# Chapter 4 Exercises

*Emily Maloney*

*January 28, 2019*

## Chapter 4

```
library(tidyverse)
library(brms)
library(tidybayes)
```

### Easy Problems

#### 4E1

In this model, the likelihood is defined by $y\ i \sim \text{Normal}(\mu, \sigma)$.

#### 4E2

There are 2 parameters in the posterior distribution of this model.

#### 4E3

omit

#### 4E4

The line describing the linear model is $\mu\ i = \alpha + \beta\ x\ i$.

#### 4E5

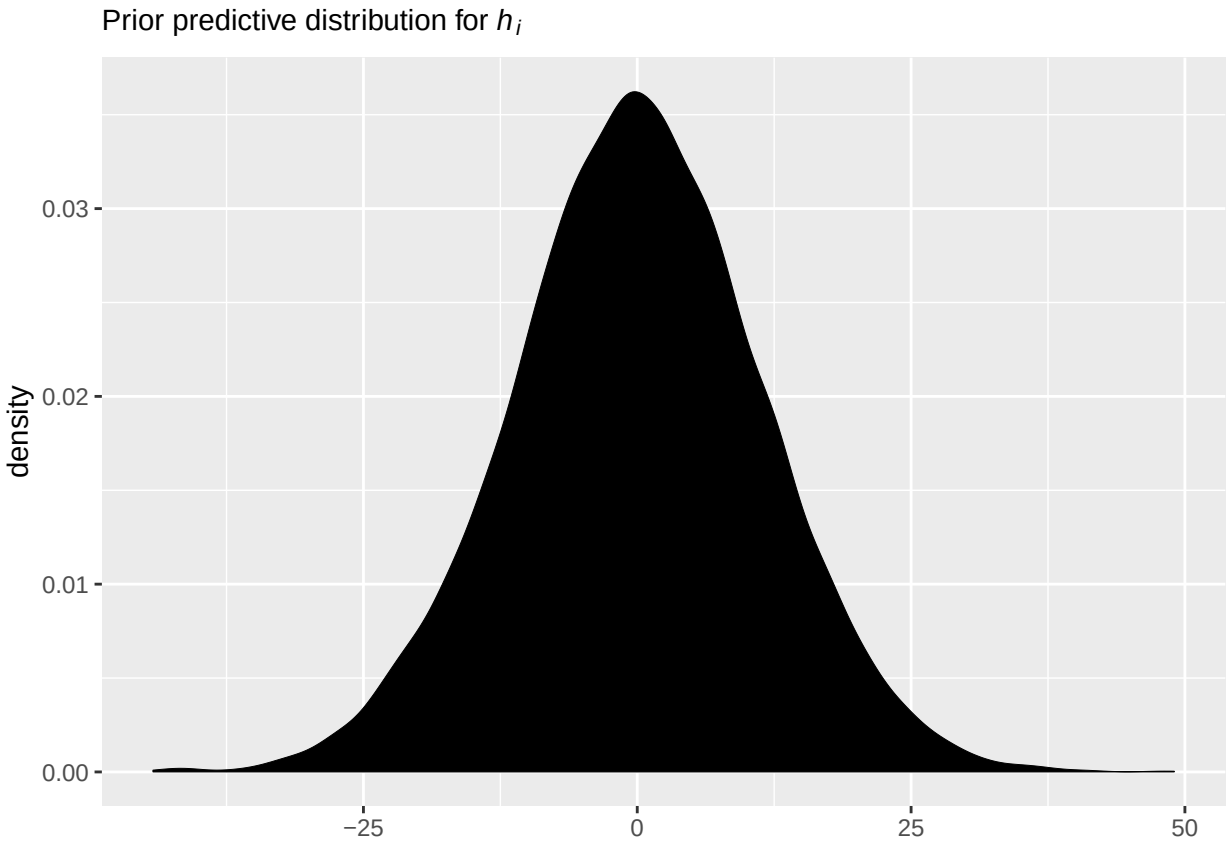
There are 3 parameters in the posterior distribution of this model.

