LMM Examples

Alexander McLain

Contents

1 Example one: Framingham data

In the Framingham study, each of 2634 participants was examined every 2 years for a 10 year period for his/her cholesterol level. Study objectives: + How does cholesterol level change over time on average as people get older? + How is the change of cholesterol level associated with sex and baseline age? A subset of 200 subjects' data is used for illustrative purpose.

```
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 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)
```

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

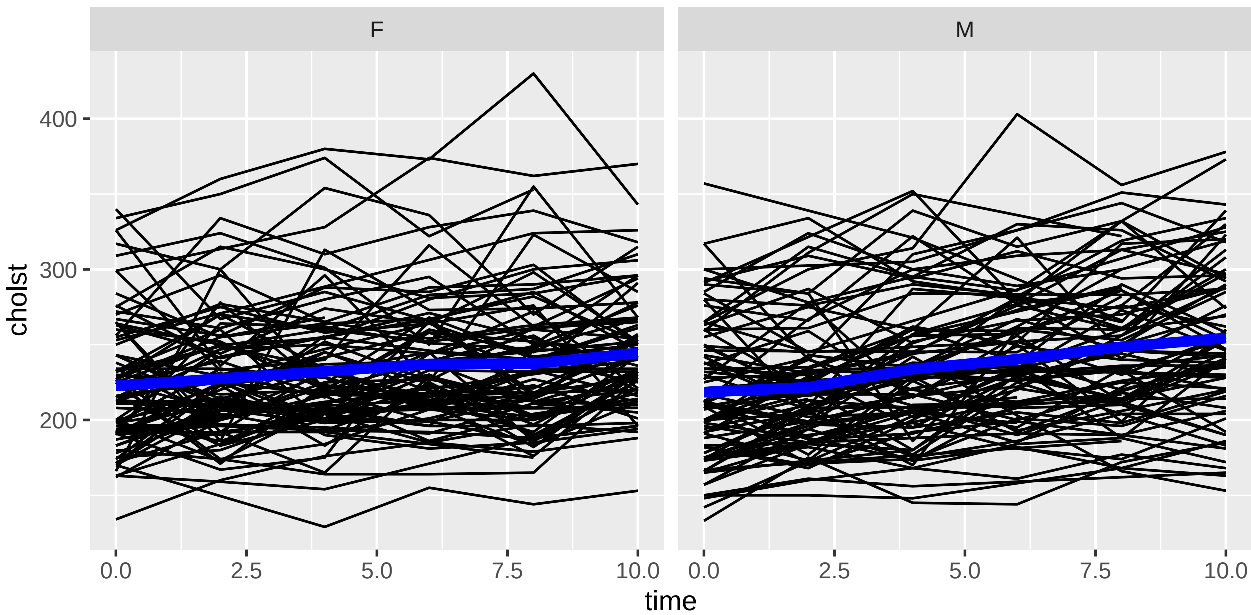
First, thing i'm going to do is to change the formatting of some of the variables and create a variable that is the subjects age at the time of their measurement.

```
library(tidyverse)
```

