Fixed Effects Models

Alexander McLain

Contents

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

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

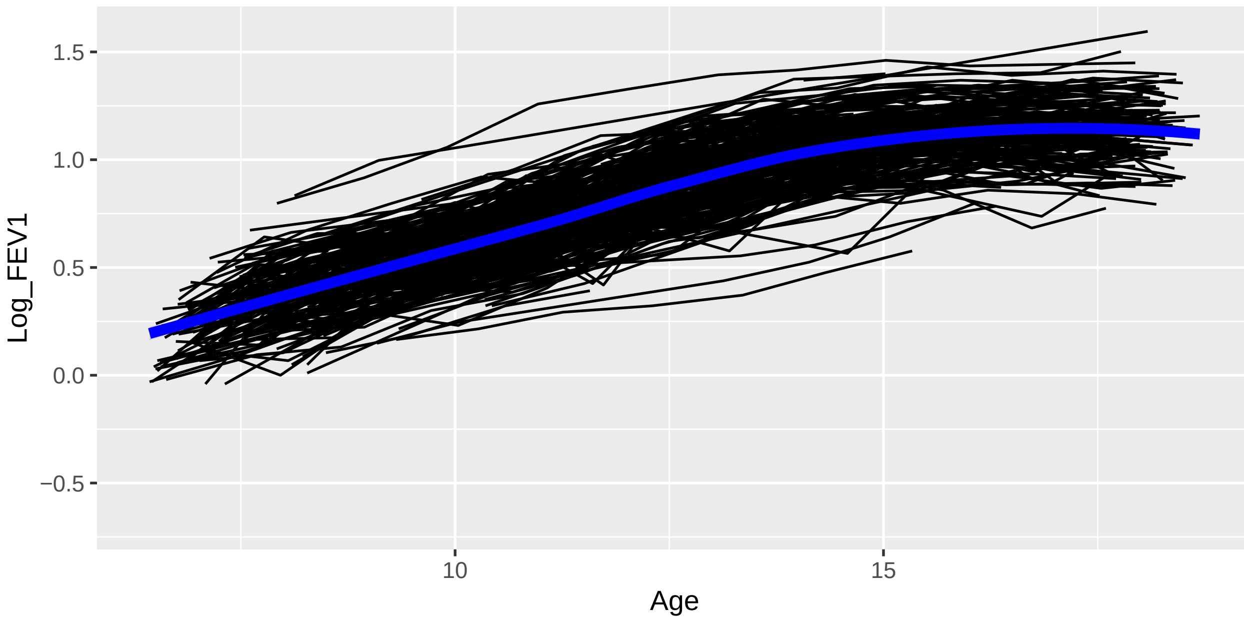
```
## 'data.frame':    1994 obs. of  6 variables:
##  $ ID        : int  1 1 1 1 1 1 1 2 2 2 ...
##  $ Height    : num  1.2 1.28 1.33 1.42 1.48 1.5 1.52 1.13 1.19 1.49 ...
##  $ Age       : num  9.34 10.39 11.45 12.46 13.42 ...
##  $ INI_Height: num  1.2 1.2 1.2 1.2 1.2 1.2 1.2 1.13 1.13 1.13 ...
##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

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

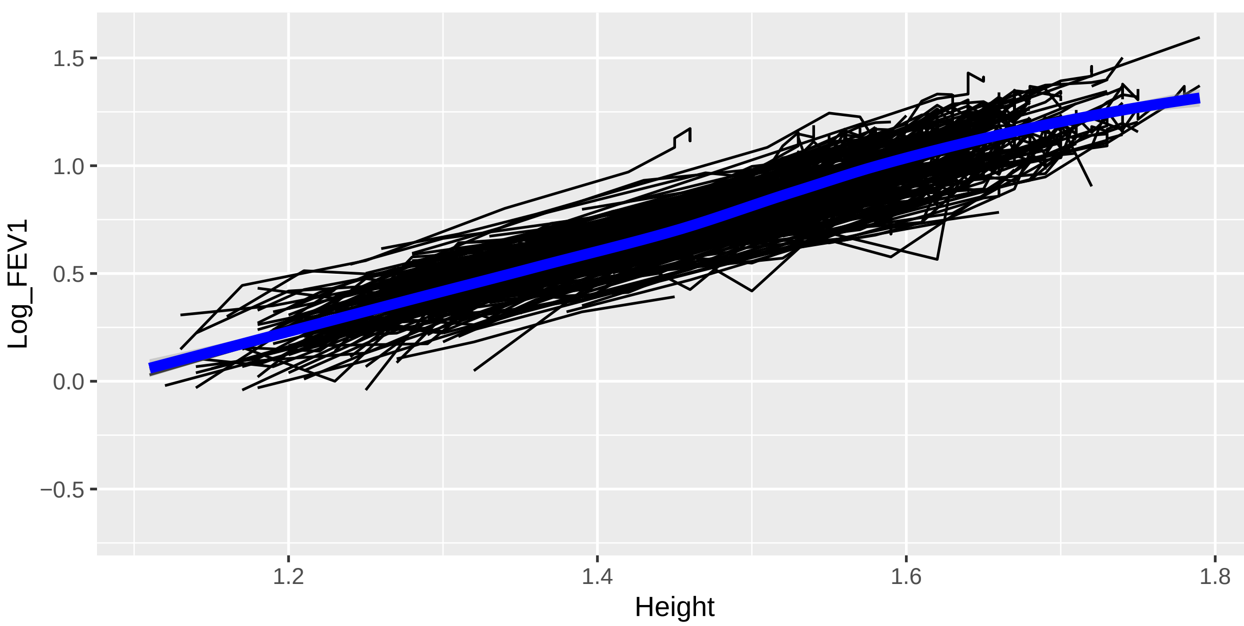
	ID	Height	Age	INI_Height	INI_Age	Log_FEV1
1987	299	1.64	17.9904	1.57	12.9555	1.09527
1988	300	1.44	11.9617	1.44	11.9617	0.68310
1989	300	1.50	12.9993	1.44	11.9617	0.85015
1990	300	1.57	13.9055	1.44	11.9617	0.81536
1991	300	1.61	14.9596	1.44	11.9617	1.11841
1992	300	1.62	15.9398	1.44	11.9617	1.08181
1993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

