LMM with B-Splines

Alexander McLain

Contents

The Six Cities Study of Air Pollution and Health example (see the first R notes for details).

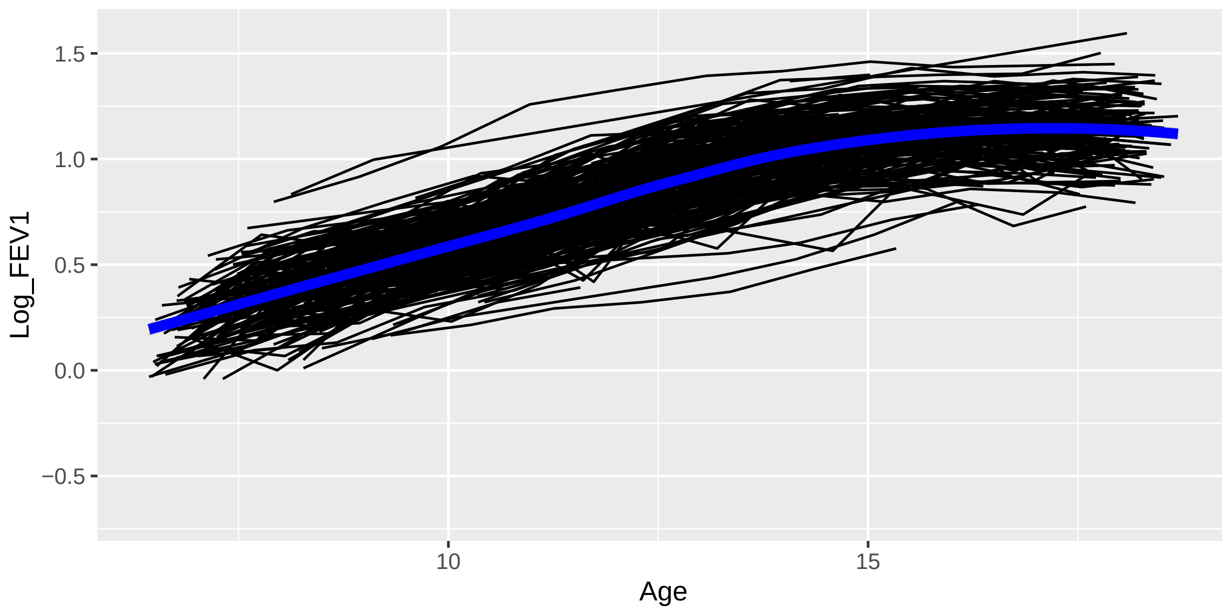
```r
Six_cities <- read.csv("Six_cities.csv", header = TRUE)
str(Six_cities)
```

```
## 'data.frame':    1994 obs. of  6 variables:
##  $ ID        : int  1 1 1 1 1 1 1 2 2 2 ...
##  $ Height    : num  1.2 1.28 1.33 1.42 1.48 1.5 1.52 1.13 1.19 1.49 ...
##  $ Age       : num  9.34 10.39 11.45 12.46 13.42 ...
##  $ INI_Height: num  1.2 1.2 1.2 1.2 1.2 1.2 1.2 1.13 1.13 1.13 ...
##  $ INI_Age   : num  9.34 9.34 9.34 9.34 9.34 ...
##  $ Log_FEV1  : num  0.215 0.372 0.489 0.751 0.833 ...
```

