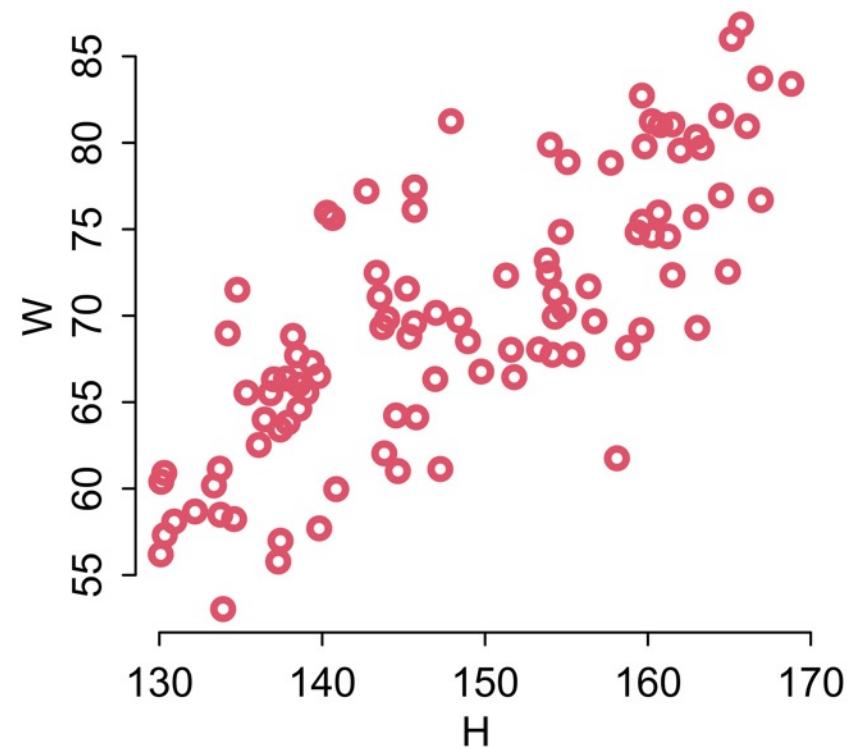


First validate with simulation

```
alpha <- 70
beta <- 0.5
sigma <- 5
n_individuals <- 100
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 )
```



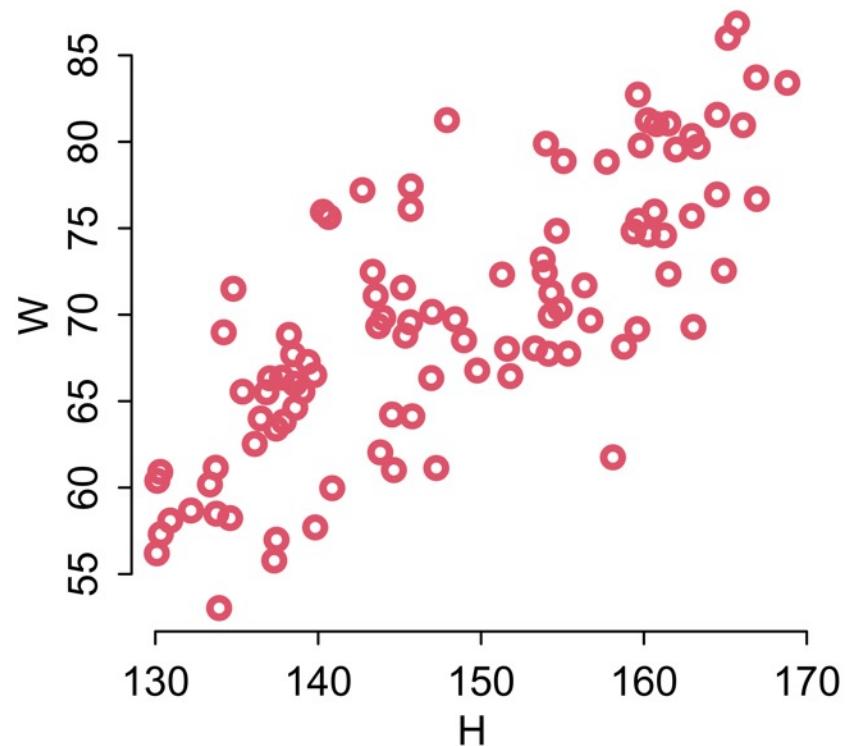
First validate with simulation

```
> 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

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 )
```