### Medium Problems

#### 4M1

```
#sampling from both priors to get simulation of observed heights
n <- 1e4
set.seed(432)
tibble(sample_mu = rnorm(n, mean = 0, sd = 10),
       sample_sigma = runif(n, min = 0, max = 10)) %>%
       mutate(x = rnorm(n, mean = sample_mu, sd = sample_sigma)) %>%

       ggplot(aes(x = x)) +
       geom_density(fill = "black", size = 0) +
       labs(subtitle = expression(paste("Prior predictive distribution for ", italic(h[i]))),
            x = NULL)
```




**4M2**

The model translated into a map formula is: flist <- alist ( y ~ dnorm(mu, sigma), mu ~ dnorm(0, 10), sigma ~ dunif(0, 10) )

**4M3**

The map model formula translated into a mathematical model definition is: $y\ i \sim \text{Normal}(\mu, \sigma)$
$\mu\ i = \alpha + \beta\ x\ i$
$\alpha \sim \text{Normal}(0, 50)$
$\beta \sim \text{Uniform}(0, 10)$
$\sigma \sim \text{Uniform}(0, 50)$

**4M4**

The mathematical model definitions for predicting height using year as a predictor I would use is:
$y\ i \sim \text{Normal}(\mu, \sigma)$
$\mu\ i = \alpha + \beta\ x\ i$
$\alpha \sim \text{Normal}(107, 10)$ #assuming kindergarten-3rd grade so ~ 3.5 feet tall mean?
$\beta \sim \text{Normal}(7, 2)$ #how many inches grow each year, assuming 3 inches and 1 std dev?
$\sigma \sim \text{Uniform}(0, 50)$

**4M5**

Yes, would change to $\alpha \sim \text{Normal}(120, 10)$, but would likely leave $\beta$ specification alone.

### 4M6

I would change the specification for $\sigma$ to $\sigma \sim \text{Uniform}(0, 64)$.

## Hard Problems

### 4H1

```
library(rethinking)
```

```
## Loading required package: rstan
```

```
## Loading required package: StanHeaders
```

```
## rstan (Version 2.18.2, GitRev: 2e1f913d3ca3)
```

```
## For execution on a local, multicore CPU with excess RAM we recommend calling
## options(mc.cores = parallel::detectCores()).
## To avoid recompilation of unchanged Stan programs, we recommend calling
## rstan_options(auto_write = TRUE)
```

```
## For improved execution time, we recommend calling
## Sys.setenv(LOCAL_CPPFLAGS = '-march=native')
## although this causes Stan to throw an error on a few processors.
```

```
##
## Attaching package: 'rstan'
```

```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
## Loading required package: parallel
```

```
## rethinking (Version 1.59)
```

```
##
## Attaching package: 'rethinking'
```

```
## The following objects are masked from 'package:brms':
##
##     LOO, stancode, WAIC
```

```
## The following object is masked from 'package:purrr':
##
##     map
```

```
library(tidyverse)
library(knitr)
data(Howell1) # load in data

d <- Howell1
#d2 <- d %>% filter(age >= 18) # filter to only adults

#fit model
mhw.1 <- map(alist(
              height ~ dnorm(mu, sigma),
              mu <- a + b*weight,
              a ~ dnorm(156, 100),
              b ~ dnorm(0, 10),
```

```
            sigma ~ dunif(0, 50)
          ),
        data = d)

#summary call for what's in the model
precis(mhw.1)
```

```
##        Mean StdDev  5.5% 94.5%
## a     75.45   1.05 73.77 77.12
## b      1.76   0.03  1.72  1.81
## sigma  9.35   0.28  8.89  9.80
```

```
#steve's code
N <- 1e4 # sample size

# Get predictive means and data
preds <-
  as.tibble(MASS::mvrnorm(mu = mhw.1@coef,
                          Sigma = mhw.1@vcov , n = N )) %>%      # rather than extract.samples
  mutate(weight = sample(c(46.95, 43.72, 64.78, 32.59, 54.63), N, replace = T),
         predmean = a + b * weight ,                             # line uncertainty
         predverb = rnorm(N, a + b*weight, sigma )) %>%          # data uncertainty
  group_by(weight) %>%
  mutate(lb_mu = rethinking::HPDI(predmean, prob = .89)[1],
         ub_mu = rethinking::HPDI(predmean, prob = .89)[2],
         lb_ht = rethinking::HPDI(predverb, prob = .89)[1],
         ub_ht = rethinking::HPDI(predverb, prob = .89)[2]) %>%
  slice(1) %>%
  mutate(yhat = mhw.1@coef["a"] + mhw.1@coef["b"] * weight) %>%       # yhat for reg line
  select(weight, yhat, lb_ht, ub_ht)
```

```
## Warning: `as.tibble()` is deprecated, use `as_tibble()` (but mind the new semantics).
## This warning is displayed once per session.
```

```
kable(preds, type = "pandoc", caption = "!Kung Predicted Heights")
```

