Chapter 4

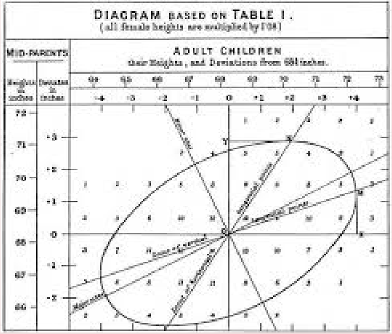
Last week

Assumption for Bayesian model:

(1) Likelihood,
(2) Parameters,
(3) Prior.

Given data preform inference using the posterior.

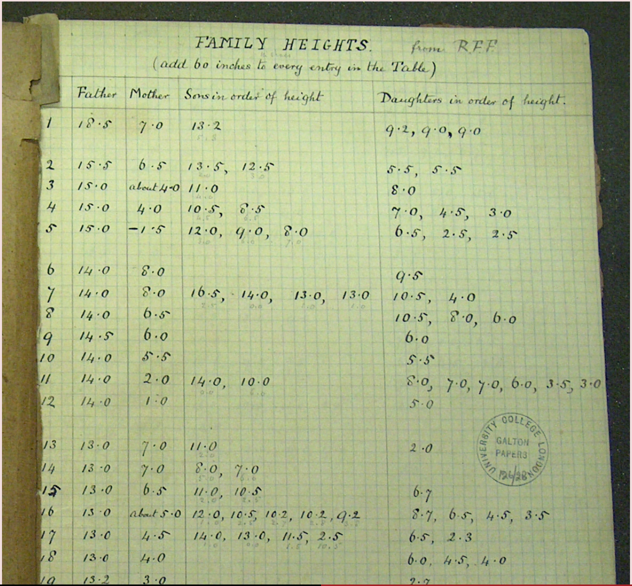
Linear model



Origin:

- Carl Friedrich Gauss perhaps inventor least square.
- Sir Francis Galton (cousin of Darwin).

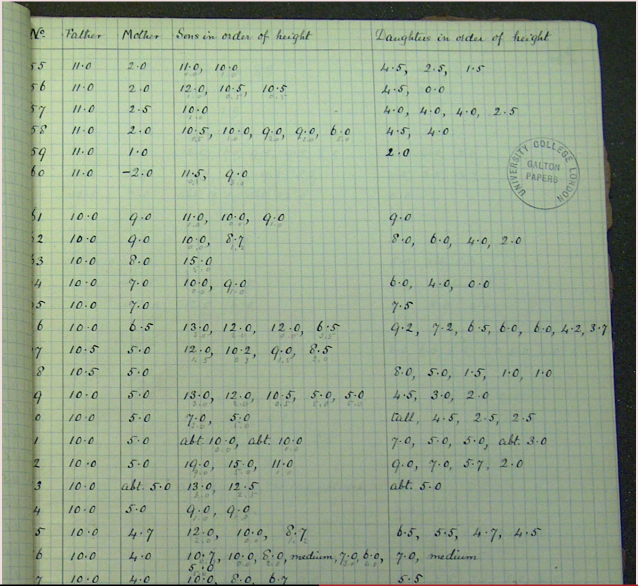
Galton Height data



FAMILY HEIGHTS. from R.F.F.
(add 60 inches to every entry in the Table)

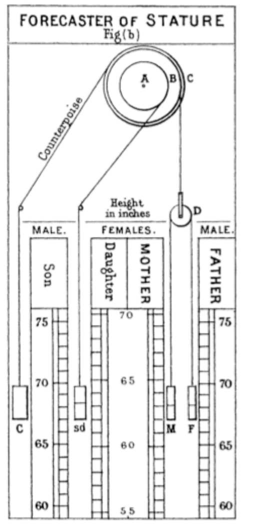
	Father	Mother	Sons in order of height	Daughters in order of height.
1	18·5	7·0	13·2	9·2, 9·0, 9·0
2	15·5	6·5	13·5, 12·5	5·5, 5·5
3	15·0	about 4·0	11·0	8·0
4	15·0	4·0	10·5, 8·5	7·0, 4·5, 3·0
5	15·0	−1·5	12·0, 9·0, 8·0	6·5, 2·5, 2·5
6	14·0	8·0		9·5
7	14·0	8·0	16·5, 14·0, 13·0, 13·0	10·5, 4·0
8	14·0	6·5		10·5, 8·0, 6·0
9	14·5	6·0		6·0
10	14·0	5·5		5·5
11	14·0	2·0	14·0, 10·0	8·0, 7·0, 7·0, 6·0, 3·5, 3·0
12	14·0	1·0		5·0
13	13·0	7·0	11·0	2·0
14	13·0	7·0	8·0, 7·0	
15	13·0	6·5	11·0, 10·5	6·7
16	13·0	about 5·0	12·0, 10·5, 10·2, 10·2, 9·2	8·7, 6·5, 4·5, 3·5
17	13·0	4·5	14·0, 13·0, 11·5, 2·5	6·5, 2·3
18	13·0	4·0		6·0, 4·5, 4·0
19	13·2	3·0		2·7

Galton Height data



No.	Father	Mother	Sons in order of height	Daughters in order of height
55	11·0	2·0	11·0, 10·0	4·5, 2·5, 1·5
56	11·0	2·0	12·0, 10·5, 10·5	4·5, 0·0
57	11·0	2·5	10·0	4·0, 4·0, 4·0, 2·5
58	11·0	2·0	10·5, 10·0, 9·0, 9·0, 6·0	4·5, 4·0
59	11·0	1·0		2·0
60	11·0	−2·0	11·5, 9·0	
61	10·0	9·0	11·0, 10·0, 9·0	9·0
62	10·0	9·0	10·0, 8·7	8·0, 6·0, 4·0, 2·0
63	10·0	8·0	15·0	
4	10·0	7·0	10·0, 9·0	6·0, 4·0, 0·0
5	10·0	7·0		7·5
6	10·0	6·5	13·0, 12·0, 12·0, 6·5	9·2, 7·2, 6·5, 6·0, 6·0, 4·2, 3·7
7	10·5	5·0	12·0, 10·2, 9·0, 8·5	
8	10·5	5·0		8·0, 5·0, 1·5, 1·0, 1·0
9	10·0	5·0	13·0, 12·0, 10·5, 5·0, 5·0	4·5, 3·0, 2·0
0	10·0	5·0	7·0, 5·0	tall, 4·5, 2·5, 2·5
1	10·0	5·0	abt. 10·0, abt. 10·0	7·0, 5·0, 5·0, abt. 3·0
2	10·0	5·0	19·0, 15·0, 11·0	9·0, 7·0, 5·7, 2·0
3	10·0	abt. 5·0	13·0, 12·5	abt. 5·0
4	10·0	5·0	9·0, 9·0	
5	10·0	4·7	12·0, 10·0, 8·7	6·5, 5·5, 4·7, 4·5
6	10·0	4·0	10·7, 10·0, 8·0, medium, 7·0, 6·0, 5·0	7·0, medium
7	10·0	4·0	10·0, 8·0, 6·7	5·5

Galton Height data



Linear model

- Descriptively accurate
- Mechanistically often wrong.
- Easy to fit.

