# Cross Sectional and Longitudinal effects

## Alexander McLain

## Contents

3	MIT Growth Study	11
2	Example 2: Framingham Heart Study	5
1	Example 1: Six Cities	1

## 1 Example 1: Six Cities

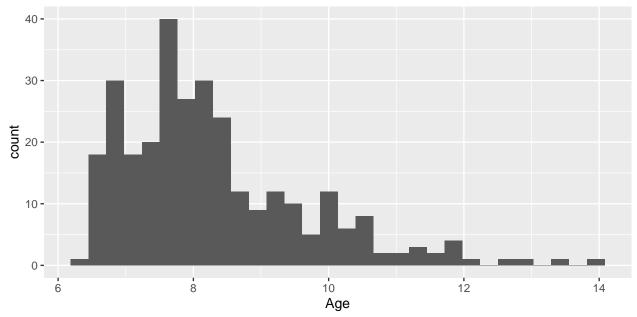
geom\_histogram()

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

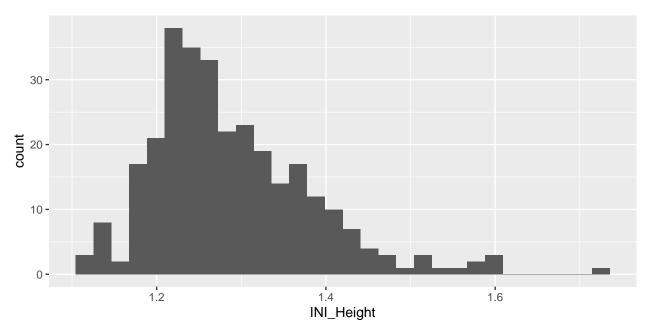
```
Six_cities <- read.csv("Six_cities.csv", header = TRUE)</pre>
```

Here we're going to look at the cross-sectional versus longitudinal effects of height and age. First, let's look at the distribution of baseline (initial) age. This will help us determine if we should use baseline or mean centered longitudinal effects. To do that, we'll create a data set with only the first observation and then look at a histrogram of the variables.

```
library(tidyverse)
Six_cities_ini <- Six_cities %>% group_by(ID) %>% filter(row_number(Age) == 1)
head(Six_cities_ini)
## # A tibble: 6 x 6
  # Groups:
               ID [6]
##
        ID Height
                    Age INI_Height INI_Age Log_FEV1
     <int>
            <dbl> <dbl>
                              <dbl>
                                      <dbl>
                                                <dbl>
             1.2
                               1.2
                                       9.34
                                               0.215
## 1
         1
                   9.34
## 2
         2
             1.13
                   6.59
                                       6.59
                                               0.307
                               1.13
## 3
         3
             1.18 6.91
                                       6.91
                                               0.432
                               1.18
         4
                  6.76
                                       6.76
                                               0.0583
             1.15
                               1.15
## 5
         5
             1.11
                   6.50
                               1.11
                                       6.50
                                               0.0296
             1.24
                   6.90
                               1.24
                                       6.90
                                               0.262
ggplot(data=Six_cities_ini, aes(Age)) +
```



```
ggplot(data=Six_cities_ini, aes(INI_Height)) +
geom_histogram()
```