Table 1: !Kung Predicted Heights

| weight | yhat | lb_ht | ub_ht |
|---|---|---|---|
| 32.59 | 132.9363 | 117.5893 | 147.0679 |
| 43.72 | 152.5704 | 137.6683 | 167.1974 |
| 46.95 | 158.2684 | 143.2099 | 173.5524 |
| 54.63 | 171.8164 | 156.3160 | 186.5088 |
| 64.78 | 189.7218 | 174.6773 | 204.7158 |

**4H2**

a)

```
#filter data to only children
d3 <- d %>% filter(age < 18)

#fit model
mhw.2 <- map(alist(
```

```
            height ~ dnorm(mu, sigma),
            mu <- a + b*weight,
            a ~ dnorm(156, 100),
            b ~ dnorm(0, 10),
            sigma ~ dunif(0, 50)
           ),
        data = d3)

#summary call for what's in the model
precis(mhw.2)
```

```
##         Mean StdDev  5.5% 94.5%
## a     58.25   1.40 56.02 60.48
## b      2.72   0.07  2.61  2.83
## sigma  8.43   0.43  7.75  9.12
```

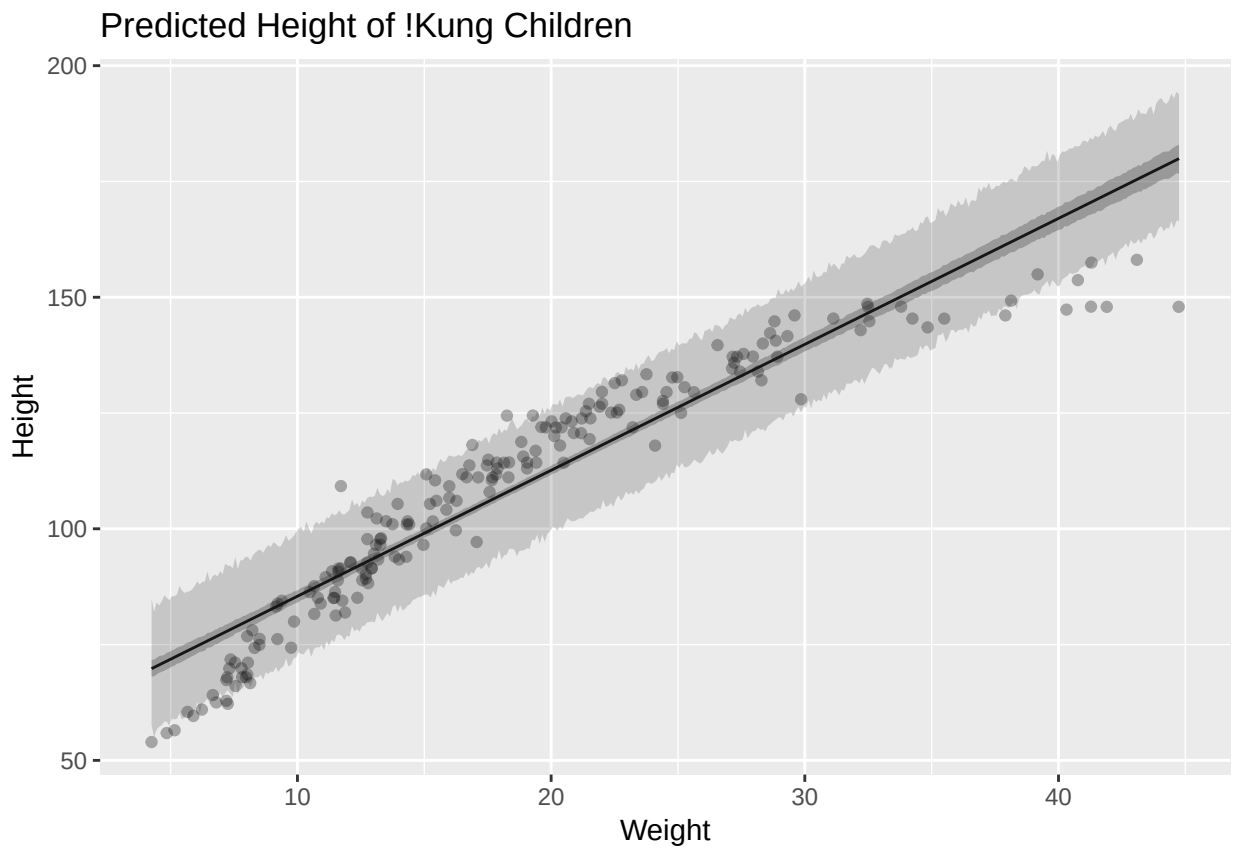
For every 10 units increase in weight, the model predicts that a child will get 27.2 cm taller.