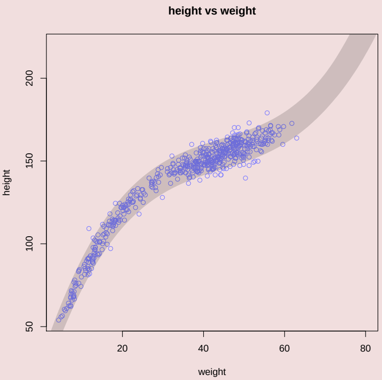


Figure: Polynomial regression

Why Normal?

- The normal distribution, is the most important distribution in statistics.
- Many mechanism creates an end product that follows a Normal distribution. Like sums of random variables, products of random variables.
- The distribution is the easiest to handle computationally.

Sum of Uniform

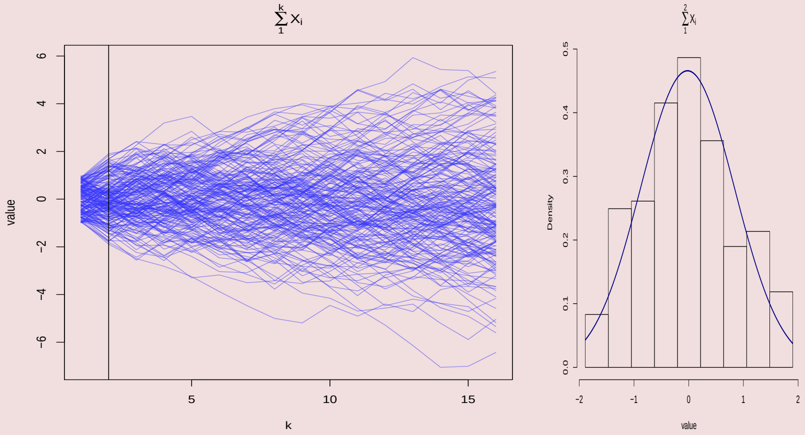


Figure: $\sum_{i=1}^{k} X_i,\ p(x) = \frac{1}{2}\mathbb{I}_{[-1,1]}(x)$

Sum of Uniform

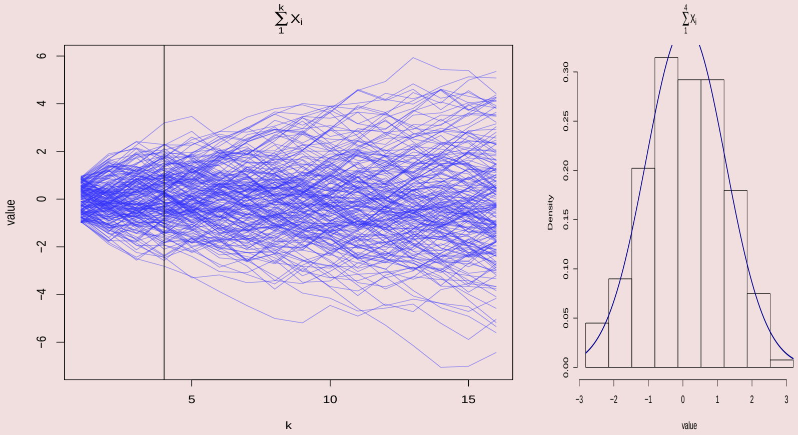


Figure: $\sum_{i=1}^{k} X_i,\ p(x) = \frac{1}{2}\mathbb{I}_{[-1,1]}(x)$

Sum of Uniform

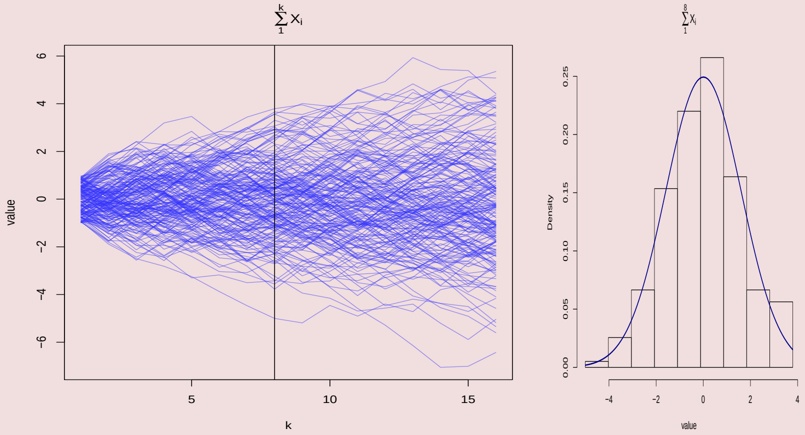


Figure: $\sum_{i=1}^{k} X_i, \ p(x) = \frac{1}{2}\mathbb{I}_{[-1,1]}(x)$

Sum of Uniform

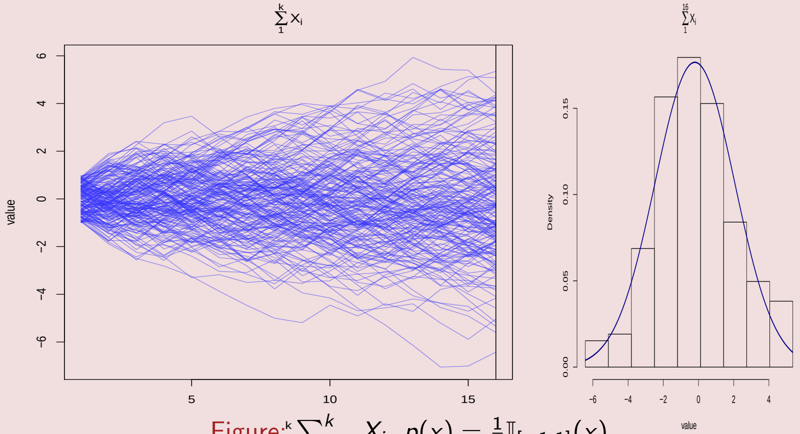