Clearly there is no consistent definition to "baseline". When this is the case, we'll use mean centered longitudinal effects and the mean cross-sectional effect.

```
head(Six_cities,10)
## # A tibble: 10 x 13
## # Groups:
               ID [2]
##
         ID Height
                     Age INI_Height INI_Age Log_FEV1 Height_L Height_C Age_L Age_C
##
      <int> <dbl> <dbl>
                              <dbl>
                                       <dbl>
                                                <dbl>
                                                         <dbl>
                                                                  <dbl> <dbl> <dbl>
##
   1
          1
              1.2
                    9.34
                               1.2
                                        9.34
                                                0.215
                                                      -0.190
                                                                   1.39 -3.36
                                                                                 12.7
   2
                                                                   1.39 -2.31
##
              1.28 10.4
                               1.2
                                        9.34
                                                0.372 -0.110
                                                                                 12.7
          1
                                                                   1.39 -1.25
                                                0.489 -0.0600
##
   3
          1
              1.33 11.5
                               1.2
                                        9.34
                                                                                 12.7
##
              1.42 12.5
                                                                   1.39 -0.242 12.7
  4
                               1.2
                                        9.34
                                                0.751
                                                        0.03
          1
## 5
              1.48 13.4
                                        9.34
                                                        0.09
                                                                   1.39 0.717 12.7
          1
                               1.2
                                                0.833
##
  6
          1
             1.5 15.5
                               1.2
                                        9.34
                                                0.892
                                                       0.11
                                                                   1.39 2.77
                                                                                 12.7
##
  7
          1
              1.52 16.4
                               1.2
                                        9.34
                                                0.871
                                                        0.13
                                                                   1.39 3.67
                                                                                 12.7
## 8
          2
                                        6.59
                                                0.307 -0.319
                                                                   1.45 -6.61
                                                                                 13.2
              1.13 6.59
                               1.13
## 9
          2
              1.19 7.65
                               1.13
                                        6.59
                                                0.351 -0.259
                                                                   1.45 -5.55
                                                                                 13.2
                                        6.59
## 10
          2
              1.49 12.7
                               1.13
                                                0.756
                                                        0.0413
                                                                   1.45 -0.457 13.2
## # ... with 3 more variables: log_Age <dbl>, log_Age_L <dbl>, log_Age_C <dbl>
We'll fit this model using random effects.
library(lme4)
library(lmerTest)
LMM_formula <- Log_FEV1 ~ Height_L + Height_C + Age_L + Age_C + ( 1 + Height_L | ID )
LMM_L_CS_both <- lmer( formula = LMM_formula , data = Six_cities)</pre>
Let's test if the cross-sectional and longitudinal effects of age are equal.
library(car)
## Loading required package: carData
## Registered S3 methods overwritten by 'car':
##
     method
                                      from
##
     influence.merMod
                                      1me4
##
     cooks.distance.influence.merMod lme4
##
     dfbeta.influence.merMod
                                      1me4
##
     dfbetas.influence.merMod
                                      lme4
##
## Attaching package: 'car'
## The following object is masked from 'package:dplyr':
##
##
## The following object is masked from 'package:purrr':
##
##
linearHypothesis( LMM_L_CS_both, "Age_L = Age_C")
```

Df	Chisq	Pr(>Chisq)
NA	NA	NA
1	0.248565	0.6180873

```
LMM_formula <- Log_FEV1 ~ Height_L + Height_C + Age + ( 1 + Height_L | ID )
LMM_L_CS_hgt <- lmer( formula = LMM_formula , data = Six_cities)
anova(LMM_L_CS_hgt,LMM_L_CS_both)</pre>
```

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMM_L_CS_hgt	8	-4604.165	-4559.382	2310.082	-4620.165	NA	NA	NA
LMM_L_CS_both	9	-4602.414	-4552.033	2310.207	-4620.414	0.2488397	1	0.6178933

Now let's test if the cross-sectional and longitudinal effects of height are equal.

linearHypothesis( LMM\_L\_CS\_hgt, "Height\_L = Height\_C")

Df	Chisq	Pr(>Chisq)
NA	NA	NA
1	0.0580345	0.8096299

```
LMM_formula <- Log_FEV1 ~ Height + Age + ( 1 + Height_L| ID )
LMM_L_CS_none <- lmer( formula = LMM_formula , data = Six_cities)
anova(LMM_L_CS_none,LMM_L_CS_hgt)</pre>
```

## refitting model(s) with ML (instead of REML)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMM_L_CS_none	7	-4606.106	-4566.921	2310.053	-4620.106	NA	NA	NA
$LMM\_L\_CS\_hgt$	8	-4604.165	-4559.382	2310.082	-4620.165	0.059017	1	0.8080564

Now we can check if the mean centered or longitudinal height is a better random effect term.

```
LMM_formula <- Log_FEV1 ~ Height + Age + ( 1 + Height | ID )
LMM_L_CS_none_H <- lmer( formula = LMM_formula , data = Six_cities)
anova(LMM_L_CS_none,LMM_L_CS_none_H)</pre>
```

## refitting model(s) with ML (instead of REML)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMM_L_CS_none	7	-4606.106	-4566.921	2310.053	-4620.106	NA	NA	NA
$LMM_L_CS_none_H$	7	-4603.339	-4564.154	2308.670	-4617.339	0	0	1