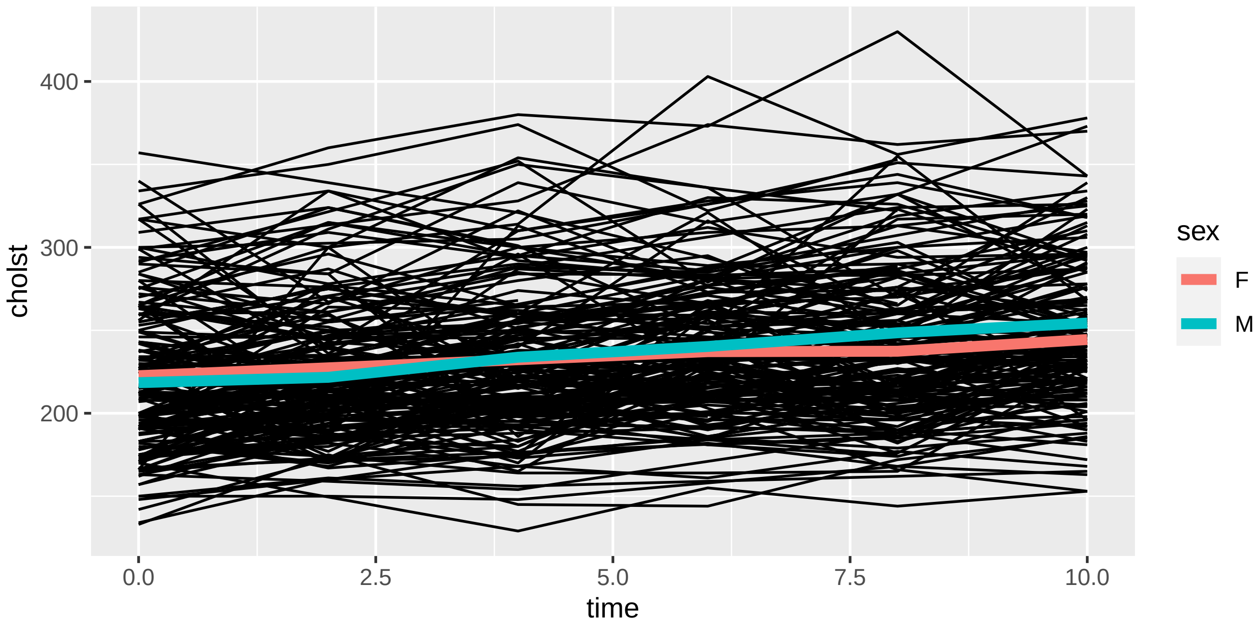
```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

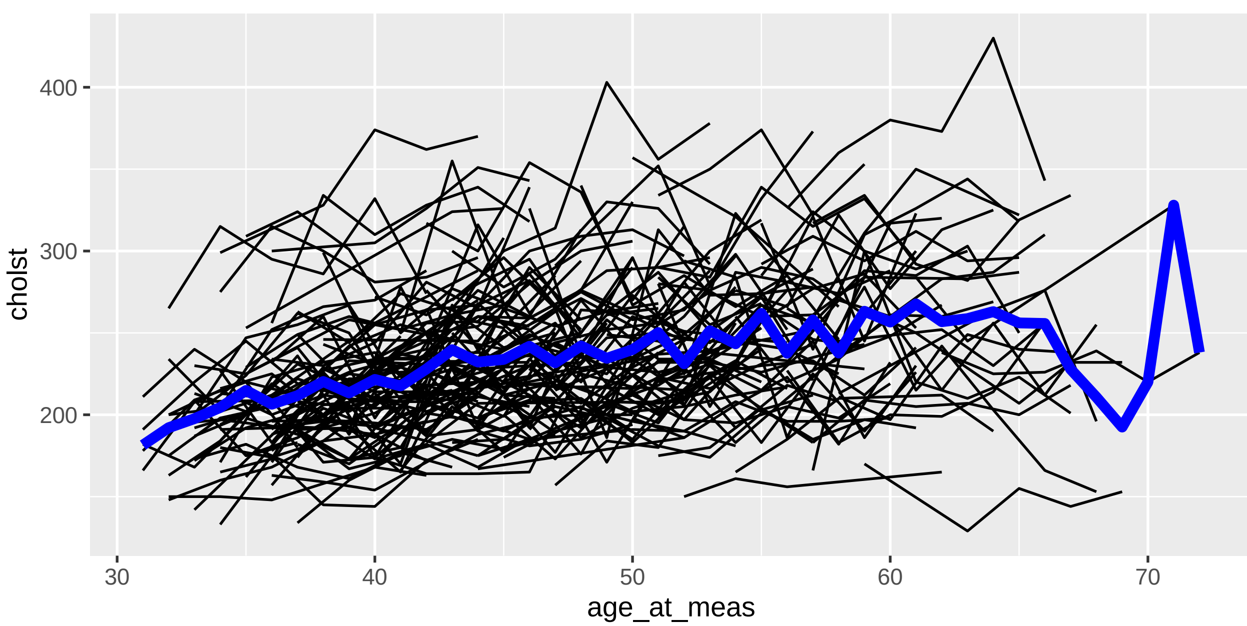
```r
Cholst <- Cholst %>% mutate( sex = factor(sex, levels = c(0,1) ,
                                          labels = c("F","M") ),
                             age_at_meas = age + time)
### Let's look at the trends by time ###
p <- ggplot(Cholst, aes(x = time, y = cholst, group = ID))
p + geom_line() +
  stat_summary(aes(group = 1), geom = "line", color = "blue",
               fun = mean, size = 2) +
  facet_grid(. ~ sex)
```




