STAT-670-Assignment 5

Prashanth Thirukkurungudi Sekar (UID: 2000376633)

2/23/2019

Worked with Pravin Sundar and Barathwaaj Parthasarathy

Loading the data:

```
cols_req = c("BMI","Age","Height")
df = NHANES[,cols_req]
```

Summary Statistitcs :

```
summary(df)
```

```
##       BMI              Age              Height     
##  Min.   :12.88   Min.   : 0.00   Min.   : 83.6  
##  1st Qu.:21.58   1st Qu.:17.00   1st Qu.:156.8  
##  Median :25.98   Median :36.00   Median :166.0  
##  Mean   :26.66   Mean   :36.74   Mean   :161.9  
##  3rd Qu.:30.89   3rd Qu.:54.00   3rd Qu.:174.5  
##  Max.   :81.25   Max.   :80.00   Max.   :200.4  
##  NA's   :366                     NA's   :353
```

Removing missing value records:

```
df = df[complete.cases(df),]
df = data.frame(df)
df$Age = as.numeric(df$Age)
```

Question 1 : Modelling BMI with Age and Height

```
bmi.lo = loess(log10(BMI) ~ Height + Age , degree = 1,span=0.5, data = df)
```

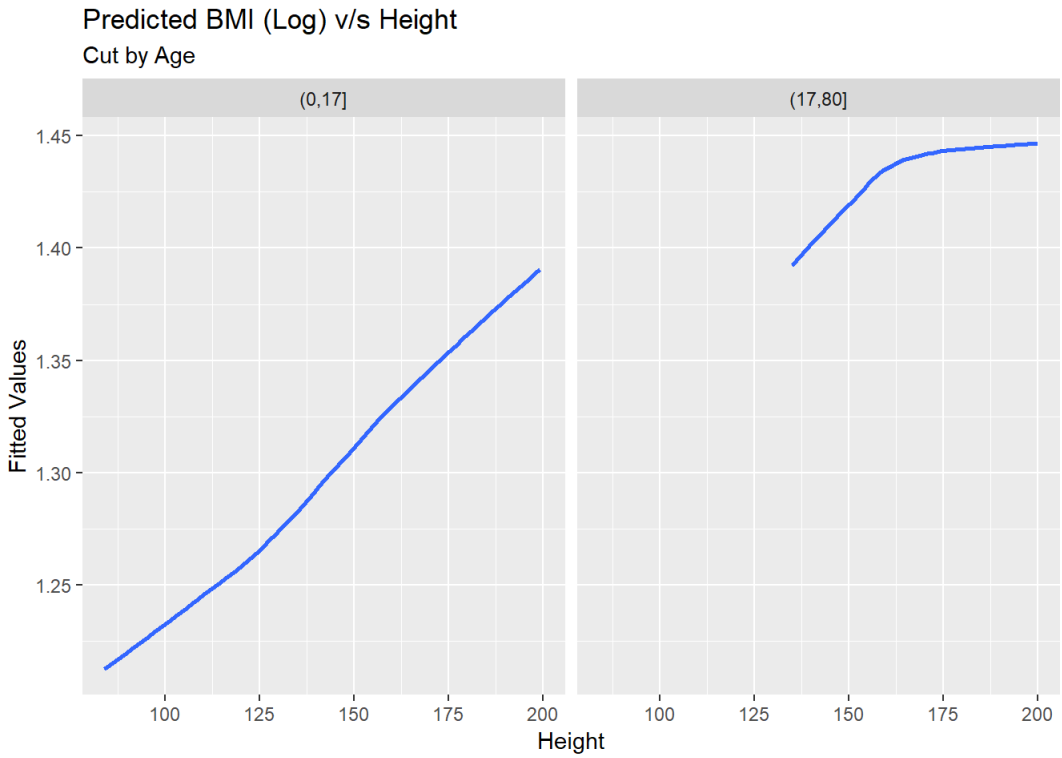
```
print(var(fitted.values(bmi.lo))/(var(fitted.values(bmi.lo)) + var(residuals(bmi.lo))))
```

```
## [1] 0.3834421
```