b)

```
#steve's code
N <- 1e6 # sample size

# Get predictive means and data
preds <-
  as.tibble(MASS::mvrnorm(mu = mhw.2@coef,
                          Sigma = mhw.2@vcov , n = N )) %>%      # rather than extract.samples
  mutate(weight = sample(seq(from = 4.25, to = 44.75, by = 0.1), N, replace = T),
         predmean = a + b * weight ,                            # line uncertainty
         predverb = rnorm(N, a + b*weight, sigma )) %>%         # data uncertainty
  group_by(weight) %>%
  mutate(lb_mu = rethinking::HPDI(predmean, prob = .89)[1],
         ub_mu = rethinking::HPDI(predmean, prob = .89)[2],
         lb_ht = rethinking::HPDI(predverb, prob = .89)[1],
         ub_ht = rethinking::HPDI(predverb, prob = .89)[2]) %>%
  slice(1) %>%
  mutate(yhat = mhw.2@coef["a"] + mhw.2@coef["b"] * weight) %>%       # yhat for reg line
  select(weight, yhat, lb_mu, ub_mu, lb_ht, ub_ht)

#plot
ggplot(d3, aes(x = weight)) +
  geom_jitter(aes(y = height), alpha = .3) +
  geom_line(data = preds, aes(y = yhat)) +
  geom_ribbon(data = preds, aes(ymin = lb_mu, ymax = ub_mu), alpha = .3) +
  geom_ribbon(data = preds, aes(ymin = lb_ht, ymax = ub_ht), alpha = .2) +
  labs(x = "Weight",
       y = "Height",
       title = "Predicted Height of !Kung Children")
```




c)

- seems curvilinear, maybe should add a squared term

**4H3**

```
#add variable of log weight
d3 <- d3 %>% mutate(logweight = log(weight))

#fit model
mhw.3 <- map(alist(
             height ~ dnorm(mu, sigma),
             mu <- a + b*logweight,
             a ~ dnorm(178, 100),
             b ~ dnorm(0, 10),
             sigma ~ dunif(0, 50)
            ),
         data = d3)

#summary call for what's in the model
precis(mhw.3)
```