```
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)
```




```
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)
```



Let's look at fitting this model with a fixed effects model. First, we'll center `Log_FEV1`, `Height`, and `Age` by the subjects mean. Then we'll plot again.

```
Six_cities <- Six_cities %>% group_by(ID) %>%
  mutate( Log_FEV1_C = Log_FEV1 - mean(Log_FEV1),
          Height_C = Height - mean(Height),
          Age_C = Age - mean(Age))
head(Six_cities,10)
```

```
## # A tibble: 10 x 9
## # Groups:   ID [2]
##    ID    Height   Age INI_Height INI_Age Log_FEV1 Log_FEV1_C Height_C  Age_C
##    <fct>  <dbl> <dbl>      <dbl>   <dbl>    <dbl>      <dbl>    <dbl>  <dbl>
##  1 1       1.2   9.34       1.2     9.34    0.215    -0.417   -0.190  -3.36
##  2 1       1.28 10.4        1.2     9.34    0.372    -0.260   -0.110  -2.31
##  3 1       1.33 11.5        1.2     9.34    0.489    -0.143   -0.0600 -1.25
##  4 1       1.42 12.5        1.2     9.34    0.751     0.120    0.03   -0.242
##  5 1       1.48 13.4        1.2     9.34    0.833     0.201    0.09    0.717
##  6 1       1.5  15.5        1.2     9.34    0.892     0.260    0.11    2.77
##  7 1       1.52 16.4        1.2     9.34    0.871     0.239    0.13    3.67
##  8 2       1.13  6.59       1.13    6.59    0.307    -0.510   -0.319  -6.61
##  9 2       1.19  7.65       1.13    6.59    0.351    -0.467   -0.259  -5.55
## 10 2       1.49 12.7        1.13    6.59    0.756    -0.0615   0.0413 -0.457
```