```r
p + geom_line() +
  stat_summary(aes(group = sex, color = sex), geom = "line",
               fun = mean, size = 2)
```



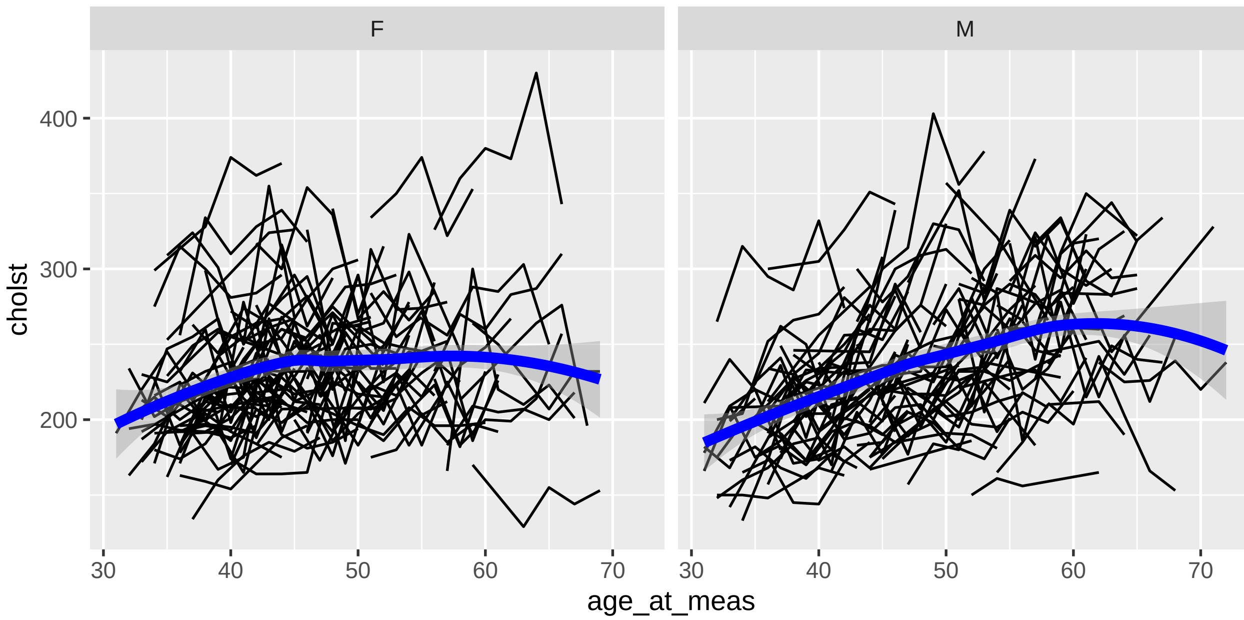