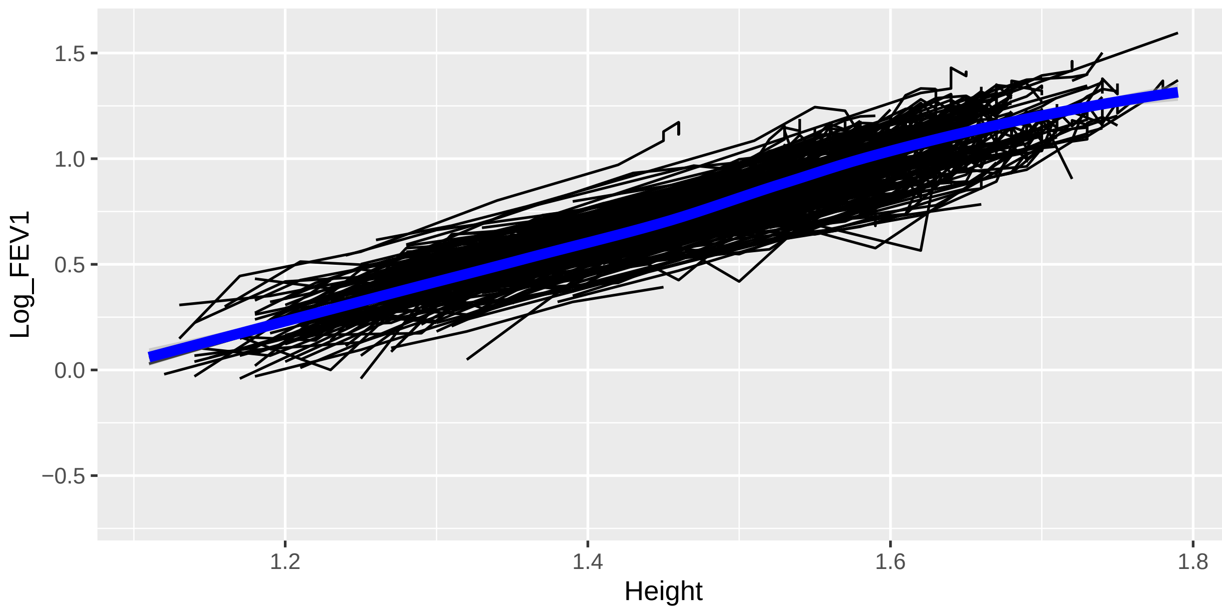
```r
tail(Six_cities,8)
```

	ID	Height	Age	INI_Height	INI_Age	Log_FEV1
1987	299	1.64	17.9904	1.57	12.9555	1.09527
1988	300	1.44	11.9617	1.44	11.9617	0.68310
1989	300	1.50	12.9993	1.44	11.9617	0.85015
1990	300	1.57	13.9055	1.44	11.9617	0.81536
1991	300	1.61	14.9596	1.44	11.9617	1.11841
1992	300	1.62	15.9398	1.44	11.9617	1.08181
1993	300	1.62	17.0075	1.44	11.9617	1.12817
1994	300	1.63	17.8645	1.44	11.9617	1.16938

```r
library(tidyverse)
Six_cities <- Six_cities %>% mutate( ID = as.factor(ID) )

### Let's look at the trends by age and height at measurement ###
par(mfrow = c(1,2))
p <- ggplot(Six_cities, aes(x = Age, y = Log_FEV1, group = ID))
p + geom_line() +
  geom_smooth(aes(group = 1), method = "loess",
                              color = "blue", size = 2)
```




```r
p <- ggplot(Six_cities, aes(x = Height, y = Log_FEV1, group = ID))
p + geom_line() +
  geom_smooth(aes(group = 1), method = "loess",
                              color = "blue", size = 2)
```



We'll fit this using cubic B-splines for Age, since it appears to have a non-linear relationship.

```r
library(lme4)
library(lmerTest)
library(splines)
#-------------------------------------------------------
# 1) Fit a linear mixed model using B-splines for 'age'
#-------------------------------------------------------
```