```
## Check ##
Six_cities$Height[Six_cities$ID==1] - mean(Six_cities$Height[Six_cities$ID==1])
```

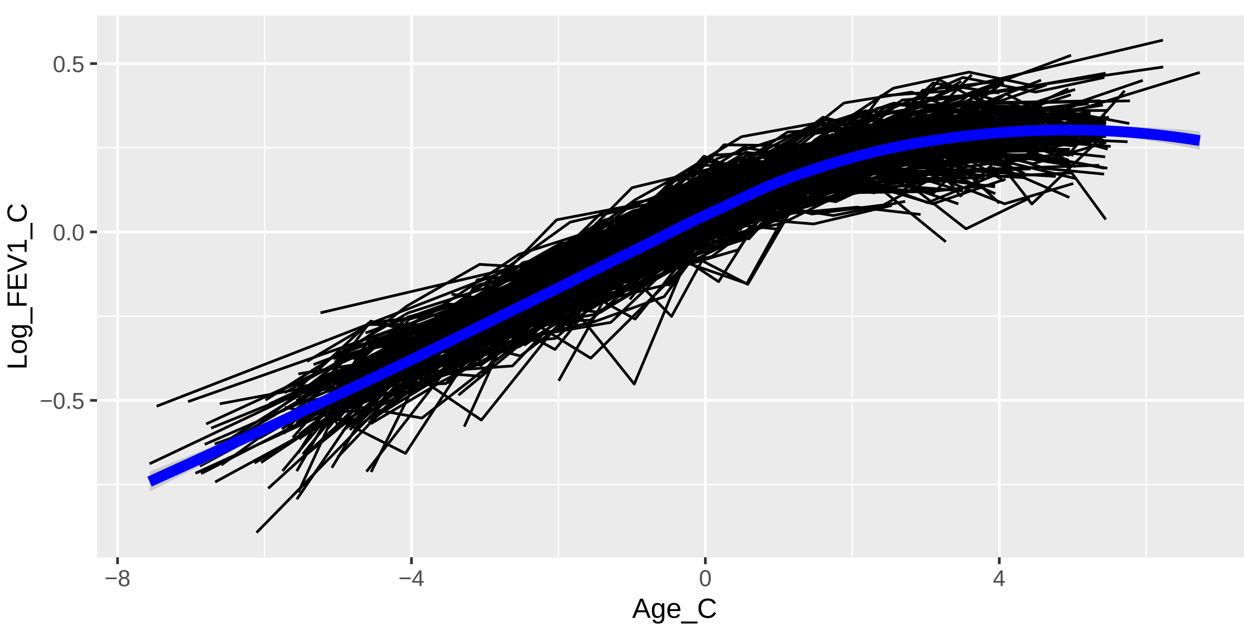
```
## [1] -0.19 -0.11 -0.06  0.03  0.09  0.11  0.13
```

```
Six_cities$Height[Six_cities$ID==2] - mean(Six_cities$Height[Six_cities$ID==2])
```

```
## [1] -0.31875 -0.25875  0.04125  0.08125  0.10125  0.11125  0.12125  0.12125
```

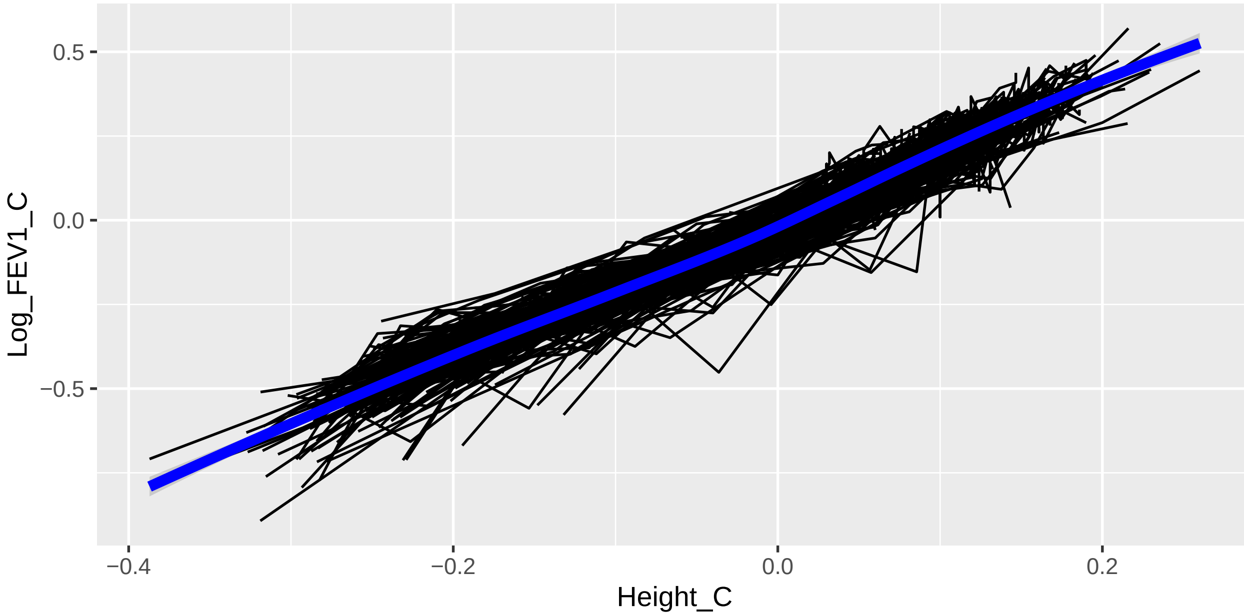
```
par(mfrow = c(1,2))
p <- ggplot(Six_cities, aes(x = Age_C, y = Log_FEV1_C, group = ID))
p + geom_line() +
  geom_smooth(aes(group = 1), method = "loess",
                              color = "blue", size = 2)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



