Assignment 3 - Height and Weight - Quadratic Approximation

Wen-Shiuan, Liang

10/14/2017

Question 1

By using the Howell1 data mentioned in the class, the weights listed below were recorded in the !Kung census, but heights were not recorded for these individuals. Provide predicted heights and 89% intervals (either HPDI or PI) for each of these individuals. That is, fill in the table below, using model-based predictions.

Import the data and build the model

```
library(rethinking)
```

```
data("Howell1")
data = Howell1
model1 = map(
  alist(
    height ~ dnorm(mu,sigma),
    mu <- a + b*weight,
    a ~ dnorm(156,100),
    b ~ dnorm(0,10),
    sigma ~ dunif(0,50)
    ),
  data = data)
precis(model1)
```

```
##        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
```

Set Posterior

```
post = extract.samples(model1)
post[1:5,]
```

```
##          a        b    sigma
## 1 75.62846 1.765557 9.696083
## 2 75.44354 1.767446 9.042853
## 3 74.33777 1.798669 8.913366
```

```
## 4 76.45392 1.740638 9.631927
## 5 74.78684 1.773455 9.407015
```

Predict the height for 46.95, 43.72, 64.78, 32.59, 54.63

```
mu.link = function(weight) post$a + post$b*weight
weight.seq = c(46.95, 43.72, 64.78, 32.59, 54.63)
mu = sapply(weight.seq, mu.link)
```

Expected height, 89% interval are as follow

```
(mu.mean = apply(mu,2,mean))

## [1] 158.2719 152.5743 189.7232 132.9415 171.8191

(mu.HPDI = apply(mu,2,HPDI))

##            [,1]     [,2]     [,3]     [,4]     [,5]
## |0.89 157.4694 151.8594 188.3150 132.3165 170.7662
## 0.89| 159.0858 153.3185 191.1561 133.6262 172.8544

(mu.PI = apply(mu,2,PI))

##          [,1]     [,2]     [,3]     [,4]     [,5]
## 5%  157.4610 151.8499 188.3079 132.2801 170.7733
## 94% 159.0807 153.3120 191.1496 133.5927 172.8660
```

Question 2

Select out all the rows in the Howell1 data with ages below 18 years of age. If you do it right, you should end up with a new data frame with 192 rows in it.

```
d2 = data[data$age<18,]
```

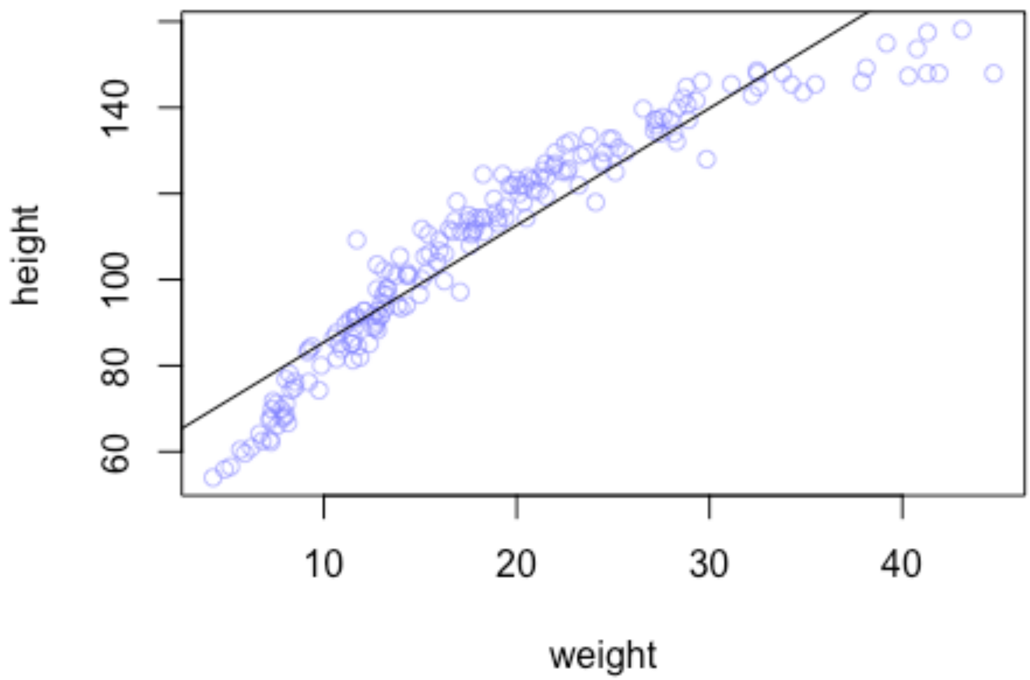
(a) Fit a linear regression to these data, using map. Present and interpret the estimates. For every 10 units of increase in weight, how much taller does the model predict a child gets?