Figure: $\sum_{i=1}^{k} X_i,\ p(x) = \frac{1}{2}\mathbb{I}_{[-1,1]}(x)$

Sum of Poisson

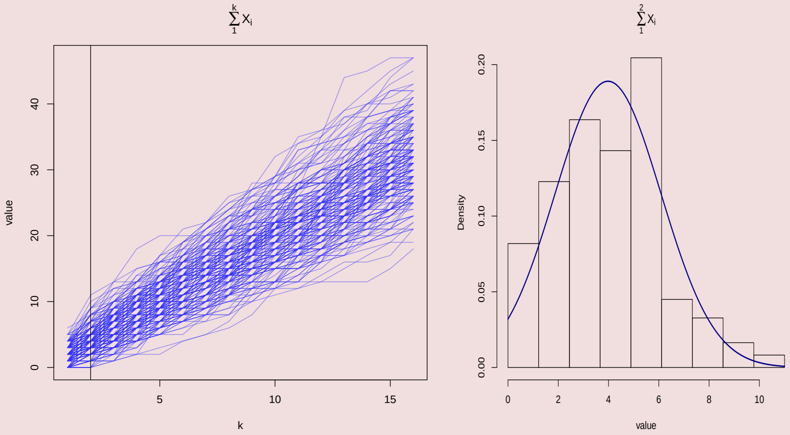


Figure: $\sum_{i=1}^{k} X_i,\ p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$

Sum of Poisson

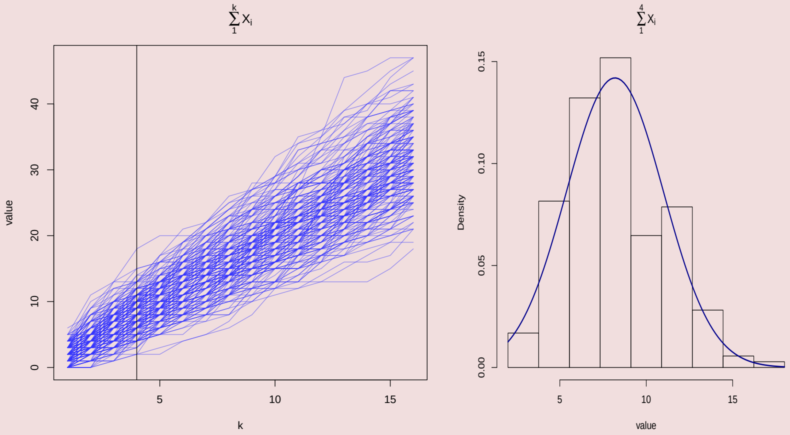


Figure: $\sum_{i=1}^{k} X_i$, $p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$

Sum of Poisson

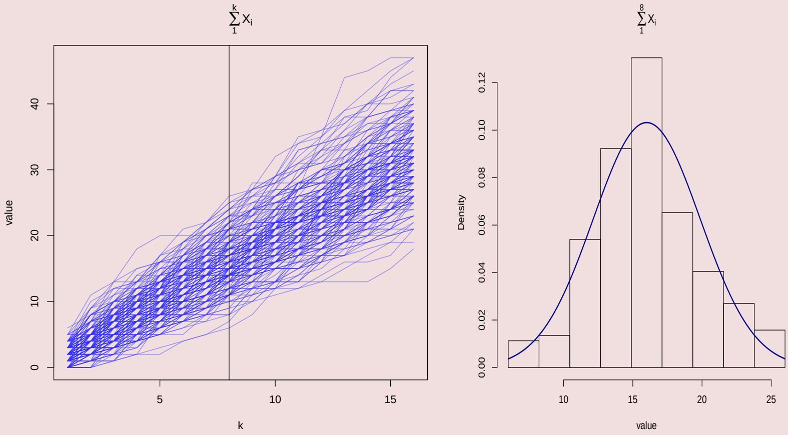


Figure: $\sum_{i=1}^{k} X_i,\ p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$

Sum of Poisson

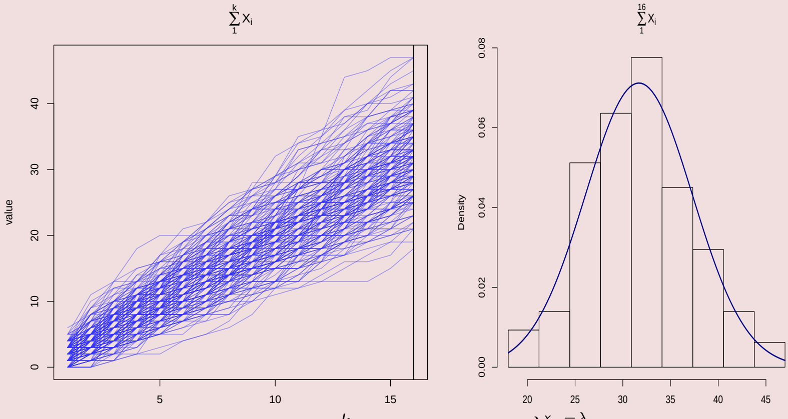
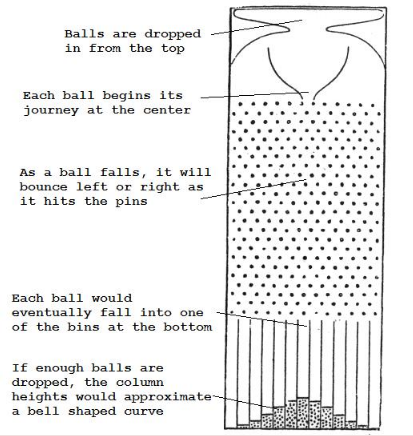


Figure: $\sum_{i=1}^{k} X_i$, $p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$

Galtons Normal distribution



Model

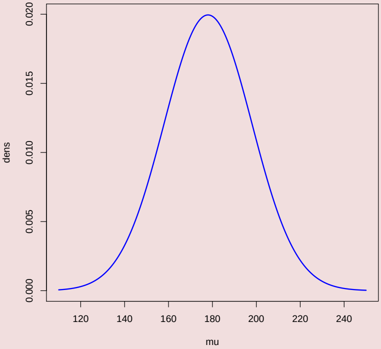
$$n_F \sim Binomial(n, p)$$
$$p \sim U[0, 1]$$