Question 2 - Variation of BMI and Height Facetted by Age

```
newdata1 = expand.grid(Height = seq(84, 200, 5),Age = seq(4,17,1)) #Age = c(4,17,40,60,80))
newdata2 = expand.grid(Height = seq(135, 200, 5),Age = seq(18,80,1)) #Age = c(4,17,40,60,80))

bmi.pred1 = predict(bmi.lo,newdata1)
bmi.pred2 = predict(bmi.lo,newdata2)
bmi.pred1 = data.frame(newdata1, fit = as.vector(bmi.pred1))
bmi.pred2 = data.frame(newdata2, fit = as.vector(bmi.pred2))
bmi.pred3 = rbind(bmi.pred1,bmi.pred2)
```

```
gg = ggplot(bmi.pred3, aes(x = Height,
y = fit)) + geom_smooth(method="loess",se=F,method.args = list(degree = 1)) + facet_grid(~cut(Age, c(0,17,8
0)))
gg + labs(title = "Predicted BMI (Log) v/s Height",x="Height",y="Fitted Values",subtitle="Cut by Age")
```



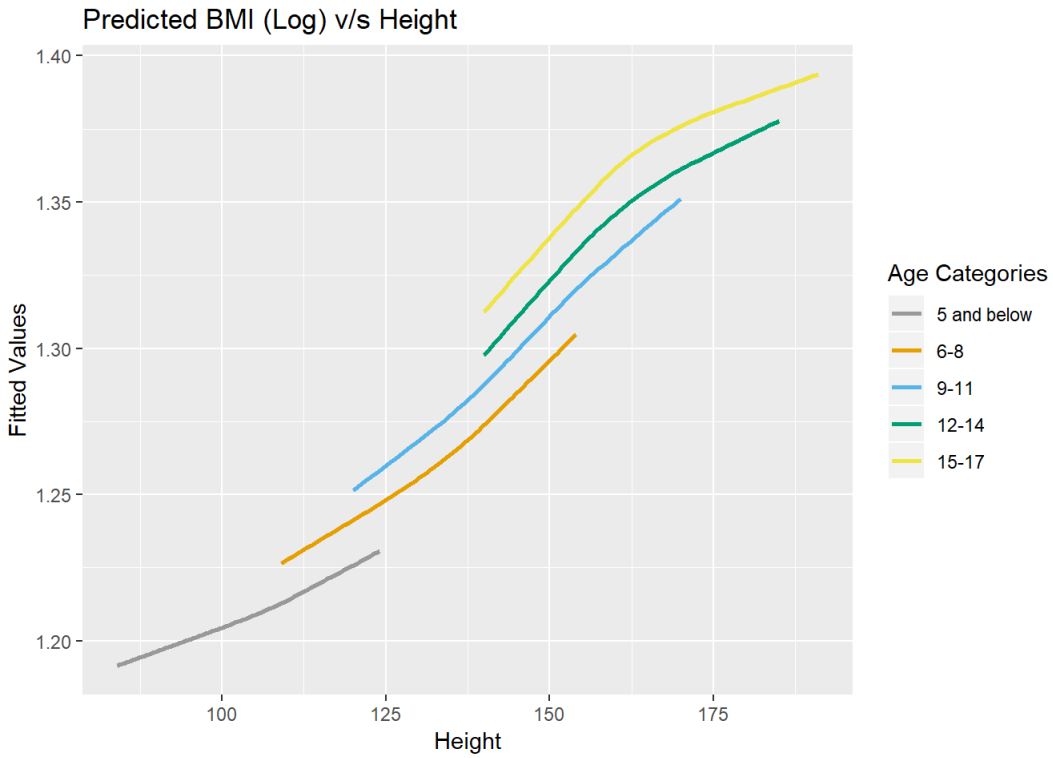
Question 3 - Variation of BMI with Height by groups of Age

```
newdata1 = expand.grid(Height = seq(84, 128, 5),Age = seq(2,5,1))
newdata2 = expand.grid(Height = seq(109, 156, 5),Age = seq(6,8,1))
newdata3 = expand.grid(Height = seq(120, 174, 5),Age = seq(9,11,1))
newdata4 = expand.grid(Height = seq(140, 187, 5),Age = seq(12,14,1))
newdata5 = expand.grid(Height = seq(146, 194, 5),Age = seq(15,17,1))


bmi.pred1 = predict(bmi.lo, newdata = newdata1)
bmi.pred2 = predict(bmi.lo, newdata = newdata2)
bmi.pred3 = predict(bmi.lo, newdata = newdata3)
bmi.pred4 = predict(bmi.lo, newdata = newdata4)
bmi.pred5 = predict(bmi.lo, newdata = newdata5)

bmi.pred1 = data.frame(newdata1, fit = as.vector(bmi.pred1))
bmi.pred2 = data.frame(newdata2, fit = as.vector(bmi.pred2))
bmi.pred3 = data.frame(newdata3, fit = as.vector(bmi.pred3))
bmi.pred4 = data.frame(newdata4, fit = as.vector(bmi.pred4))
bmi.pred5 = data.frame(newdata5, fit = as.vector(bmi.pred5))

bmi.pred6 = rbind(bmi.pred1,bmi.pred2,bmi.pred3,bmi.pred4,bmi.pred5)
```

```
bmi.pred6$Age_Group = ifelse(bmi.pred6$Age<=5, "5 and below",ifelse(bmi.pred6$Age>5 & bmi.pred6$Age<=8, "6-8
",ifelse(bmi.pred6$Age>8 & bmi.pred6$Age<=11,"9-11",ifelse(bmi.pred6$Age>11 & bmi.pred6$Age<14,"12-14","15-1
7"))))

neworder <- c("5 and below","6-8","9-11","12-14","15-17")
bmi.pred6 <- arrange(mutate(bmi.pred6,
            Age_Group=factor(Age_Group,levels=neworder)),Age_Group)
```