Let's look at log(Age) and see what we get.

```
LMM_formula <- Log_FEV1 ~ Height + log_Age_L + log_Age_C + ( 1 + Height | ID )
LMM_L_CS_log_age <- lmer( formula = LMM_formula , data = Six_cities)
```

Let's **test** if the cross-sectional and longitudinal effects of age are equal.

```
linearHypothesis( LMM_L_CS_log_age, "log_Age_L = log_Age_C")
```

Df	Chisq	Pr(>Chisq)
NA	NA	NA
1	0.5599279	0.4542893

```
LMM_formula <- Log_FEV1 ~ Height + log_Age + ( 1 + Height | ID )
LMM_L_CS_log_none <- lmer( formula = LMM_formula , data = Six_cities)
anova(LMM_L_CS_log_none,LMM_L_CS_none_H)</pre>
```

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMM_L_CS_log_none	7	-4580.667	-4541.482	2297.333	-4594.667	NA	NA	NA
$LMM\_L\_CS\_none\_H$	7	-4603.339	-4564.154	2308.670	-4617.339	22.67242	0	0

```
summary(LMM_L_CS_none_H)
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: LMM formula
##
      Data: Six_cities
##
## REML criterion at convergence: -4591.2
##
## Scaled residuals:
                1Q
                   Median
                                3Q
                                       Max
##
  -6.4921 -0.4966
                   0.0800
                            0.5660
                                    2.9045
##
## Random effects:
                         Variance Std.Dev. Corr
   Groups
             Name
             (Intercept) 0.085212 0.29191
##
   ID
##
             Height
                         0.038371 0.19589
                                           -0.94
   Residual
                         0.003387 0.05819
## Number of obs: 1994, groups: ID, 300
##
## Fixed effects:
                 Estimate Std. Error
                                              df t value Pr(>|t|)
## (Intercept) -1.903e+00 3.499e-02
                                                  -54.39
                                      6.037e+02
                                                           <2e-16 ***
## Height
                1.658e+00 3.188e-02
                                      1.229e+03
                                                   51.99
                                                           <2e-16 ***
## Age
                1.876e-02 1.249e-03
                                      1.770e+03
                                                   15.03
                                                           <2e-16 ***
##
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Correlation of Fixed Effects:
          (Intr) Height
## Height -0.962
           0.708 -0.848
## Age
```

## 2 Example 2: Framingham Heart Study

In the Framingham study, each of 2634 participants was examined every 2 years for a 10 year period for his/her cholesterol level. We already examined the effect of time, age and age at measurement. This time

let's test if the cross-sectional and longitudinal effects of age are equal.

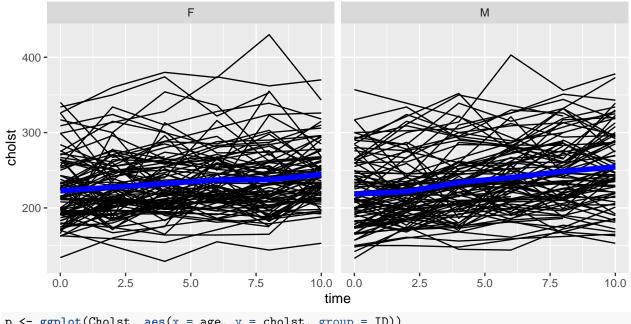
What is the cross-sectional effect of age? What is the longitudinal effect of age?

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

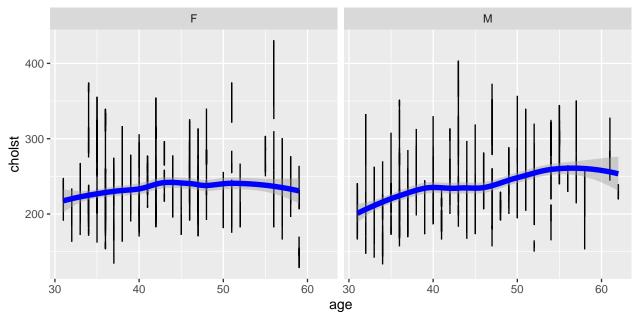
## 'data.frame': 1044 obs. of 5 variables:
## $ ID : int 1 1 1 1 1 1 2 2 2 2 ...
## $ cholst: int 175 198 205 228 214 214 299 328 374 362 ...
## $ sex : int 1 1 1 1 1 1 1 0 0 0 0 ...
## $ age : int 32 32 32 32 32 32 34 34 34 34 ...
## $ time : int 0 2 4 6 8 10 0 4 6 8 ...
head(Cholst,8)</pre>
```

ID	cholst	sex	age	time
1	175	1	32	0
1	198	1	32	2
1	205	1	32	4
1	228	1	32	6
1	214	1	32	8
1	214	1	32	10
2	299	0	34	0
2	328	0	34	4

Some data cleaning



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