Orignal Linear Regression is as follow(only change data)

```
model1 = map(
  alist(
    height ~ dnorm(mu,sigma),
    mu <- a + b*weight,
    a ~ dnorm(156,100),
    b ~ dnorm(0,10),
    sigma ~ dunif(0,50)
    ),
  data = d2)
precis(model1)

##         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  8.44   0.43  7.75  9.13
```

```
post = extract.samples(model1)
mu.link = function(weight) post$a + post$b*weight
weight.seq = seq( from = 2, to = 70, by = 1)
mu = sapply(weight.seq, mu.link)
mu.mean = apply(mu,2,mean)
plot(height ~ weight, data = d2, col = col.alpha(rangi2,0.5))
lines(weight.seq,mu.mean)
```




```
model2 = map(
  alist(
    height ~ dnorm(mu,sigma),
    mu <- a + b*weight,
    a ~ dnorm(110,100),
    b ~ dnorm(0,10),
    sigma ~ dunif(0,50)
    ),
  data = d2)
precis(model2)
```

```
##         Mean StdDev  5.5% 94.5%
## a      58.24   1.40 56.01 60.47
```

```
## b       2.72   0.07  2.61  2.83
## sigma   8.43   0.43  7.75  9.12
```

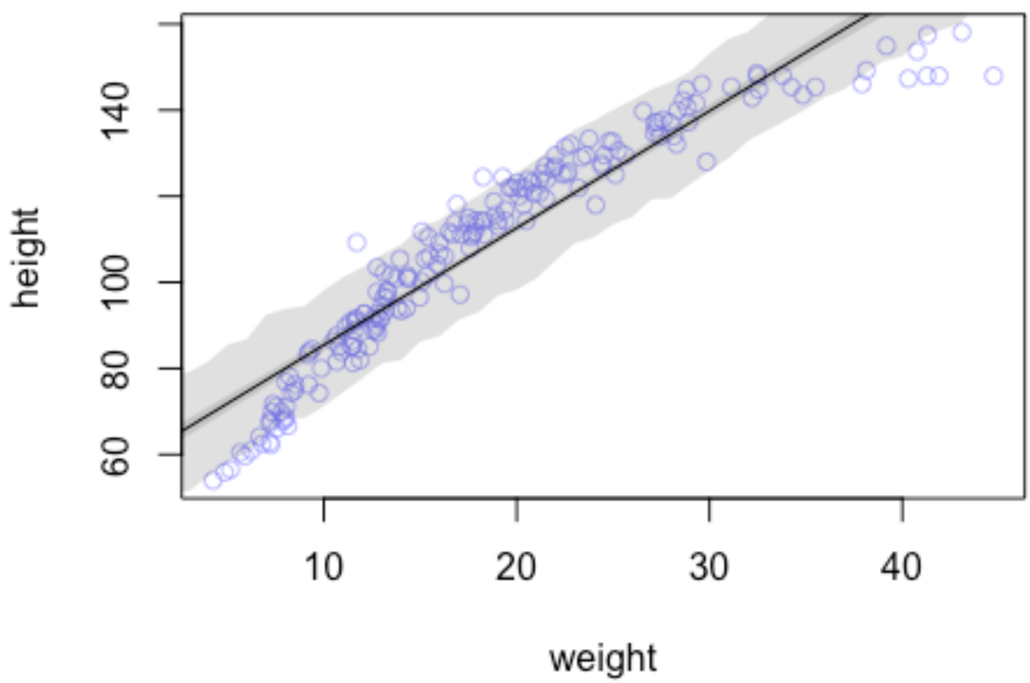
It seems that there is no change in beta between model1 and model2, with changing the dnorm parameters.