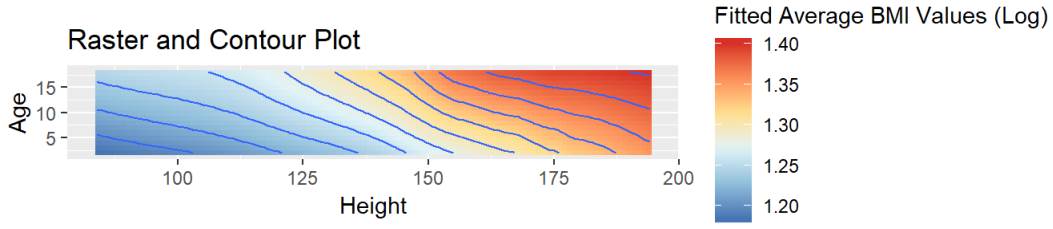
```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

```
legend_title <- "Age Categories"
ggplot(bmi.pred6, aes(x = Height, y = fit, group = Age_Group, color = factor(Age_Group))) + geom_smooth(meth
od="loess",se=F,method.args = list(degree = 1)) + scale_color_manual(legend_title,values = cb_palette)+ labs
(title = "Predicted BMI (Log) v/s Height",x="Height",y="Fitted Values")
```



Question 4 - Contour Plot for variation of Avg BMI with Height and Age:

```
newdata1 = expand.grid(Height = seq(84,194, 1), Age = seq(2,18,1))#Age = c(4,17,40,60,80))
bmi.pred1 = predict(bmi.lo, newdata = newdata1)
bmi.pred1 = data.frame(newdata1, predicted_log10_bmi = as.vector(bmi.pred1))
legend_title <- "Fitted Average BMI Values (Log)"
ggplot(bmi.pred1, aes(x = Height, y = Age, z = predicted_log10_bmi)) + geom_raster(aes(fill = predicted_log1
0_bmi)) +coord_fixed() +  scale_fill_distiller(legend_title,palette = "RdYlBu") + geom_contour() +  labs(tit
le="Raster and Contour Plot",x="Height",y="Age")
```



Question 5 - Research Questions :

How does average BMI vary with height and age for children 17 and under?

We will be considering the first two graphs for this research question.
From the first graph, we can observe that for children (Age 17 and under), there is an evident relationship between height and average BMI. It can be seen that the avgerage BMI increases as height increases.
Next, let's consider the second graph to explore the variation of average BMI with age. We can observe that there is pretty evident variation between BMI and age. Average BMI increases steadily (along with increase in height) as children grow older.Among the different children groups that have been chosen, we can see that the age groups 6-8 and 9-11 have the steepest increase of average BMI with respect to height along with the widest range of average BMI. Finally, we can tell that children below age 5 and between age 12-17 do not show such a steep change in average BMI that was observed in the former 2 categories (6-8 and 9-11) but still there is an increase in avg. BMI as age grows.

Is BMI a good measure of body mass for children?

BMI being defined as a measure of weight (in Kg) divided by height squared (in m) is not a good measure of body mass for children. Through the exploratory analysis performed on the NHANES data, we can observe that average BMI for children (under the age of 17) varies with both height and age. It can be seen that for children in the age group of 2-5, the average BMI varies in the range of 13-29 which actually does not correspond to it's relative body mass. Finally, a BMI calculation for children with height and weight alone will be as accurate as they are for adults (Age 17+) since it is evident that children's body fat percentage varies/ changes as they grow and their BMI varies based on their age too and this trend can be observed from the second and third graph.