```
### Let's look at the trends by Age at measurement ###
p <- ggplot(Cholst, aes(x = age_at_meas, y = cholst, group = ID))
p + geom_line() +
  stat_summary(aes(group = 1), geom = "line", color = "blue",
               fun = mean, size = 2)
```



Once we look at the data as a function of "age at measurement" it's clear how unbalenced the data are. To plot unbalenced data with a mean trend, we need to use `geom_smooth` with the lowess function.

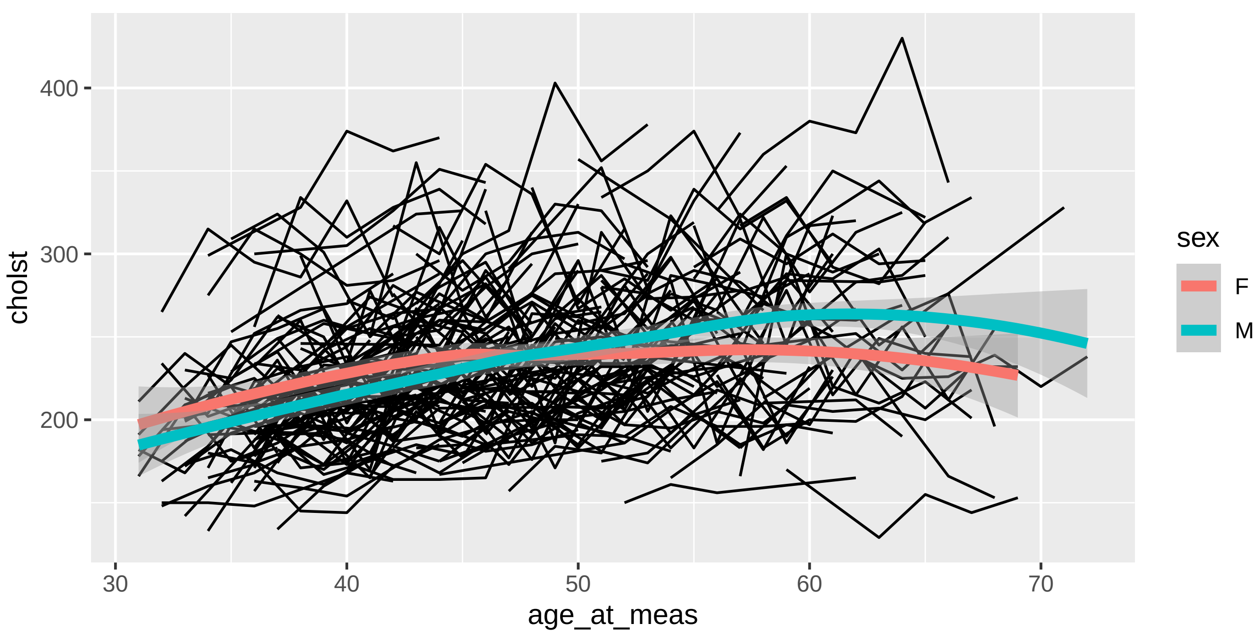
```
p + geom_line() +
  geom_smooth(aes(group = 1), method = "loess",
                              color = "blue", size = 2) +
  facet_grid(. ~ sex)
```

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




```r
p + geom_line() +
  geom_smooth(aes(group = sex, color = sex), method = "loess",
                                            size = 2)
```

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



Does the mean actually go down? Do we want to model this by “age at measurement”, “time” or both?

```r
library(lme4)
```

```
## Loading required package: Matrix
##
```

```
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```r
library(lmerTest) # Added to get p-values
```

```
##
## Attaching package: 'lmerTest'

## The following object is masked from 'package:lme4':
##
##     lmer

## The following object is masked from 'package:stats':
##
##     step
```

```r
LMM_formula <- cholst ~ time + age + (1|ID)
LMM_time_age <- lmer( formula = LMM_formula , data = Cholst)


LMM_formula <- cholst ~ time + age_at_meas + (1|ID)
LMM_time_age_at_meas <- lmer( formula = LMM_formula , data = Cholst)
anova(LMM_time_age,LMM_time_age_at_meas)
```

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

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMM_time_age	5	9974.247	9999.001	-4982.123	9964.247	NA	NA	NA
LMM_time_age_at_meas	5	9974.247	9999.001	-4982.123	9964.247	0	0	NA

Why are they the same?