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:

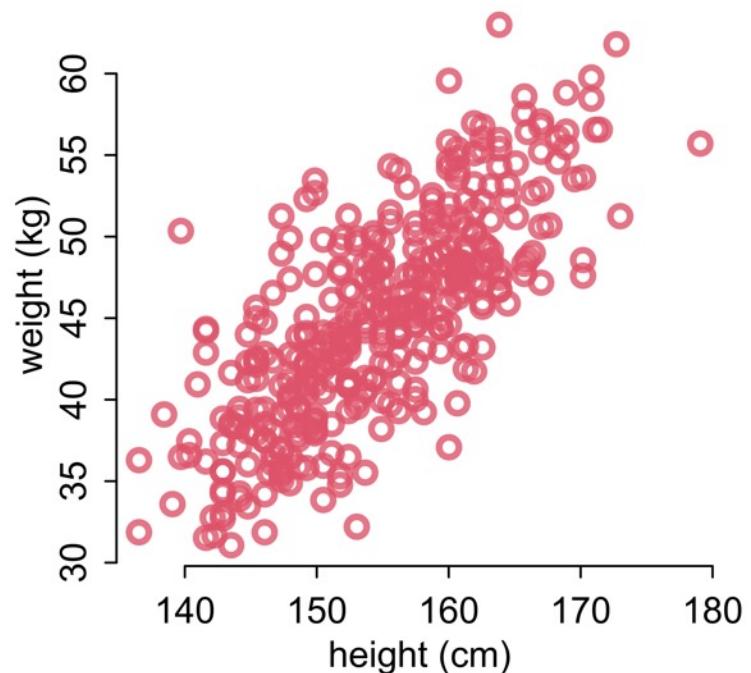
Push out **posterior predictions** and describe/interpret those

```
> 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
>
```