```
LMM_formula <- cholst ~ time + age + sex + sex*age + sex*time + (1|ID)
LMM_TAbS <- lmer( formula = LMM_formula , data = Cholst)
linearHypothesis( LMM_TAbS, "time = age")</pre>
```

Df	Chisq	Pr(>Chisq)
NA	NA	NA
1	5.287729	0.0214762

### linearHypothesis( LMM\_TAbS, "age:sexM = time:sexM")

Df	Chisq	Pr(>Chisq)
NA	NA	NA
1	0.0208743	0.885122

```
LMM_formula <- cholst ~ time + log(age) + sex +
   sex*log(age) + sex*time + (1 | ID)
LMM_TlAbS <- lmer( formula = LMM_formula , data = Cholst)
anova(LMM_TAbS, LMM_TlAbS)</pre>
```

#### ## refitting model(s) with ML (instead of REML)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMM_TAbS	8	9959.684	9999.291	-4971.842	9943.684	NA	NA	NA
$LMM\_TlAbS$	8	9959.288	9998.895	-4971.644	9943.288	0.3957945	0	0

```
LMM_formula <- cholst ~ time + log(age) + sex +
  sex*log(age) + sex*time + (1 + time | ID)
LMM_TlAbS_sl <- lmer( formula = LMM_formula , data = Cholst)
anova(LMM_TlAbS_sl, LMM_TlAbS)</pre>
```

#### ## refitting model(s) with ML (instead of REML)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMM_TlAbS	8	9959.288	9998.895	-4971.644	9943.288	NA	NA	NA
$LMM\_TlAbS\_sl$	10	9940.624	9990.132	-4960.312	9920.624	22.66458	2	1.2e-05

```
LMM_formula <- cholst ~ time + log(age) + sex +
   sex*log(age) + sex*time + (0 + time + log(age) | ID)
LMM_TlAbS_Dsl <- lmer( formula = LMM_formula , data = Cholst)
anova(LMM_TlAbS_Dsl, LMM_TlAbS_sl)</pre>
```

#### ## refitting model(s) with ML (instead of REML)

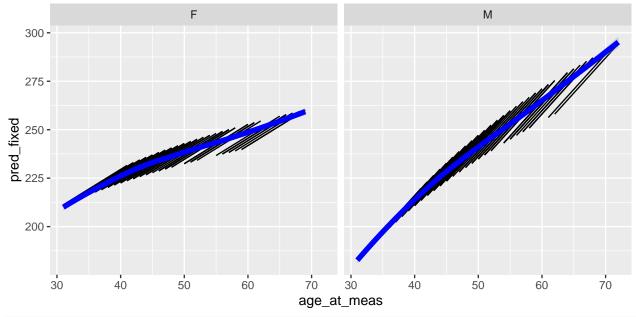
	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMM_TlAbS_Dsl	10	9938.169	9987.677	-4959.085	9918.169	NA	NA	NA
$LMM\_TlAbS\_sl$	10	9940.624	9990.132	-4960.312	9920.624	0	0	1

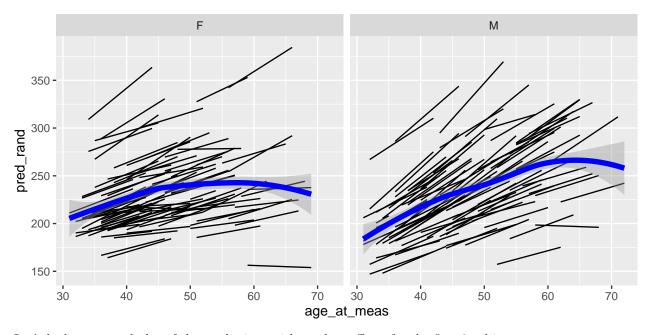
```
summary(LMM_TlAbS_Dsl)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [ ^{-}
```

## lmerModLmerTest]