```r
fixef(LMM_time_age)
```

```
## (Intercept)        time         age
##  157.083561    2.825912    1.492383
```

```r
fixef(LMM_time_age_at_meas)
```

```
## (Intercept)        time age_at_meas
##  157.083561    1.333529    1.492383
```

```r
1.333529 + 1.492383
```

```
## [1] 2.825912
```

```r
LMM_formula <- cholst ~ time + age + sex + sex*time + (1|ID)
LMM_timebysex <- lmer( formula = LMM_formula , data = Cholst)

LMM_formula <- cholst ~ time + age + sex + sex*age + (1|ID)
LMM_agebysex <- lmer( formula = LMM_formula , data = Cholst)

LMM_formula <- cholst ~ time + age + sex + sex*age + sex*time + (1|ID)
LMM_timeagebysex <- lmer( formula = LMM_formula , data = Cholst)
```

```
LMM_formula <- cholst ~ time + age_at_meas + sex + sex*age_at_meas + (1|ID)
LMM_age_at_meas_sex <- lmer( formula = LMM_formula , data = Cholst)

anova(LMM_timebysex, LMM_agebysex,
      LMM_timeagebysex,LMM_age_at_meas_sex)
```

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

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMM_timebysex	7	9962.253	9996.909	-4974.127	9948.253	NA	NA	NA
LMM_agebysex	7	9973.648	10008.304	-4979.824	9959.648	0.0000000	0	NA
LMM_age_at_meas_sex	7	9957.706	9992.361	-4971.853	9943.706	15.9424963	0	NA
LMM_timeagebysex	8	9959.684	9999.291	-4971.842	9943.684	0.0214397	1	0.8835874

Now, let's look at what we really wanted to from the plots we saw.

```
LMM_formula <- cholst ~ time + log(age_at_meas) + sex +
  sex*log(age_at_meas) + (1|ID)
LMM_age_log <- lmer( formula = LMM_formula , data = Cholst)


LMM_formula <- cholst ~ log(time+1) + log(age_at_meas) + sex +
  sex*log(age_at_meas) + (1|ID)
LMM_all_log <- lmer( formula = LMM_formula , data = Cholst)

anova(LMM_age_log,LMM_all_log)
```

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

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMM_age_log	7	9947.257	9981.913	-4966.629	9933.257	NA	NA	NA
LMM_all_log	7	9950.271	9984.926	-4968.135	9936.271	0	0	NA

```
LMM_formula <- cholst ~ time + log(age_at_meas) + sex +
  sex*time + (1 + time |ID)
LMM_age_log_rand_time <- lmer( formula = LMM_formula , data = Cholst)


LMM_formula <- cholst ~ time + log(age_at_meas) + sex +
  sex*log(age_at_meas) + (1 + log(age_at_meas) |ID)
LMM_age_log_rand_age <- lmer( formula = LMM_formula , data = Cholst)
anova(LMM_age_log,LMM_age_log_rand_time)
```

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

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMM_age_log	7	9947.257	9981.913	-4966.629	9933.257	NA	NA	NA
LMM_age_log_rand_time	9	9935.682	9980.239	-4958.841	9917.682	15.57512	2	0.0004149

It say's that random log age fits better. Let's look into the results a little further.

```r
VarCorr(LMM_age_log_rand_time)
```

```
##  Groups   Name        Std.Dev. Corr
##  ID       (Intercept) 34.7685
##           time         1.6894  0.236
##  Residual             20.8342
```

```r
confint(LMM_age_log_rand_time)
```

```
## Computing profile confidence intervals ...
```

	2.5 %	97.5 %
.sig01	30.5983043	38.8804240
.sig02	-0.0638044	0.7194118
.sig03	0.9421964	2.2554851
.sigma	19.7624645	22.0048151
(Intercept)	-225.3094008	-5.1944775
time	-0.8837506	0.9251927
log(age_at_meas)	61.6253365	120.6798372
sexM	-20.0906493	1.1375084
time:sexM	0.7594824	2.5586478