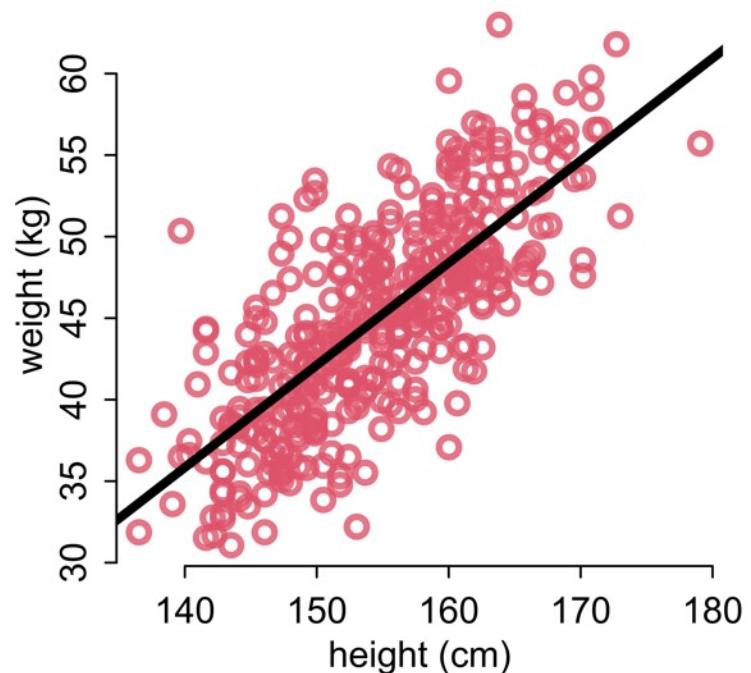
Posterior predictive distribution

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



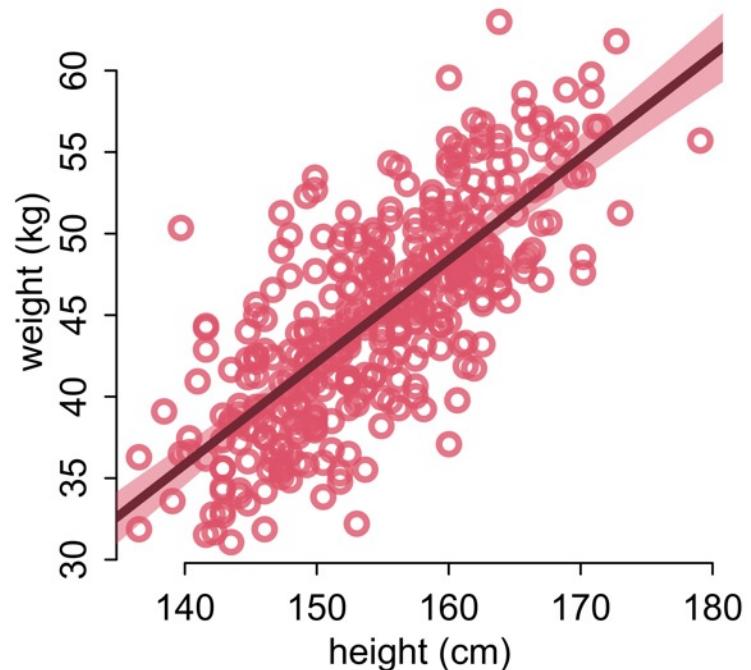
Posterior predictive distribution

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



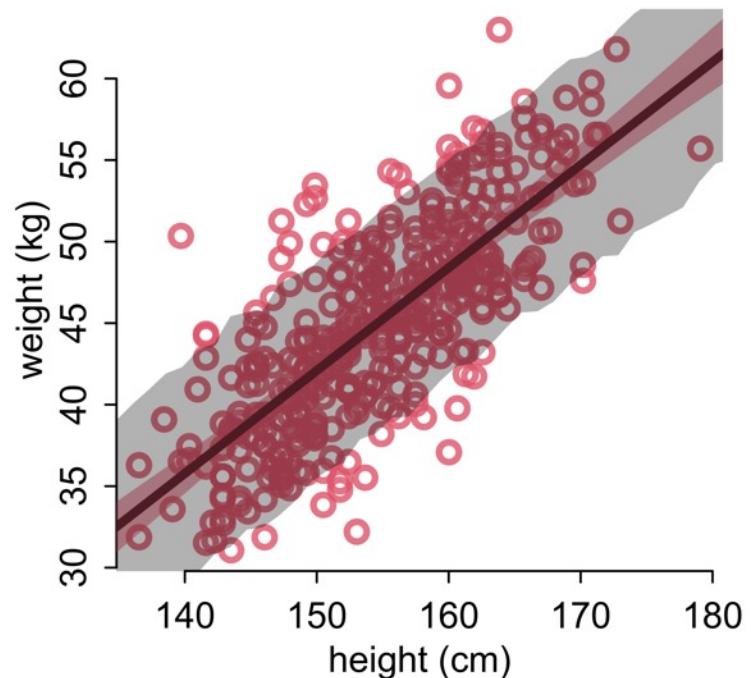
Posterior predictive distribution

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



Posterior predictive distribution

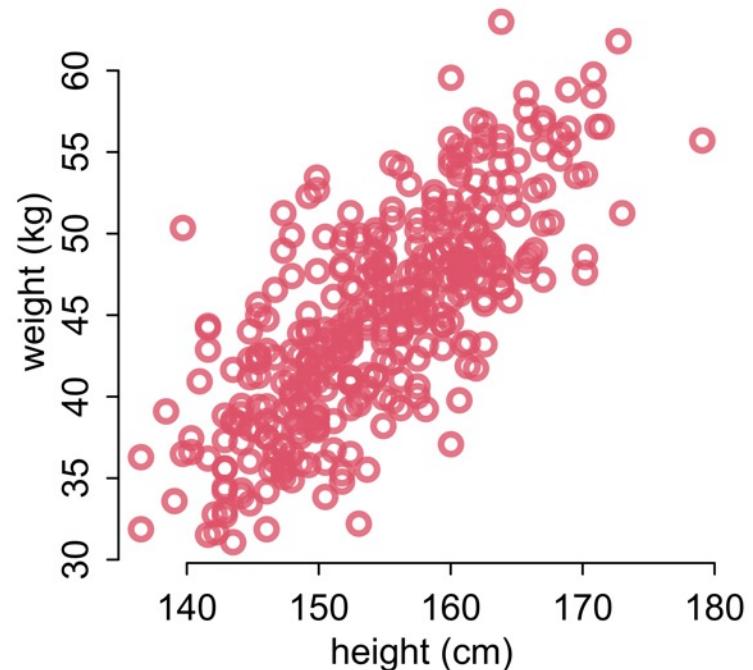
- (1) Plot the sample
- (2) Plot the posterior mean
- (3) Plot uncertainty of the mean
- (4) Plot uncertainty of predictions**