```
## Formula: LMM_formula
##
     Data: Cholst
##
## REML criterion at convergence: 9894.3
## Scaled residuals:
             10 Median
                               30
                                      Max
## -4.3010 -0.4997 -0.0151 0.5319 3.9078
##
## Random effects:
## Groups
            Name
                     Variance Std.Dev. Corr
                              1.807
## ID
                       3.267
            time
            log(age) 83.740
                              9.151
##
                                       0.21
                     434.190 20.837
## Residual
## Number of obs: 1044, groups: ID, 200
##
## Fixed effects:
##
                 Estimate Std. Error
                                            df t value Pr(>|t|)
## (Intercept)
                  59.0537
                             81.6100 193.8028
                                                0.724 0.470179
                              0.3292 173.2995
## time
                   2.0322
                                                 6.173 4.61e-09 ***
## log(age)
                  44.3223
                             22.0209 193.3444
                                                2.013 0.045529 *
## sexM
                -245.7750
                           109.1012 195.2389 -2.253 0.025389 *
## log(age):sexM
                 63.4377
                             29.3207 194.6924
                                                 2.164 0.031714 *
## time:sexM
                   1.5974
                              0.4689 177.6668
                                                3.407 0.000813 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Correlation of Fixed Effects:
##
              (Intr) time
                            log(g) sexM
                                          lg():M
              -0.014
## time
## log(age)
              -0.999 0.006
## sexM
              -0.748 0.010 0.747
## log(ag):sxM 0.750 -0.005 -0.751 -0.999
               0.010 -0.702 -0.004 -0.014 0.006
## time:sexM
Let's look at some predicted values
Cholst <- Cholst %>% mutate(pred_rand = predict(LMM_TlAbS_Dsl),
                           pred_fixed = predict(LMM_TlAbS_Dsl, re.form = ~0))
p <- ggplot(Cholst, aes(x = age_at_meas, y = pred_fixed, group = ID))</pre>
p + geom_line() +
 geom_smooth(aes(group = 1), method = "loess",
                             color = "blue", size = 2) +
facet_grid(. ~ sex)
```



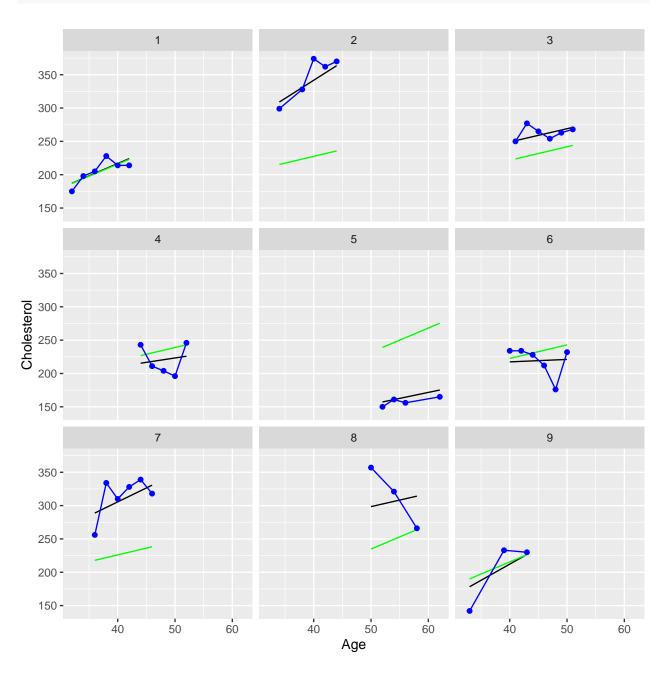


Let's look at a panel plot of the predictions with random effects for the first 9 subjects

```
Cholst_sm <- Cholst %>% filter(ID < 10)

p <- ggplot(Cholst_sm, aes(x = age_at_meas, y = pred_rand, group = ID))
p + geom_line() +
   geom_line(aes(y = pred_fixed), color = "green") +
   geom_point(aes(y = cholst), color = "blue") +</pre>
```

