HW3

Matt Kaye

9/27/2018

4H1

```
##   individual weight expected.height           hpdi.89
## 1          1  46.95        156.3699 155.912 - 156.809
## 2          2  43.72        153.4482 152.996 - 153.872
## 3          3  64.78        172.4978 171.094 - 173.885
## 4          4  32.59        143.3807 142.466 - 144.327
## 5          5  54.63        163.3167 162.541 - 164.087
```

4H2

a

```r
rm(list = ls())
data("Howell1")
df <- Howell1 %>%
  filter(age < 18)


flist <- alist(
    height ~ dnorm(mu , sigma),
    mu <- a + b*weight,
    a ~ dnorm(100, 50),
    b ~ dnorm(0, 10),
    sigma ~ dunif(0, 50)
)

mod <- rethinking::map(flist, data = df)
precis(mod)
```

```
##        Mean StdDev  5.5% 94.5%
## a     58.27   1.40 56.03 60.50
## b      2.72   0.07  2.61  2.83
## sigma  8.44   0.43  7.75  9.13
```

The coefficients produced by this model are above. The intercept coefficient a is not particularly relevant, but implies that a weightless child is 58.27 units of height tall, on average. The slope coefficient b means that for every unit increase in weight, we expect a 2.72 unit increase in height, on average. For every 10 units increase in weight, the model predicts that a child gets 27.2 units of height taller, on average. The sigma coefficient is the expected standard deviation of our height estimates.

I chose the values for my hyperparameters in a way so that they made sense in the context of the problem. I figured that the height of a child is, on average, smaller than the height of an adult at all levels of weight, so I decreased the mean of the distribution of a to 100cm. I also assumed that the relationships between height and weight among children relative ot adults would be about the same, so I left the parameters for the distribution of b the same.

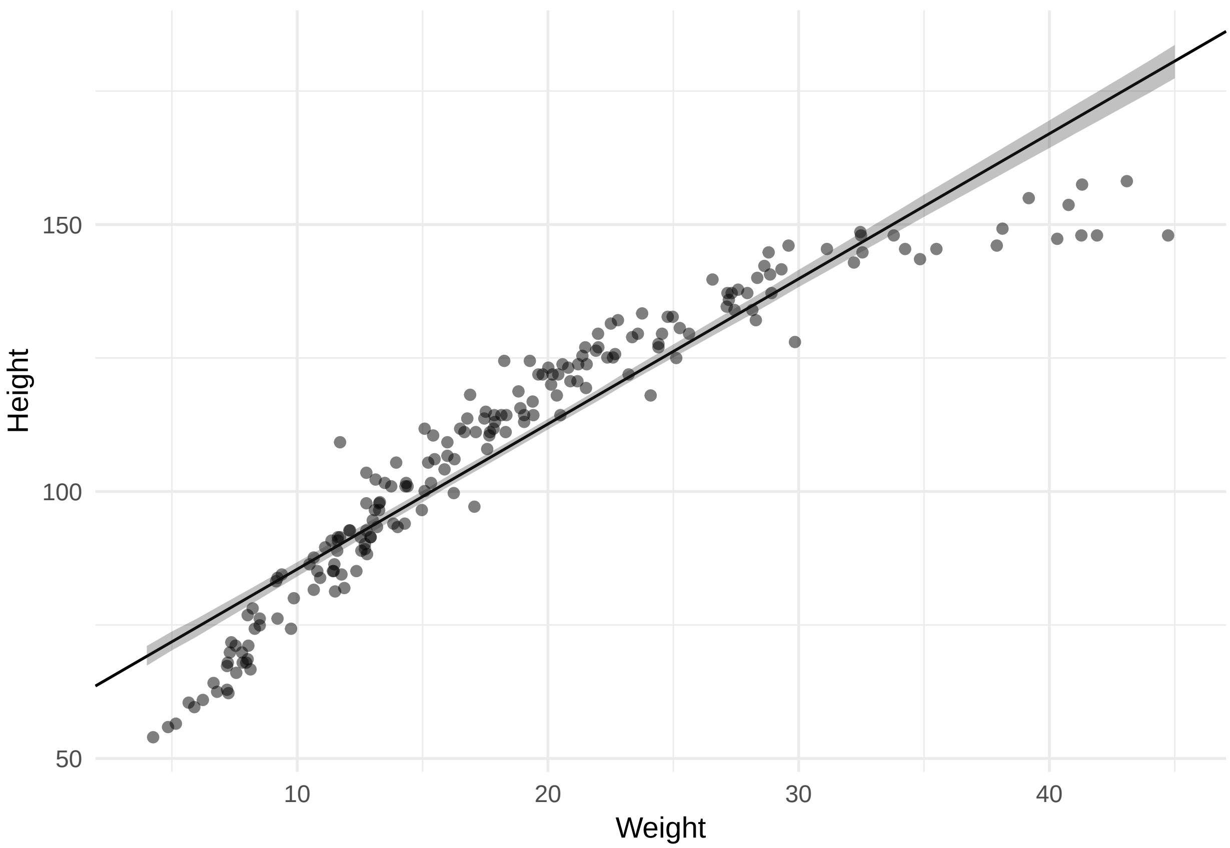
b

```
weights <- data.frame(weight = seq(4, 45, by = 1))
mu <- link(mod,
           data = weights, refresh = 0) %>%
  data.frame()

mu.bands <- cbind(
  weight = weights,
  mu.mean = apply(mu, 2, mean),
  mu.hpdi = t(apply(mu, 2, HPDI))
)



ggplot(data=df)+
  geom_point(mapping=aes(x=weight,y=height),alpha=0.5)+
  geom_abline(intercept=coef(mod)[1],slope=coef(mod)[2])+
  geom_ribbon(data = mu.bands, mapping=aes(x=weight,ymin=`mu.hpdi.|0.89`,ymax=`mu.hpdi.0.89|`),alpha=0.
  labs(x="Weight",y="Height")+
  theme_minimal()
```