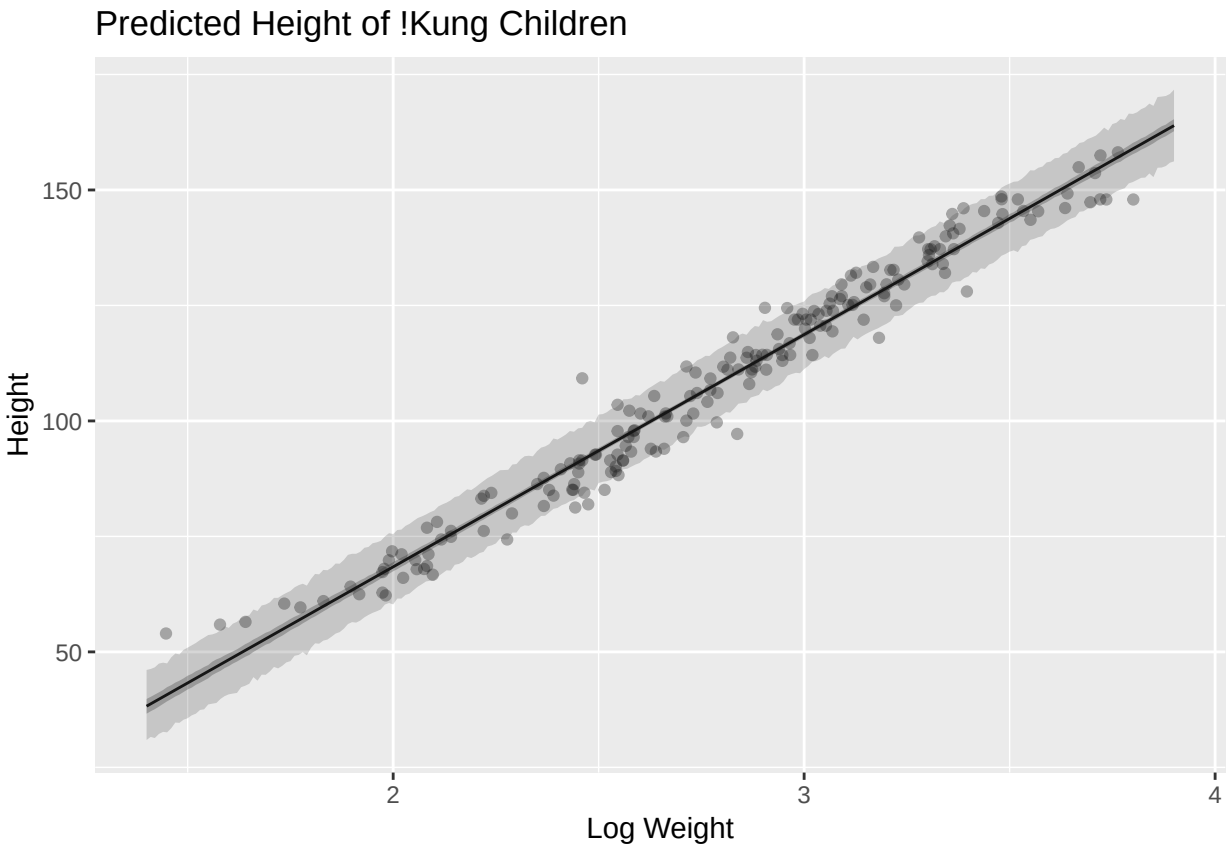
```
##          Mean StdDev   5.5%  94.5%
## a     -32.14   1.91 -35.19 -29.10
## b      50.28   0.67  49.20  51.35
## sigma   4.66   0.24   4.28   5.04
```

b)

```
#steve's code
N <- 1e6 # sample size

# Get predictive means and data
preds <-
  as.tibble(MASS::mvrnorm(mu = mhw.3@coef,
                          Sigma = mhw.3@vcov , n = N )) %>%      # rather than extract.samples
  mutate(logweight = sample(seq(from = 1.4, to = 3.9, by = 0.01), N, replace = T),
         predmean = a + b * logweight,                           # line uncertainty
         predverb = rnorm(N, a + b*logweight, sigma )) %>%        # data uncertainty
  group_by(logweight) %>%
  mutate(lb_mu = rethinking::HPDI(predmean, prob = .89)[1],
         ub_mu = rethinking::HPDI(predmean, prob = .89)[2],
         lb_ht = rethinking::HPDI(predverb, prob = .89)[1],
         ub_ht = rethinking::HPDI(predverb, prob = .89)[2]) %>%
  slice(1) %>%
  mutate(yhat = mhw.3@coef["a"] + mhw.3@coef["b"] * logweight) %>%       # yhat for reg line
  select(logweight, yhat, lb_mu, ub_mu, lb_ht, ub_ht)

#plot
ggplot(data = d3, aes(x = logweight)) +
  geom_jitter(aes(y = height), alpha = .3) +
  geom_line(data = preds, aes(y = yhat)) +
  geom_ribbon(data = preds, aes(ymin = lb_mu, ymax = ub_mu), alpha = .3) +
  geom_ribbon(data = preds, aes(ymin = lb_ht, ymax = ub_ht), alpha = .2) +
  labs(x = "Log Weight",
       y = "Height",
       title = "Predicted Height of !Kung Children")
```


Predicted Height of !Kung Children
Height
150
100
50
2
3
4
Log Weight