A first model of height

$$
\begin{aligned}
h_i &\sim N(\mu, \sigma) \\
\mu &\sim N(178, 20) \\
\sigma &\sim U[0, 50]
\end{aligned}
$$

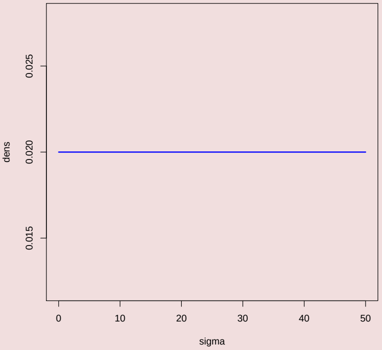
A first model of height

$$h_i \sim N(\mu, \sigma)$$
$$\mu \sim N(178, 20)$$
$$\sigma \sim U[0, 50]$$



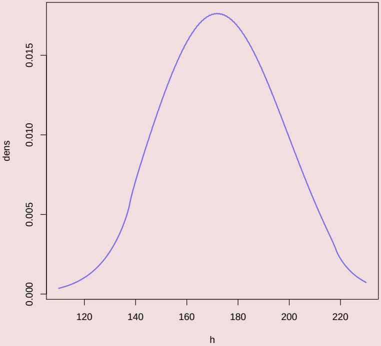
A first model of height

$$h_i \sim N(\mu, \sigma)$$
$$\mu \sim N(178, 20)$$
$$\sigma \sim U[0, 50]$$



A first model of height

$$h_i \sim N(\mu, \sigma)$$
$$\mu \sim N(178, 20)$$
$$\sigma \sim U[0, 50]$$



Model to distribution

Model	density
$h_i \sim N(\mu, \sigma)$	$N(h_i \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(h_i - \mu)^2}$
$\mu \sim N(178, 20)$	$p(\beta) = \frac{1}{\sqrt{2\pi 20}} e^{-\frac{1}{2 \cdot 20^2}(\beta - 178)^2}$
$\sigma \sim U(0, 50)$	$p(\sigma) = \frac{1}{50} \mathbb{I}_{[0,50]}(\sigma)$

Unless stated variables are independent.

Fitting the distribution

- Observing a height h_1, the posterior distribution is

$$
\begin{aligned}
p(\mu, \sigma | h_1) &= \frac{N(h_1|\mu,\sigma)p(\mu)p(\sigma)}{\int\int N(h_1|\tilde{\mu},\tilde{\sigma})p(\tilde{\mu})p(\tilde{\sigma})d\tilde{\mu}d\tilde{\sigma}} \\
&\propto N(h_1|\mu,\sigma)p(\mu)p(\sigma)
\end{aligned}
$$

- Observing two heights h_1, h_2, the posterior distribution is

$$
\begin{aligned}
p(\mu, \sigma | h_1, h_2) &= \frac{N(h_1|\mu,\sigma)N(h_2|\mu,\sigma)p(\mu)p(\sigma)}{\int\int N(h_1|\tilde{\mu},\tilde{\sigma})N(h_2|\tilde{\mu},\tilde{\sigma})p(\tilde{\mu})p(\tilde{\sigma})d\tilde{\mu}d\tilde{\sigma}} \\
&\propto N(h_1|\mu,\sigma)N(h_2|\mu,\sigma)p(\mu)p(\sigma)
\end{aligned}
$$