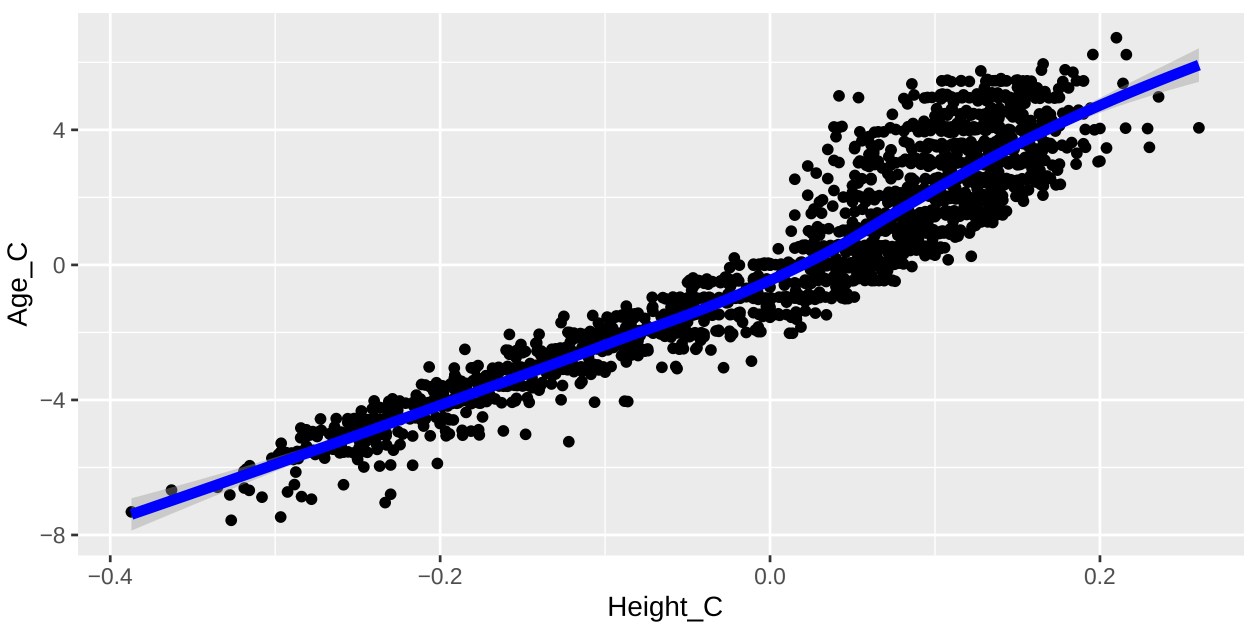

```
p <- ggplot(Six_cities, aes(x = Height_C, y = Log_FEV1_C, group = ID))
p + geom_line() +
  geom_smooth(aes(group = 1), method = "loess",
                              color = "blue", size = 2)
```

```
## `geom_smooth()` using formula 'y ~ x'
```




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

```
## `geom_smooth()` using formula 'y ~ x'
```