```
geom_line(aes(y = cholst), color = "blue") +
facet_wrap(. ~ ID) + xlab("Age") + ylab("Cholesterol")
```



# 3 MIT Growth Study

These data are from a prospective study on body fat accretion in a cohort of 162 girls from the MIT Growth and Development Study. The study was designed to look at changes in percent body fat in girls before and after menarche. All subjects had to be pre-menarche and non-obese to enter the study. Observations were taken annually until 4 years after menarche. At each observation percent body fat was measured.

Two time-scales are included: age, and time since menarche (which can be negative). Time since menarche is the more biologically relevant time scale to use. The variables (in order) are: Subject ID, Current Age

(years), Age at Menarche (years), time relative to Menarche (years), Percent Body Fat.

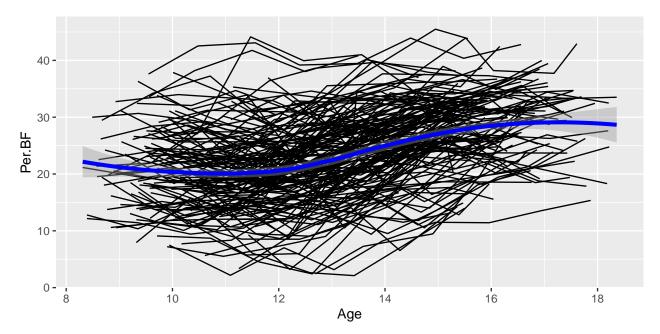
```
MITgrowth <- read.csv("MITgrowth.csv", header = FALSE, na.strings = "",
                      stringsAsFactors = FALSE)
names(MITgrowth) <- c("ID", "Age", "Age.men", "time.r.men", "Per.BF")</pre>
str(MITgrowth)
## 'data.frame':
                    1049 obs. of 5 variables:
    $ ID
                : int 1 1 1 1 1 1 2 2 2 2 ...
##
## $ Age
                : num 9.32 10.33 11.24 12.19 13.24 ...
                : num 13.2 13.2 13.2 13.2 13.2 ...
## $ Age.men
## $ time.r.men: num -3.87 -2.86 -1.95 -1 0.05 1.05 -4.44 -3.2 -2.25 -0.51 ...
  $ Per.BF
                : num 7.94 15.65 13.51 23.23 10.52 ...
head(MITgrowth)
```

ID	Age	Age.men	time.r.men	Per.BF
1	9.32	13.19	-3.87	7.94
1	10.33	13.19	-2.86	15.65
1	11.24	13.19	-1.95	13.51
1	12.19	13.19	-1.00	23.23
1	13.24	13.19	0.05	10.52
1	14.24	13.19	1.05	20.45

First, we'll look at age.

```
p <- ggplot(data = MITgrowth, aes(x = Age, y = Per.BF, group = ID))
p + geom_line() +
geom_smooth(aes(group = 1), method = "loess", color = "blue", size = 1.5)</pre>
```

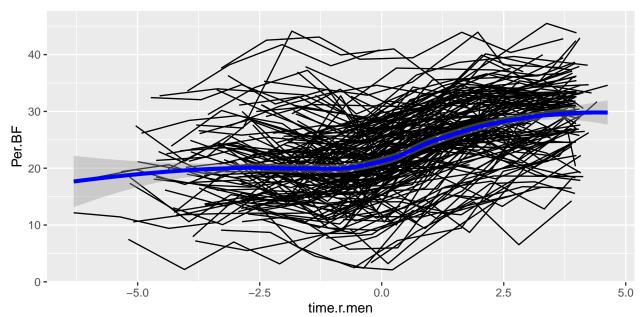
## `geom\_smooth()` using formula 'y ~ x'



Second, we'll look at time relative to menarche.