```r
VarCorr(LMM_age_log_rand_age)
```

```
##  Groups   Name             Std.Dev. Corr
##  ID       (Intercept)      223.755
##           log(age_at_meas)  62.517  -0.990
##  Residual                   21.057
```

```r
confint(LMM_age_log_rand_age)
```

```
## Computing profile confidence intervals ...
```

```
## Warning in FUN(X[[i]], ...): non-monotonic profile for .sig02
```

```
## Warning in confint.thpr(pp, level = level, zeta = zeta): bad spline fit
## for .sig02: falling back to linear interpolation
```

	2.5 %	97.5 %
.sig01	98.2127990	311.0346463
.sig02	-0.9942734	-0.9673241
.sig03	33.4396309	84.8109487
.sigma	20.0075750	22.1863650
(Intercept)	-108.0487611	157.4196654
time	0.1145259	1.6161199
log(age_at_meas)	18.0378621	89.5543423
sexM	-415.3336492	-153.5784020
log(age_at_meas):sexM	38.8366592	108.3629878

```r
data.frame( VarCorr(LMM_age_log_rand_time) )
```

grp	var1	var2	vcov	sdcor
ID	(Intercept)	NA	1208.845982	34.7684625
ID	time	NA	2.854119	1.6894139
ID	(Intercept)	time	13.858769	0.2359408
Residual	NA	NA	434.063220	20.8341839

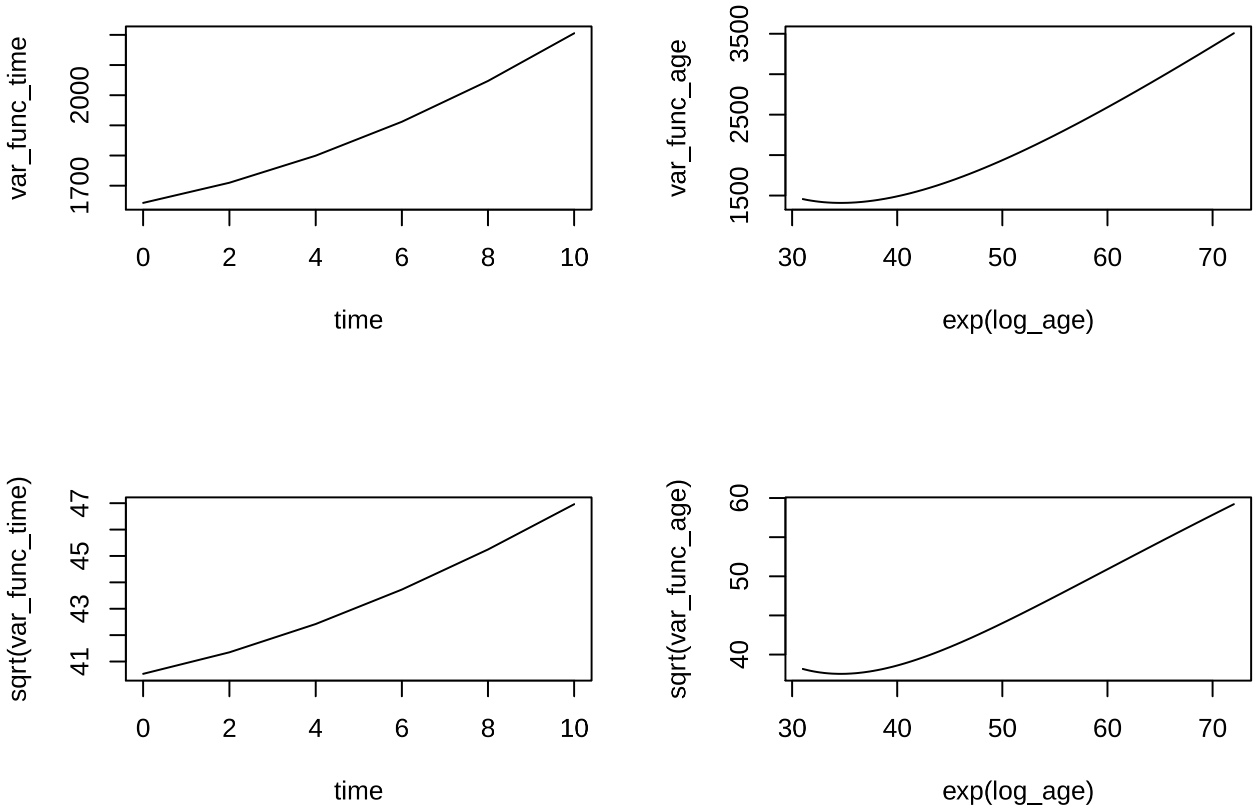
```
vcov_est_time <- data.frame( VarCorr(LMM_age_log_rand_time) )
str(vcov_est_time)
```