The linear regression model predict that every 10 addtional kg would increase 27 cm in height

(b) Plot the raw data, with height on the vertical axis and weight on the horizontal axis. Superimpose the MAP regression line and 89% HPDI for the mean. Also superimpose the 89% HPDI for predicted heights.

take a look at the plot

```
post2 = extract.samples(model2)
mu.link2 = function(weight) post2$a + post2$b*weight
weight.seq2 = seq( from = 2, to = 70, by = 1)
mu2 = sapply(weight.seq2, mu.link2)
mu.mean2 = apply(mu2,2,mean)
#The MAP line
plot(height ~ weight, data = d2, col = col.alpha(rangi2,0.5))
lines(weight.seq2,mu.mean2)

#Plot a shaded region for 89% HPDI
mu.HPDI2 = apply(mu2, 2, HPDI)
shade(mu.HPDI2, weight.seq2)

# superimpose the 89% HPDI for predicted heights.
sim.height = sim(model2, data = list(weight = weight.seq2))

## [ 100 / 1000 ]
[ 200 / 1000 ]
[ 300 / 1000 ]
[ 400 / 1000 ]
[ 500 / 1000 ]
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
[ 900 / 1000 ]
[ 1000 / 1000 ]

height.HPDI = apply(sim.height, 2, HPDI)
shade(height.HPDI, weight.seq2)
```


height
60
80
100
140
10
20
30
40
weight

(c) What aspects of the model fit concern you? Describe the kinds of assumption you would change, if any, to improve your model. You don't have to write any new code. Just explain what the model appears to be doing a bad job of, and what you hypothesize would be a better model.

The model predict badly those children who weights more than 35 kg, yet it does well between the weight from 8 - 35. From question 1 we can see that: It seems !Kung people are not likely to grow higher than 160 cm. That's why the gain in weight could not imply higher height after 160 cm. So a linear model could not fit well.

Base on this, I would rule out the out those childern who gain more than 35 kg and re-run the model.

Question 3

Suppose a colleague of yours, who works on allometry, glances at the practice problems just above. (In Question 2). You colleague exclaims, "That's silly. Everyone knows that it's only the logarithm of body weight that scales with height!" Let's take your colleague's advice and see what happens.

(a) Model the relationship between height (cm) and the natural logarithm of weight (log-kg). Use the entire Howell1 data frame, all 544 rows, adults and non-adults. Fit this model, using quadratic approximation:h ~ (,)

= + log( ) (178,100) (0,100) ~ (0,50) ####where h is the height of individual and is the weight (in kg) of individual . The function for computing a natural log in R is just log. Can you interpret the resulting estimates?