sampling from posterior is the same as doing a posterior predictive distribution

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)" )
```

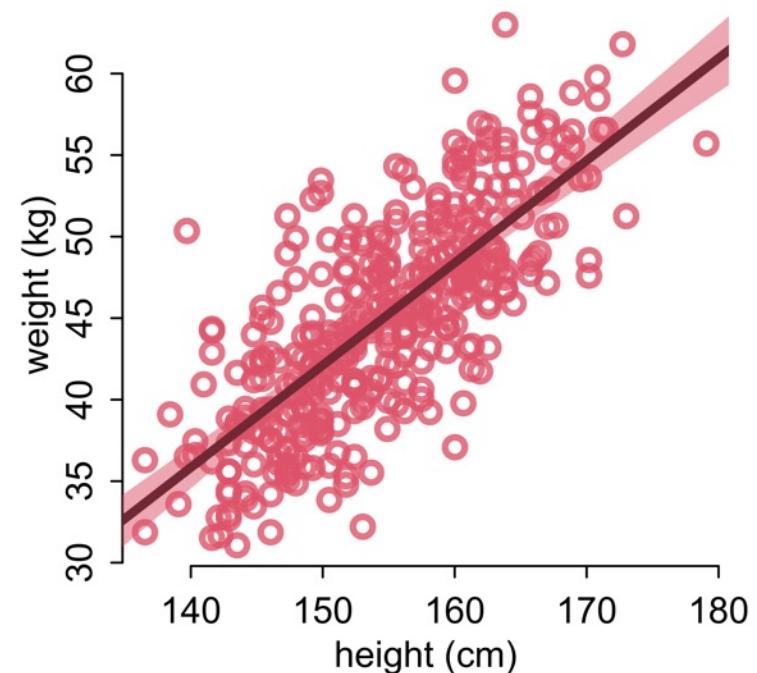


See 4.4.3 starting page 98 in book

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) )
```



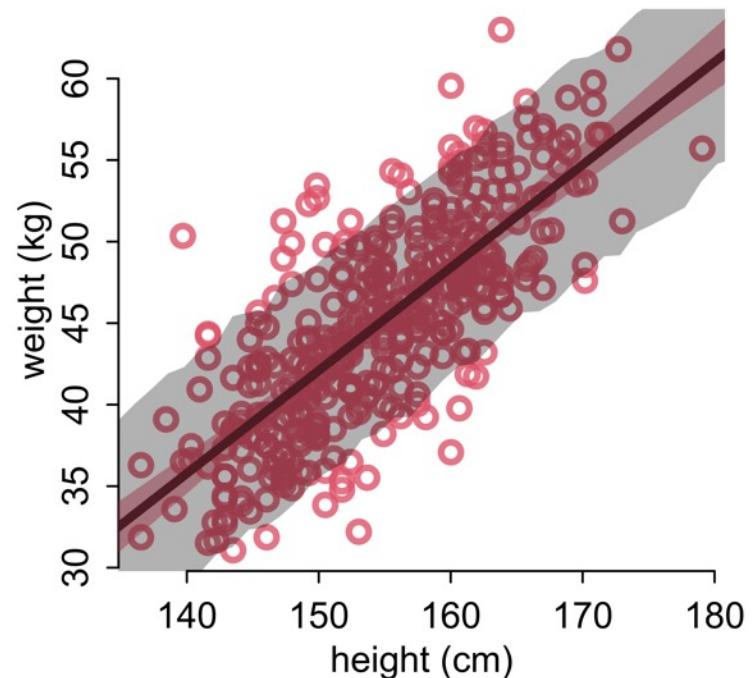
See 4.4.3 starting page 98 in book

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) )
```



See 4.4.3 starting page 98 in book

Flexible Linear Thermometers

(2) Scientific model

How does **height** influence
weight?

$$H \longrightarrow W$$

$$W = f(H)$$

“Weight is some function of height”

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

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

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

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

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