```
## 'data.frame':    4 obs. of  5 variables:
##  $ grp  : chr  "ID" "ID" "ID" "Residual"
##  $ var1 : chr  "(Intercept)" "time" "(Intercept)" NA
##  $ var2 : chr  NA NA "time" NA
##  $ vcov : num  1208.85 2.85 13.86 434.06
##  $ sdcor: num  34.768 1.689 0.236 20.834
```

```
vcov_est_time <- vcov_est_time$vcov
time <- seq(0,10, 2)

var_func_time <- vcov_est_time[2]*time^2 + 2*vcov_est_time[3]*time +
  vcov_est_time[1] + vcov_est_time[4]

vcov_est_age <- data.frame( VarCorr(LMM_age_log_rand_age) )
vcov_est_age <- vcov_est_age$vcov

range( log(Cholst$age_at_meas) )
```

```
## [1] 3.433987 4.276666
```

```
log_age <- seq(3.434, 4.277, length.out = 200)

var_func_age <- vcov_est_age[2]*log_age^2 + 2*vcov_est_age[3]*log_age +
  vcov_est_age[1] + vcov_est_age[4]

par( mfrow = c(2,2))
plot( time, var_func_time, type="l")
plot(exp( log_age) , var_func_age, type="l")

plot( time, sqrt(var_func_time), type="l")
plot(exp( log_age) , sqrt(var_func_age), type="l")
```




```r
LMM_formula <- cholst ~ time + log(age_at_meas) + sex +
  sex*log(age_at_meas) + (1 + time + log(age_at_meas) |ID)
LMM_age_log_rand_timeage <- lmer( formula = LMM_formula , data = Cholst)
```

```
## boundary (singular) fit: see ?isSingular

## Warning: Model failed to converge with 2 negative eigenvalues: -5.5e-03 -1.5e+00
```

```r
VarCorr(LMM_age_log_rand_timeage)
```

```
##  Groups   Name            Std.Dev. Corr
##  ID       (Intercept)      0.0000
##           time             1.6434    NaN
##           log(age_at_meas)  9.1159    NaN 0.138
##  Residual                 20.8334
```

```r
LMM_formula <- cholst ~ time + log(age_at_meas) + sex +
  sex*log(age_at_meas) + (0 + time + log(age_at_meas) |ID)
LMM_age_log_rand_timeage <- lmer( formula = LMM_formula , data = Cholst)
VarCorr(LMM_age_log_rand_timeage)
```

```
##  Groups   Name            Std.Dev. Corr
##  ID       time             1.6434
##           log(age_at_meas)  9.1160  0.138
##  Residual                 20.8334
```