To fit the fixed effects model we'll use the `gls` function that we used previously Note that when we use this model we **do not** include an intercept. First, i'm going to fit the data by height and look at the residuals by age (all using the centered variables).

```r
library(nlme)
```

```r
Model_form <- Log_FEV1_C ~ 0 + Height_C + Age_C
cor_fun <- corIdent(form = ~ 1|ID)

lm_indep <- gls(model = Model_form, data = Six_cities, correlation = cor_fun)
summary(lm_indep)
```

```
## Generalized least squares fit by REML
##   Model: Model_form
##   Data: Six_cities
##         AIC       BIC   logLik
##   -5638.943 -5622.152 2822.471
##
## Correlation Structure: Independent
##  Formula: ~1 | ID
##  Parameter estimate(s):
## numeric(0)
##
## Coefficients:
##              Value  Std.Error  t-value p-value
## Height_C 1.6135411 0.02874199 56.13882       0
## Age_C    0.0199618 0.00124361 16.05151       0
##
##  Correlation:
##       Hght_C
## Age_C -0.937
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -6.38698429 -0.54781812  0.03431464  0.62763430  3.19497703
##
## Residual standard error: 0.05850395
## Degrees of freedom: 1994 total; 1992 residual
```

Another model we might want to try is log transformed age. To do this, we'll first create the log transformed variable and then mean center it.

```r
Six_cities <- Six_cities %>% group_by(ID) %>%
  mutate( log_Age = log(Age))%>%
  mutate( log_Age_C = log_Age - mean(log_Age))

Model_form <- Log_FEV1_C ~ 0 + Height_C + log_Age_C
cor_fun <- corIdent(form = ~ 1|ID)

lm_indep_log <- gls(model = Model_form, data = Six_cities, correlation = cor_fun)
anova(lm_indep, lm_indep_log)
```

	call	Model	df	AIC	BIC	logLik
lm_indep	gls(model = Model_form, data = Six_cities, correlation = cor_fun)	1	3	-5638.943	-5622.152	2822.471
lm_indep_log	gls(model = Model_form, data = Six_cities, correlation = cor_fun)	2	3	-5639.973	-5623.183	2822.987

```r
summary(lm_indep_log)
```

```
## Generalized least squares fit by REML
##   Model: Model_form
##   Data: Six_cities
##         AIC       BIC   logLik
##   -5639.973 -5623.183 2822.987
##
## Correlation Structure: Independent
##  Formula: ~1 | ID
##  Parameter estimate(s):
## numeric(0)
##
## Coefficients:
##               Value  Std.Error  t-value p-value
## Height_C  1.4677786 0.03773612 38.89585       0
## log_Age_C 0.3112462 0.01958350 15.89329       0
##
##  Correlation:
##           Hght_C
## log_Age_C -0.964
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -6.55804835 -0.57183519  0.03678813  0.63210271  3.00530086
##
## Residual standard error: 0.05856978
## Degrees of freedom: 1994 total; 1992 residual
```

Let's compare this to one of the mixed effects models we considered.

```r
Six_cities <- Six_cities %>% mutate( INI_log_Age = log( INI_Age ) )

library(lme4)
library(lmerTest)
LMM_formula <- Log_FEV1 ~ INI_Height + Height + INI_log_Age + log_Age + (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: -4582
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.6258 -0.4939  0.0856  0.5685  2.9134
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  ID       (Intercept) 0.079370 0.28173
##           Height      0.036043 0.18985  -0.93
##  Residual             0.003428 0.05855
```

```
## Number of obs: 1994, groups:  ID, 300
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)   -2.09856    0.09273  304.79008 -22.630  < 2e-16 ***
## INI_Height     0.39140    0.12102  336.11240   3.234  0.00134 **
## Height         1.50669    0.04215 1435.06842  35.748  < 2e-16 ***
## INI_log_Age   -0.28257    0.07211  306.96596  -3.919  0.00011 ***
## log_Age        0.29914    0.02059 1684.12517  14.526  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) INI_Hg Height INI__A
## INI_Height  -0.301
## Height      -0.030 -0.312
## INI_log_Age -0.281 -0.817  0.238
## log_Age     -0.024  0.301 -0.918 -0.252
```