```
p <- ggplot(data = MITgrowth, aes(x = time.r.men, y = Per.BF, group = ID))
p + geom_line() +
geom_smooth(aes(group = 1), method = "loess", color = "blue", size = 1.5)</pre>
```

## `geom\_smooth()` using formula 'y ~ x'



ID	Age	Age.men	time.r.men	Per.BF	time.r.men.pos
1	9.32	13.19	-3.87	7.94	0.00
1	10.33	13.19	-2.86	15.65	0.00
1	11.24	13.19	-1.95	13.51	0.00
1	12.19	13.19	-1.00	23.23	0.00
1	13.24	13.19	0.05	10.52	0.05
1	14.24	13.19	1.05	20.45	1.05

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMM	7	6094.803	6129.492	-3040.402	6080.803	NA	NA	NA
$LMM\_ba$	10	6078.313	6127.869	-3029.156	6058.313	22.49076	3	5.16e-05

#### summary(LMM\_ba)

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: LMM_formula
##
     Data: MITgrowth
##
## REML criterion at convergence: 6062.4
##
## Scaled residuals:
##
      Min
              1Q Median
                                ЗQ
                                       Max
## -2.7743 -0.5901 -0.0359 0.5947 3.3798
##
## Random effects:
##
   Groups
            Name
                            Variance Std.Dev. Corr
##
             (Intercept)
                            45.940
                                     6.778
##
            time.r.men
                             1.631
                                     1.277
                                               0.29
##
                            2.750
                                     1.658
                                              -0.54 -0.83
            time.r.men.pos
                                     3.078
## Residual
                             9.473
## Number of obs: 1049, groups: ID, 162
##
## Fixed effects:
                  Estimate Std. Error
                                            df t value Pr(>|t|)
##
                   21.3614
                               0.5645 161.5659 37.838 < 2e-16 ***
## (Intercept)
                   0.4171
                               0.1572 108.4569
                                                2.654 0.00915 **
## time.r.men
## time.r.men.pos
                   2.0471
                               0.2280 132.6744
                                                8.980 2.32e-15 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##
               (Intr) tm.r.m
              0.351
## time.r.men
## tim.r.mn.ps -0.515 -0.872
```

Were going to compare these models with one that uses an interaction with an indicator that the girl has reached menarche.

```
MITgrowth <- MITgrowth %>% mutate( Men_ind = if_else( time.r.men > 0, 1, 0) )
head( MITgrowth , 10 )
```

ID	Age	Age.men	time.r.men	Per.BF	time.r.men.pos	Men_ind
1	9.32	13.19	-3.87	7.94	0.00	0
1	10.33	13.19	-2.86	15.65	0.00	0
1	11.24	13.19	-1.95	13.51	0.00	0
1	12.19	13.19	-1.00	23.23	0.00	0
1	13.24	13.19	0.05	10.52	0.05	1
1	14.24	13.19	1.05	20.45	1.05	1
2	8.84	13.28	-4.44	16.17	0.00	0
2	10.08	13.28	-3.20	13.34	0.00	0

ID	Age	Age.men	time.r.men	Per.BF	time.r.men.pos	Men_ind
2	11.03	13.28	-2.25	16.05	0.00	0
2	12.77	13.28	-0.51	15.26	0.00	0

```
LMM_formula <- Per.BF ~ time.r.men + Men_ind*time.r.men - Men_ind +
   (1 + time.r.men + Men_ind*time.r.men - Men_ind | ID)
LMM_alt_ba <- lmer( formula = LMM_formula , data = MITgrowth)
### Compare with the model with the linear spline and 3 random effects.
anova(LMM_ba,LMM_alt_ba)</pre>
```

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMM_ba	10	6078.313	6127.869	-3029.156	6058.313	NA	NA	NA
$LMM\_alt\_ba$	10	6078.313	6127.869	-3029.156	6058.313	0	0	1

```
LMM_formula <- Per.BF ~ Age + Men_ind*Age - Men_ind + (1 + Age |ID)
LMM_alt <- lmer( formula = LMM_formula , data = MITgrowth)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00380159 (tol = 0.002, component 1)

### Compare with the model with the linear spline and 2 random effects.
anova(LMM,LMM_alt)</pre>
```

#### ## refitting model(s) with ML (instead of REML)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMM	7	6094.803	6129.492	-3040.402	6080.803	NA	NA	NA
$LMM\_alt$	7	6155.182	6189.871	-3070.591	6141.182	0	0	1

### summary(LMM\_alt)

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: LMM formula
      Data: MITgrowth
##
## REML criterion at convergence: 6149.1
##
## Scaled residuals:
##
      Min
                1Q Median
                                3Q
                                       Max
## -3.0949 -0.5721 -0.0082 0.6213 3.4202
##
## Random effects:
## Groups
                        Variance Std.Dev. Corr
## ID
             (Intercept) 170.1814 13.0454
##
                          0.6358 0.7974
                                          -0.88
## Residual
                          10.9493 3.3090
## Number of obs: 1049, groups: ID, 162
##
```