Posterior distribution of μ

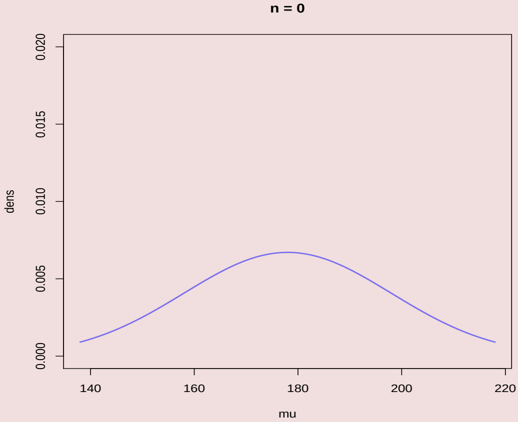


Figure: $p(\mu|h_1, \ldots, h_n)$

Posterior distribution of μ

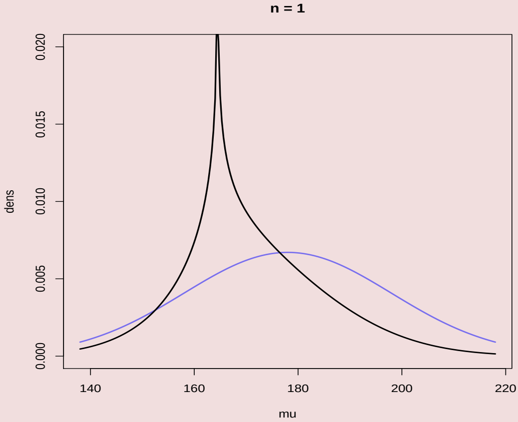


Figure: $p(\mu|h_1, \ldots, h_n)$

Posterior distribution of μ

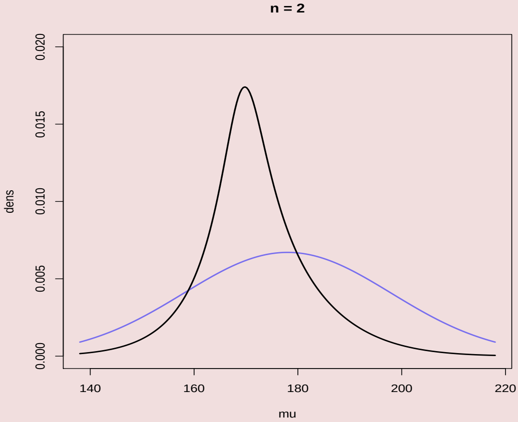


Figure: $p(\mu|h_1, \ldots, h_n)$

Posterior distribution of μ

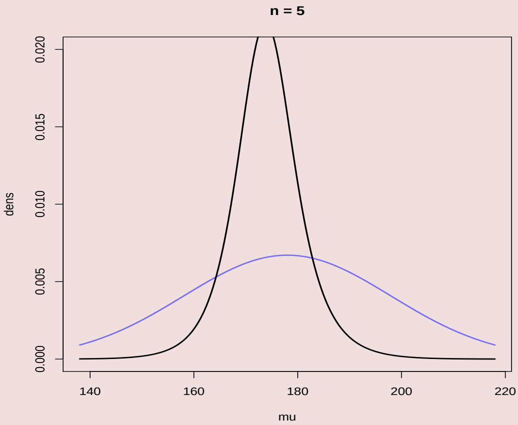


Figure: $p(\mu|h_1, \ldots, h_n)$

Posterior distribution of μ

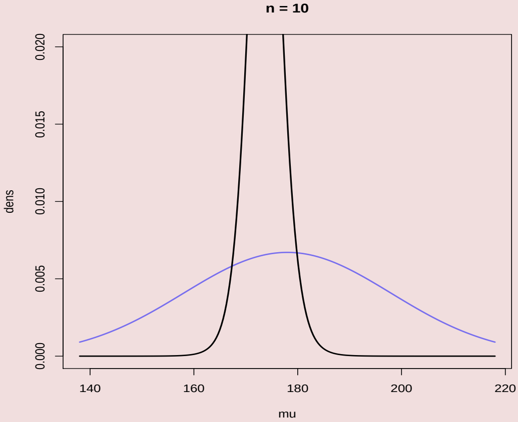


Figure: $p(\mu|h_1, \ldots, h_n)$

Fitting the distribution in R

- We will use the **map** function.
- Uses quadratic approximation of the posterior. We will treat it as true samples from the posterior distribution.

predictive Height model

$$h_i \sim N(\mu, \sigma)$$
$$\mu \sim N(178, 20)$$
$$\sigma \sim U[0, 50]$$

```
library(rethinking)
h<- c(...) # vector of heights
model  <- map(
        flist = alist(
              height ~  dnorm(
              mu
~  dnorm(178, 20),
              sigma
~  dunif(0,50)
              ),
        data  = list(height = h))
```

Posterior Samples from the model

```
post <- extract.samples(model, n = 100)
head(post, n = 3)
```

Output:

	mu	sigma
1	154.6512	
2	154.2872	
3	154.1929	

Posterior Samples from the predictive distribution

```
post  <- extract.samples(model, n = 100)
hstar <- rnorm(n = 100, mean = post$mu, sd = post$sigma)
head(hstar, n = 3)
```

Output:

```
[1] 176.6431 183.7880 178.33
```

Predictive distribution, $p(h^*|\cdot)$

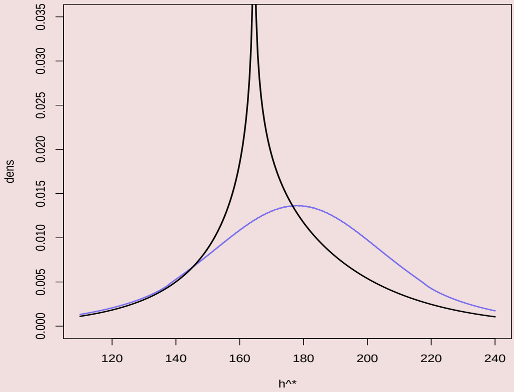


Figure: $p(h^*|h_1, \ldots, h_n)$

Predictive distribution, $p(h^*|\cdot)$

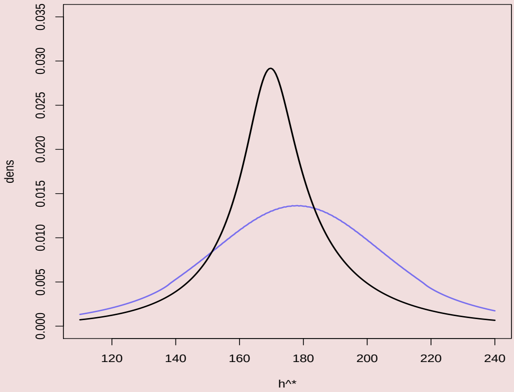


Figure: $p(h^*|h_1, \ldots, h_n)$

Predictive distribution, $p(h^*|\cdot)$

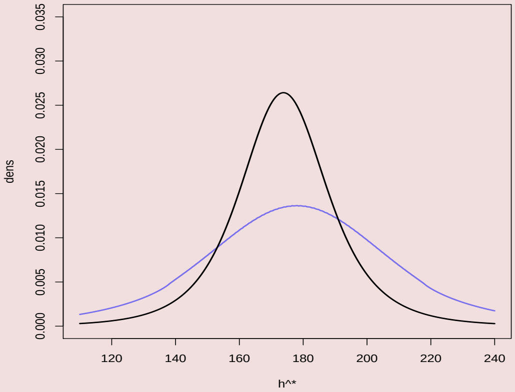


Figure: $p(h^*|h_1, \ldots, h_n)$

Predictive distribution, $p(h^*|\cdot)$

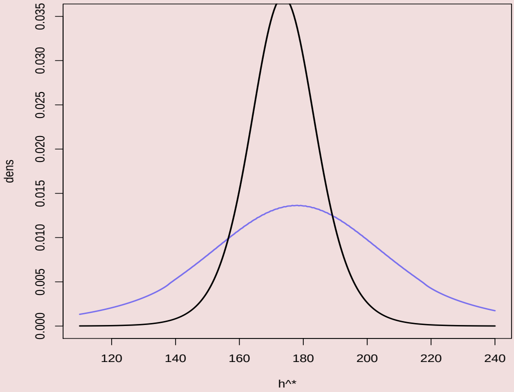


Figure: $p(h^*|h_1, \ldots, h_n)$

Predictive distribution, $p(h^*|\cdot)$

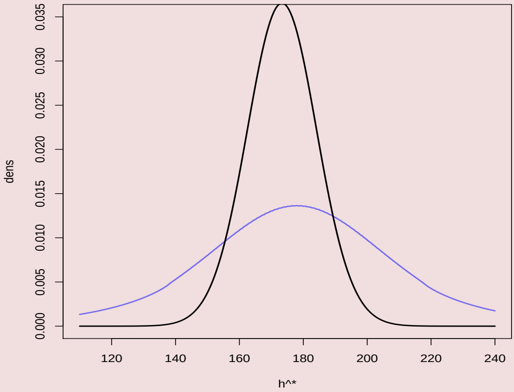
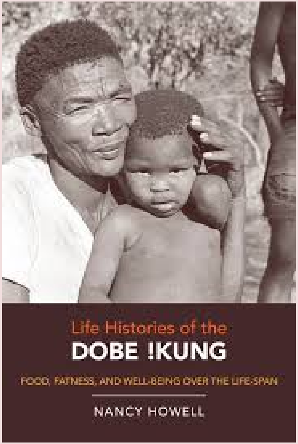