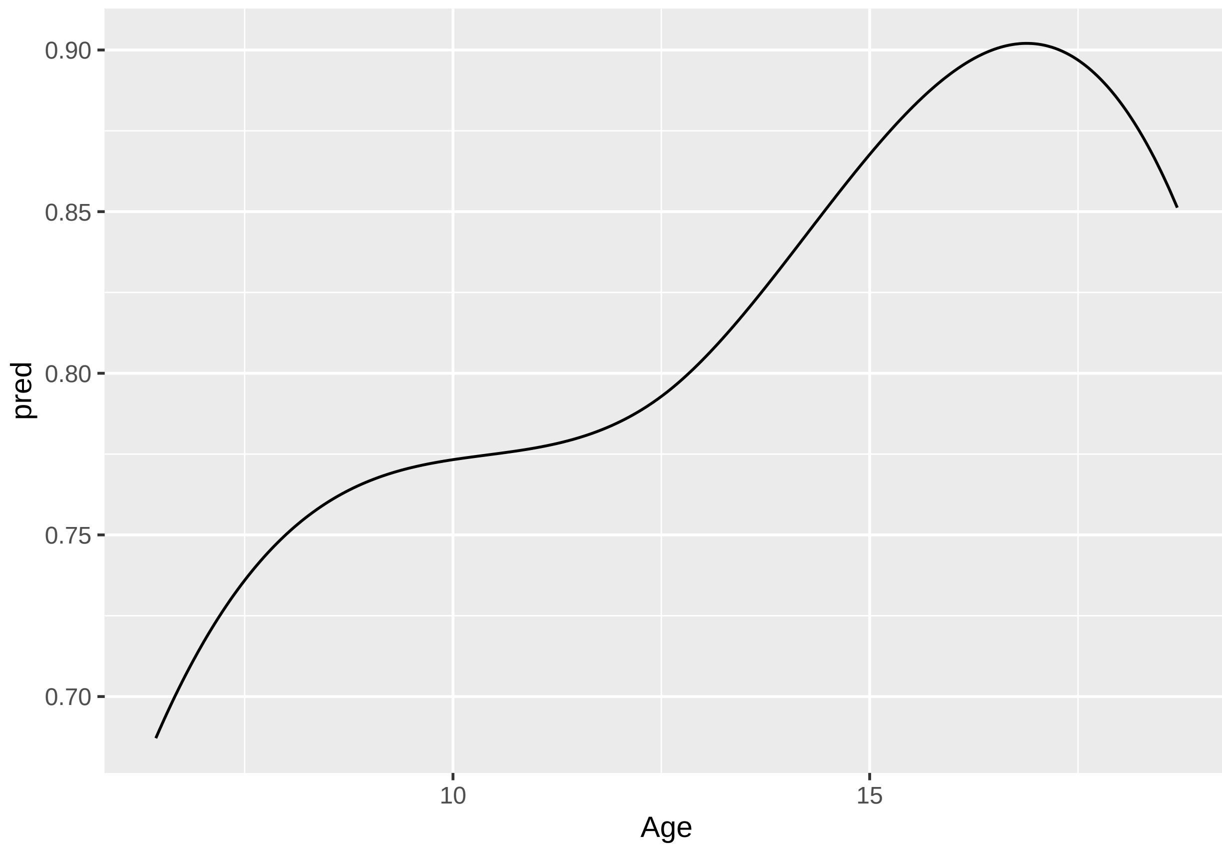
```r
# The 'bs()' function creates B-spline basis functions for the 'time' variable.
# You can adjust 'df' (degrees of freedom) or specify knots directly.

LMM_formula <- Log_FEV1 ~ Height + bs(Age, df = 4) + (1 + Height|ID)
LMM_int_slp <- lmer( formula = LMM_formula , data = Six_cities)
summary(LMM_int_slp)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: LMM_formula
##    Data: Six_cities
##
## REML criterion at convergence: -4696.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.4252 -0.4730  0.0657  0.5655  2.8502
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  ID       (Intercept) 0.095711 0.3094
##           Height      0.043392 0.2083   -0.94
##  Residual             0.003114 0.0558
## Number of obs: 1994, groups:  ID, 300
##
## Fixed effects:
##                     Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)       -1.881e+00  7.223e-02  1.349e+03 -26.036  < 2e-16 ***
## Height             1.715e+00  5.868e-02  1.475e+03  29.222  < 2e-16 ***
## bs(Age, df = 4)1   1.223e-01  1.711e-02  1.726e+03   7.152 1.26e-12 ***
## bs(Age, df = 4)2   8.666e-03  2.820e-02  1.789e+03   0.307    0.759
## bs(Age, df = 4)3   2.890e-01  2.856e-02  1.928e+03  10.121  < 2e-16 ***
## bs(Age, df = 4)4   1.641e-01  2.697e-02  1.841e+03   6.085 1.41e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Height b(A,d=4)1 b(A,d=4)2 b(A,d=4)3
## Height      -0.987
## bs(Ag,d=4)1  0.017 -0.140
## bs(Ag,d=4)2  0.856 -0.893  0.155
## bs(Ag,d=4)3  0.760 -0.833  0.571     0.765
## bs(Ag,d=4)4  0.823 -0.881  0.376     0.926     0.856
```

Now, let's plot the results

```r
#------------------------------------------------------
# 2) Plot the fitted B-spline curves
#------------------------------------------------------



# Create a small grid of age points for prediction
new_age <- data.frame(Age = sort(unique(Six_cities$Age)), Height = mean(Six_cities$Height))

# Predict using the fixed effects (excluding random intercepts)
```

```r
# If you want subject-specific predictions, include `re.form = NULL` or specify subject ID
new_age$pred <- predict(LMM_int_slp, newdata = new_age, re.form = NA)

# Plot
ggplot(new_age, aes(x = Age, y = pred)) + geom_line()
```



This plot gives you an idea of how the mean response (averaging across random intercepts) changes over age according to the B-spline basis.