MBA5011 Multivariate Analysis: Model-Based Statistics

Assignment 2

Individual Assignment 2

Question 1 (20%). 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. (Hint:`library(rethinking);data(Howell1)`)

`answer.sheet`

```
##   Individual Weight Expected height 89% interval
## 1          1  46.95              NA           NA
## 2          2  43.72              NA           NA
## 3          3  64.78              NA           NA
## 4          4  32.59              NA           NA
## 5          5  54.63              NA           NA
```

Question 2 (40%). 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.

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

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

(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

Question 3 (40%). 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 MCMC (stan):

$$h_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta \log(w_i)$$

$$\alpha \sim Normal(178, 100)$$

$$\beta \sim Normal(0, 100)$$

$$\sigma \sim Uniform(0, 50)$$

where h_i is the height of individual i and w_i is the weight (in kg) of individual i The function for computing a natural log in R is just log. Can you interpret the resulting estimates?

(b) Begin with this plot:

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

Then use samples from the 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.