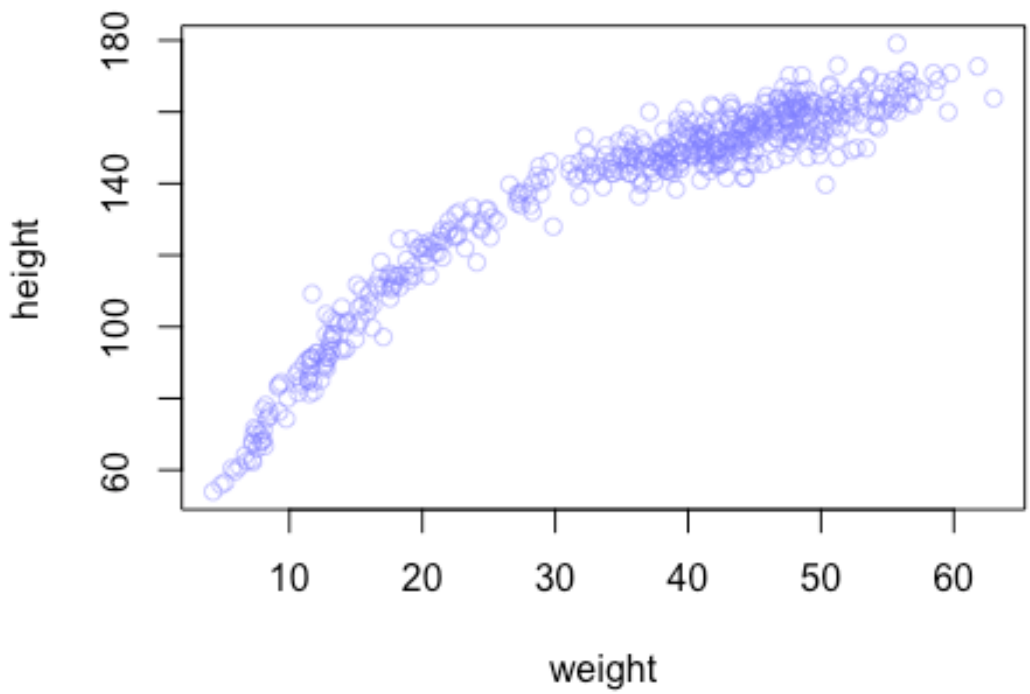
```
model3 = map(
  alist(
    height ~ dnorm(mu,sigma),
    mu <- a + b*log(weight),
    a ~ dnorm(178,100),
    b ~ dnorm(0,100),
    sigma ~ dunif(0,50)
    ),
  data = data)
precis(model3)

##          Mean StdDev   5.5%  94.5%
## a      -23.78   1.34 -25.92 -21.65
## b       47.08   0.38  46.46  47.69
## sigma    5.13   0.16   4.89   5.38
```

Beta has larger because of log, while the interpretation would change. In this case, the model predicts that 1% gained in weight would increase 0.47% height.

With reference from this site:http://www.cazaar.com/ta/econ113/interpreting-beta "If we increase x by one percent, we expect y to increase by (β1/100) units of y."

(b) Begin with this plot:

```
plot(height~weight, data=Howell1, col=col.alpha(rangi2, 0.4))
```



Then use samples from the quadratic approximate posterior of the model in (a) to superimpose on the plot: (1) the predicted mean height as a function of weight, (2) the 97% HPDI for the mean, and (3) the 97% HPDI for predicted heights.

```
post3 = extract.samples(model3)
mu.link3 = function(weight) post3$a + post3$b*weight
weight.seq3 = seq( from = 2, to = 70, by = 1)
mu3 = sapply(weight.seq3, mu.link2)
mu.mean3 = apply(mu3,2,mean)
#The MAP line
plot(height~weight, data=Howell1, col=col.alpha(rangi2, 0.4))
lines(weight.seq3,mu.mean3)

#Plot a shaded region for 97% HPDI
mu.HPDI3 = apply(mu3, 2, HPDI,prob = .97)
```

```
shade(mu.HPDI3, weight.seq3)

# superimpose the 97% HPDI for predicted heights.
sim.height3 = sim(model3, data = list(weight = weight.seq3))

## [ 100 / 1000 ]
[ 200 / 1000 ]
[ 300 / 1000 ]
[ 400 / 1000 ]
[ 500 / 1000 ]
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
[ 900 / 1000 ]
[ 1000 / 1000 ]

height.HPDI3 = apply(sim.height3, 2, HPDI,prob = .97)
shade(height.HPDI3, weight.seq3)
```