c

I think that the linearity assumption in the data is a little bit questionable. It seems like a log or quadratic model might fit a little bit better here becuase of the apparent curvature in the data, and might end up produced results that fit the data better. Additionally, it is possible that the relationship between height and weight is not the same for the adult group as the child group, which means that we might want to adjust the expected value of our b parameter.

4H3

```r
rm(list = ls())
data("Howell1")
all.data <- Howell1

flist <- alist(
    height ~ dnorm(mu , sigma),
    mu <- a + b*log(weight),
    a ~ dnorm(178, 100),
    b ~ dnorm(0, 100),
    sigma ~ dunif(0 , 50)
)

mod <- rethinking::map(flist, data = all.data)
precis(mod)
```

```
##         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
```

These results look better than the ones from the previous part where we used a linear model, and we can see in the plot below that the log model seems to fit the data significantly better. The negative a value is not particularly important, since we do not expect to have any weightless babies. The b value means that for every one percent increase in weight, we expect a .4707 unit increase in height. Again, the sigma value is the expected standard deviation of our height predictions.

b

```r
weights <- data.frame(weight = seq(4, 63, by = 1))
mu <- link(mod,
           data = weights, refresh = 0) %>%
  data.frame()

mu.bands <- cbind(
  weight = weights,
  mu.mean = apply(mu, 2, mean),
  mu.hpdi = t(apply(mu, 2, HPDI, p = .97))
)

ggplot(data=mu.bands)+
  geom_point(data = all.data, mapping=aes(x=weight,y=height),alpha=0.5)+
```

```
geom_ribbon(mapping=aes(x=weight,ymin=`mu.hpdi.|0.97`,ymax=`mu.hpdi.0.97|`, color = 'red'),alpha=0.3)
labs(x="Weight",y="Height") +
theme_minimal()
```