Figure: $p(h^*|h_1, \ldots, h_n)$

Second height data set




```r
library(rethinking)
data(Howell1)
head(Howell1)
```

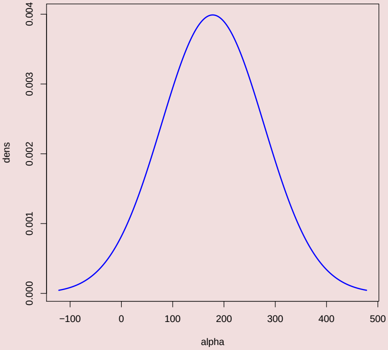
```
height
weight age male
1 151.7
63 1
2 139.7
63 0
3 136.5
65 0
4 156.8
41 1
5 145.4
51 0
```

Adding a predictor

$$
\begin{aligned}
h_i &\sim N(\mu_i, \sigma) \\
\mu_i &= \alpha + w_i \beta \\
\alpha &\sim N(178, 100) \\
\beta &\sim N(0, 20) \\
\sigma &\sim U[0, 50]
\end{aligned}
$$

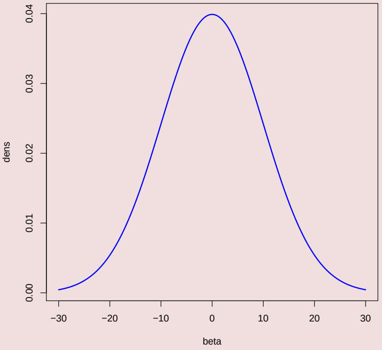
Adding a predictor

$$h_i \sim N(\mu_i, \sigma)$$
$$\mu_i = \alpha + w_i \beta$$
$$\alpha \sim N(178, 100)$$
$$\beta \sim N(0, 20)$$
$$\sigma \sim U[0, 50]$$



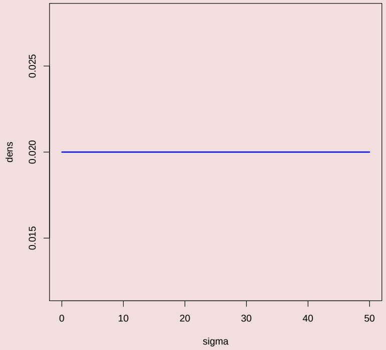
Adding a predictor

$$h_i \sim N(\mu_i, \sigma)$$
$$\mu_i = \alpha + w_i \beta$$
$$\alpha \sim N(178, 100)$$
$$\beta \sim N(0, 20)$$
$$\sigma \sim U[0, 50]$$



Adding a predictor

$$h_i \sim N(\mu_i, \sigma)$$
$$\mu_i = \alpha + w_i \beta$$
$$\alpha \sim N(178, 100)$$
$$\beta \sim N(0, 20)$$
$$\sigma \sim U[0, 50]$$



Model to distribution

Model	density
$h_i \sim N(\mu_i, \sigma)$	$p(h_i \mid \mu_i, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(h_i - \mu_i)^2}$
$\mu_i = \alpha + w_i \cdot \beta$	
$\alpha \sim N(156, 100)$	$p(\alpha) = \frac{1}{\sqrt{2\pi}100} e^{-\frac{1}{2 \cdot 100^2}(\alpha - 156)^2}$
$\beta \sim N(0, 20)$	$p(\beta) = \frac{1}{\sqrt{2\pi}20} e^{-\frac{1}{2 \cdot 20^2}(\beta)^2}$
$\sigma \sim U(0, 1)$	$p(\sigma) = \frac{1}{50} \mathbb{I}_{[0,50]}(\sigma)$

Unless stated variables are independent.

predictive Height model

$$h_i \sim N(\mu_i, \sigma)$$
$$\mu_i = \alpha + x_i\beta$$
$$\alpha \sim N(178, 100)$$
$$\beta \sim N(0, 20)$$
$$\sigma \sim U[0, 50]$$

```
library(rethinking)
data(Howell1)
dataHeight <- Howell1[Howell1$age >= 18,]

# building the model
model2 <- map(
flist = alist(
height ~ dnorm(mu, sigma),
mu     <- alpha + weight * beta ,
alpha  ~ dnorm(156, 100),
beta   ~ dnorm(0   , 10),
sigma  ~ dunif(0,50)
),
data  = dataHeight)
```

Posterior Samples from the models

```r
post <- extract.samples(model2, n = 100)
head(post, n = 3)
```

Output:

```
     alpha      beta       sigma
1 115.8987 0.8559246 5.17037
2 114.2855 0.8911623 5.18144
3 111.4015 0.9479790 5.14385
```

Posterior of α, β in Table form

```
precis(model2, prob = 0.95)
```

Output:

	Mean	StdDev	2.5%	97.5%
alpha	113.89	1.91	110.16	117.63
beta	0.90	0.04	0.82	0.99
sigma	5.07	0.19	4.70	5.45

Posterior of α, β in Table form

```
precis(model2, prob = 0.95, corr =T)
```

Output:

```
                 Mean   StdDev    2.5%   97.5%
alpha  beta  sigma
          alpha  113.89    1.91  110.16  117.63
1.00  −0.99   0
          beta     0.90    0.04    0.82    0.99  −0.99
1.00   0
          sigma    5.07    0.19    4.70    5.45
0.00   0.00   1
```

Posterior of α, β plot

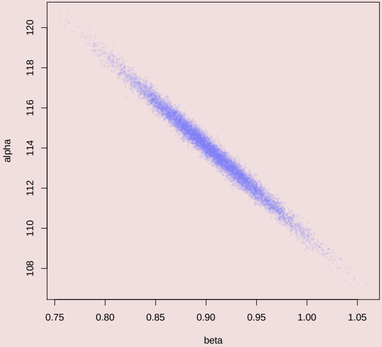


Figure: samples from $p(\alpha, \beta)$

Predictive function $\mu(x)$

- In the book, weight is from 31 to 63.
- We have a prior on the function:

$$\mu(x) = \alpha + x\beta$$

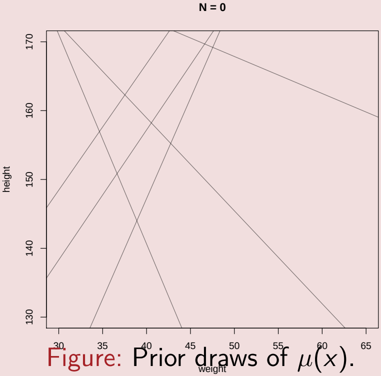


Figure: Prior draws of $\mu(x)$.

Posterior Samples from the models

```
mu.function      <- function(x) post$alpha + post$beta * x
head(mu.function(1))
```

```
[1] 115.4309 116.2358 116.2933
[4] 118.6777 112.4493 117.0376
```

Posterior $\mu(x)$

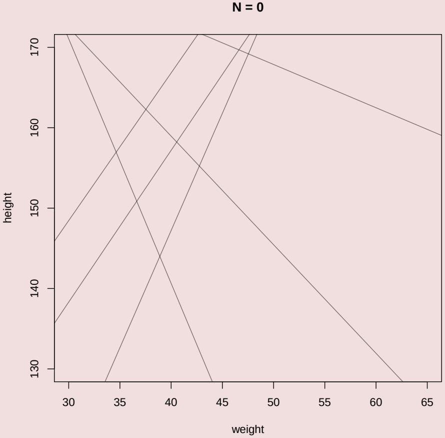


Figure: $\mu(x) = \alpha + x\beta | h_1, \ldots, h_n$

Posterior $\mu(x)$

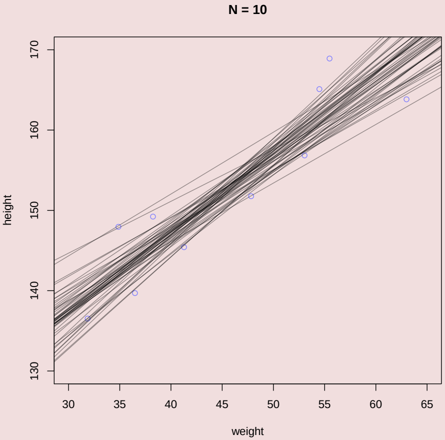


Figure: $\mu(x) = \alpha + x\beta | h_1, \ldots, h_n$

Posterior $\mu(x)$

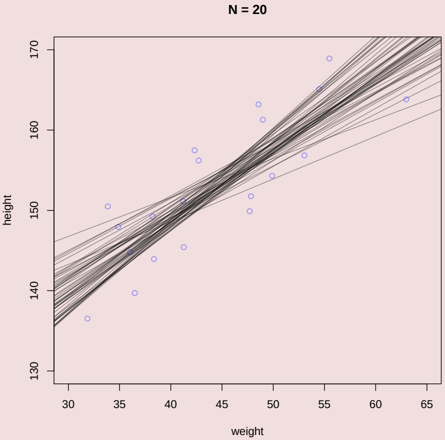


Figure: $\mu(x) = \alpha + x\beta | h_1, \ldots, h_n$

Posterior $\mu(x)$

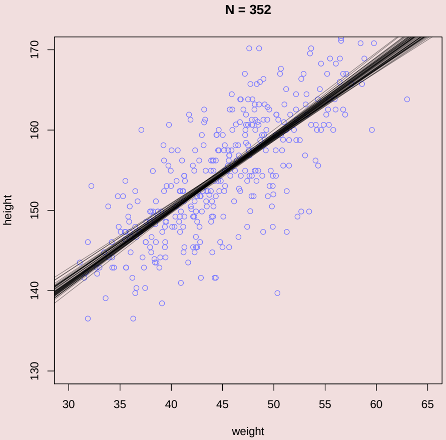


Figure: $\mu(x) = \alpha + x\beta | h_1, \ldots, h_n$

Posterior $\mu(x)$ shady version PI_{80}

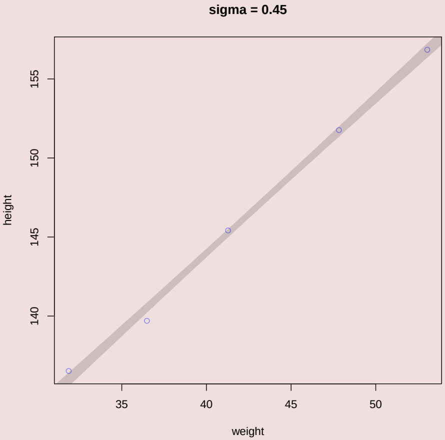


Figure: $\mu(x) = \alpha + x\beta | h_1, \ldots, h_n$

Posterior $\mu(x)$ shady version PI_{80}

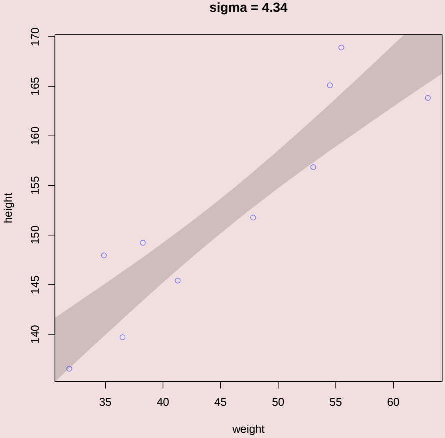


Figure: $\mu(x) = \alpha + x\beta | h_1, \ldots, h_n$

Posterior $\mu(x)$ shady version PI_{80}

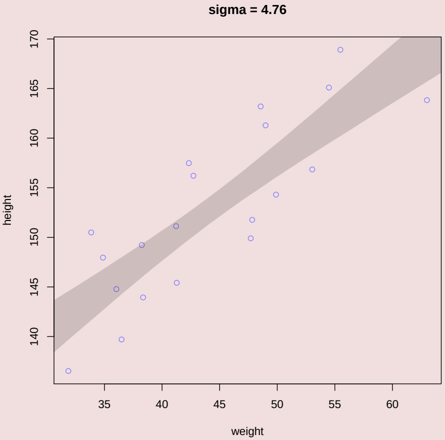


Figure: $\mu(x) = \alpha + x\beta | h_1, \ldots, h_n$

Posterior $\mu(x)$ shady version PI_{80}

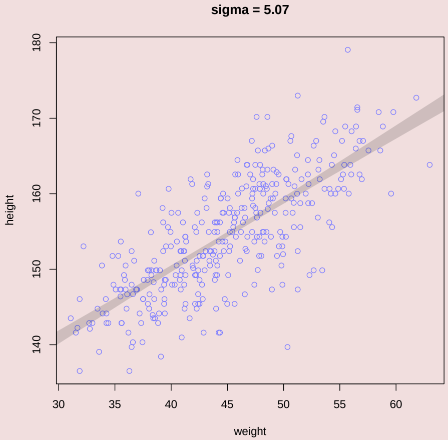


Figure: $\mu(x) = \alpha + x\beta | h_1, \ldots, h_n$

Posterior Samples from the models

```
mu.function    <- function(x) post$alpha + post$beta * x
weight.seq     <- seq(from  =25, 70, by = 0.1)
mus            <- sapply(weight.seq, mu.function)
mus[1:4,1:4]
```

```
         [,1]     [,2]     [,3]     [,4]
[1,] 136.5234 136.6112 136.6991 136.7870
[2,] 137.1122 137.1992 137.2862 137.3732
[3,] 137.0881 137.1747 137.2613 137.3480
[4,] 138.3417 138.4236 138.5056 138.5875
```

Posterior Samples from the models

```
mu.function      <- function(x) post$alpha + post$beta * x
weight.seq       <- seq(from  =25, 70, by = 0.1)
mus              <- sapply(weight.seq, mu.function)
mus.PI           <- apply(mus, 2, function(x){ quantile(x,c(0.1,0.9))})
mus.PI[,1:4]
```

```
             [,1]     [,2]     [,3]
[,4]
10% 135.3868 135.4827 135.57
90% 137.6392 137.7245 137.81
```

Posterior Samples from the models

```
mu.function   <- function(x) post$alpha + post$beta * x
weight.seq    <- seq(from =25, 70, by = 0.1)
mus           <- sapply(weight.seq, mu.function)
mus.PI        <- apply(mus, 2, function(x){ quantile(x,c(0.1,0.9))})
```

Posterior Samples from the models

```
mu.function      <- function(x) post$alpha + post$beta * x
weight.seq       <- seq(from  =25, 70, by = 0.1)
mus              <- sapply(weight.seq, mu.function)
mus.PI           <- apply(mus, 2, function(x){ quantile(x,c(0.1,0.9))})
plot(.)
shade(mus.PI, weight.seq)
```

Posterior Samples from the models

```
heights          <- mus+rnorm(n = length(post.samples$sigma),
sd = post$sigma)
height.PI        <- apply(heights, 2, function(x){ quantile(x,c(0.1,0.9))})
plot(.)
shade(mus.PI, weight.seq)
shade(height.PI, weight.seq)
```

Posterior h^* shady version

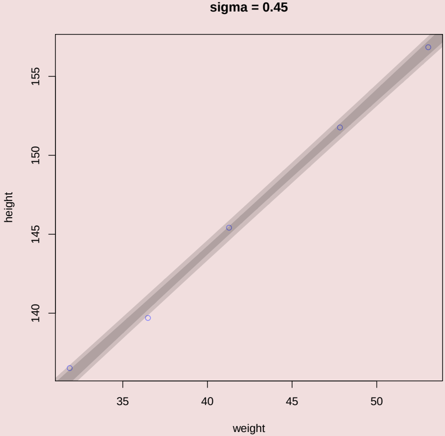


Figure: $p(h^*|x, h_1, \ldots, h_n)$

Posterior h^* shady version

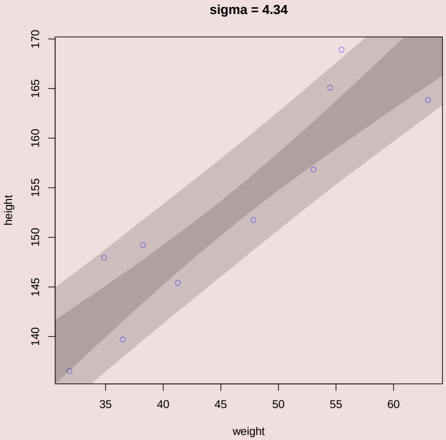


Figure: $p(h^* | x, h_1, \ldots, h_n)$

Posterior h^* shady version

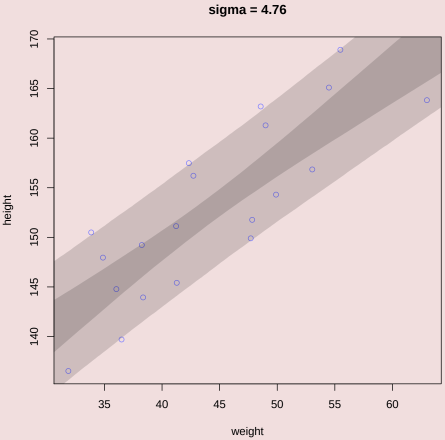


Figure: $p(h^*|x, h_1, \ldots, h_n)$

Posterior h^* shady version

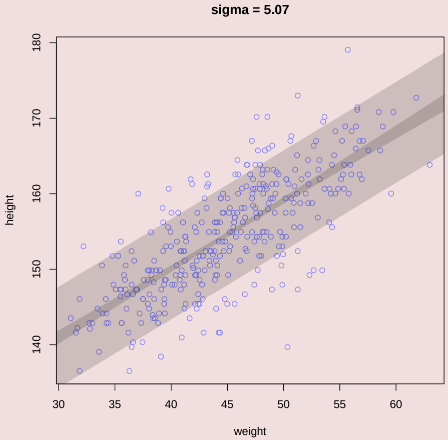


Figure: $p(h^*|x, h_1, \ldots, h_n)$

Comparing the models

The posterior distribution for h^* given $x^* = 30, 50$:

Figure:

Comparing the models

The posterior distribution for h^* given $x^* = 30, 50$:

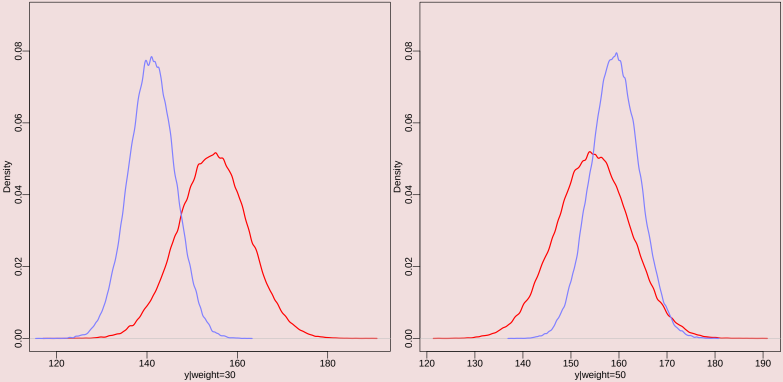


Figure: $p(y^*|x^* = \cdot, h_1, \ldots, h_n)$, blue predictive model and red simple model.