```r
anova(LMM_age_log_rand_timeage,LMM_age_log_rand_time)
```

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

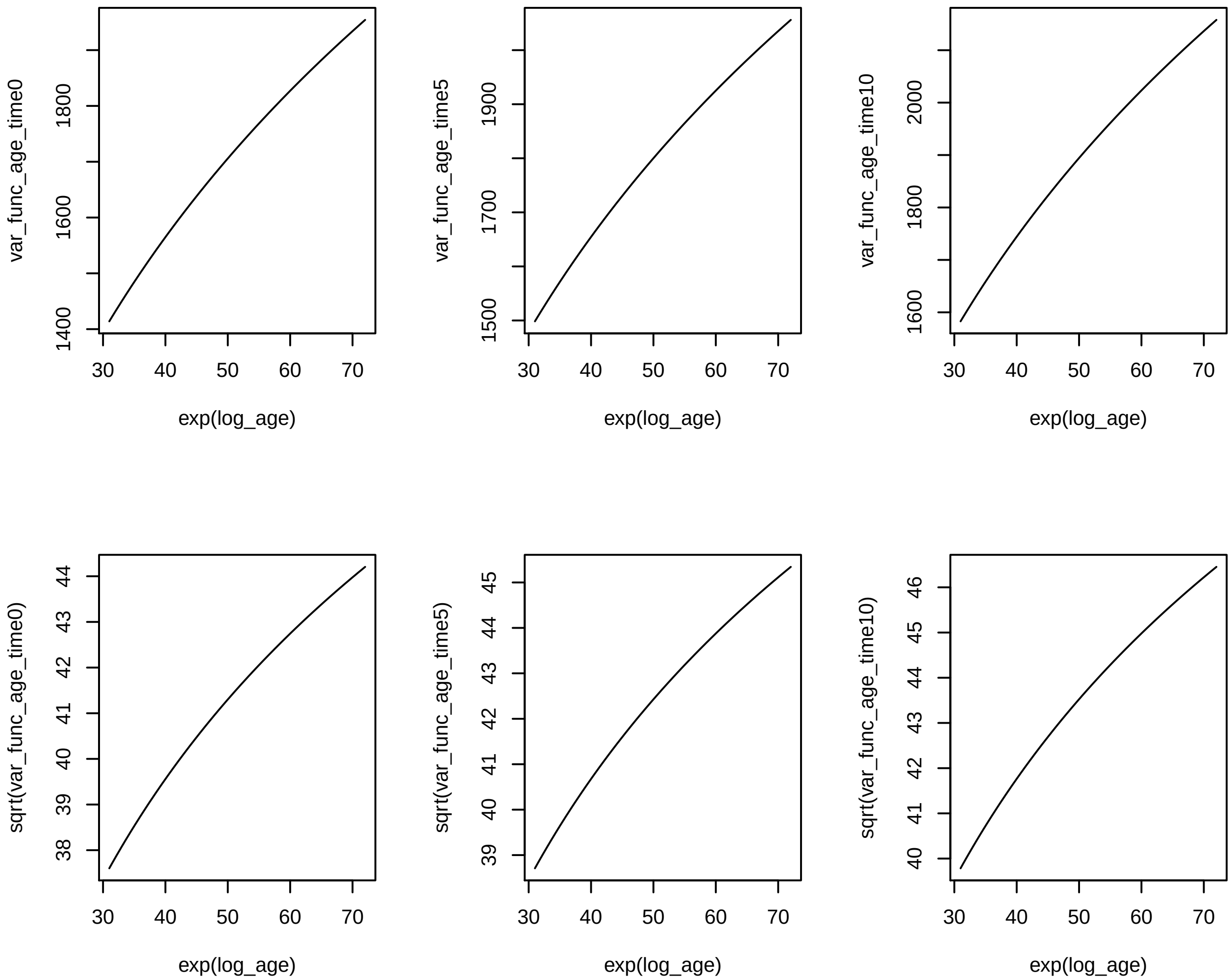
	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMM_age_log_rand_timeage	9	9928.281	9972.838	-4955.140	9910.281	NA	NA	NA
LMM_age_log_rand_time	9	9935.682	9980.239	-4958.841	9917.682	0	0	NA

Let's look at the variance function.

```
vcov_est_timeage <- data.frame( VarCorr(LMM_age_log_rand_timeage) )
vcov_est_timeage
```

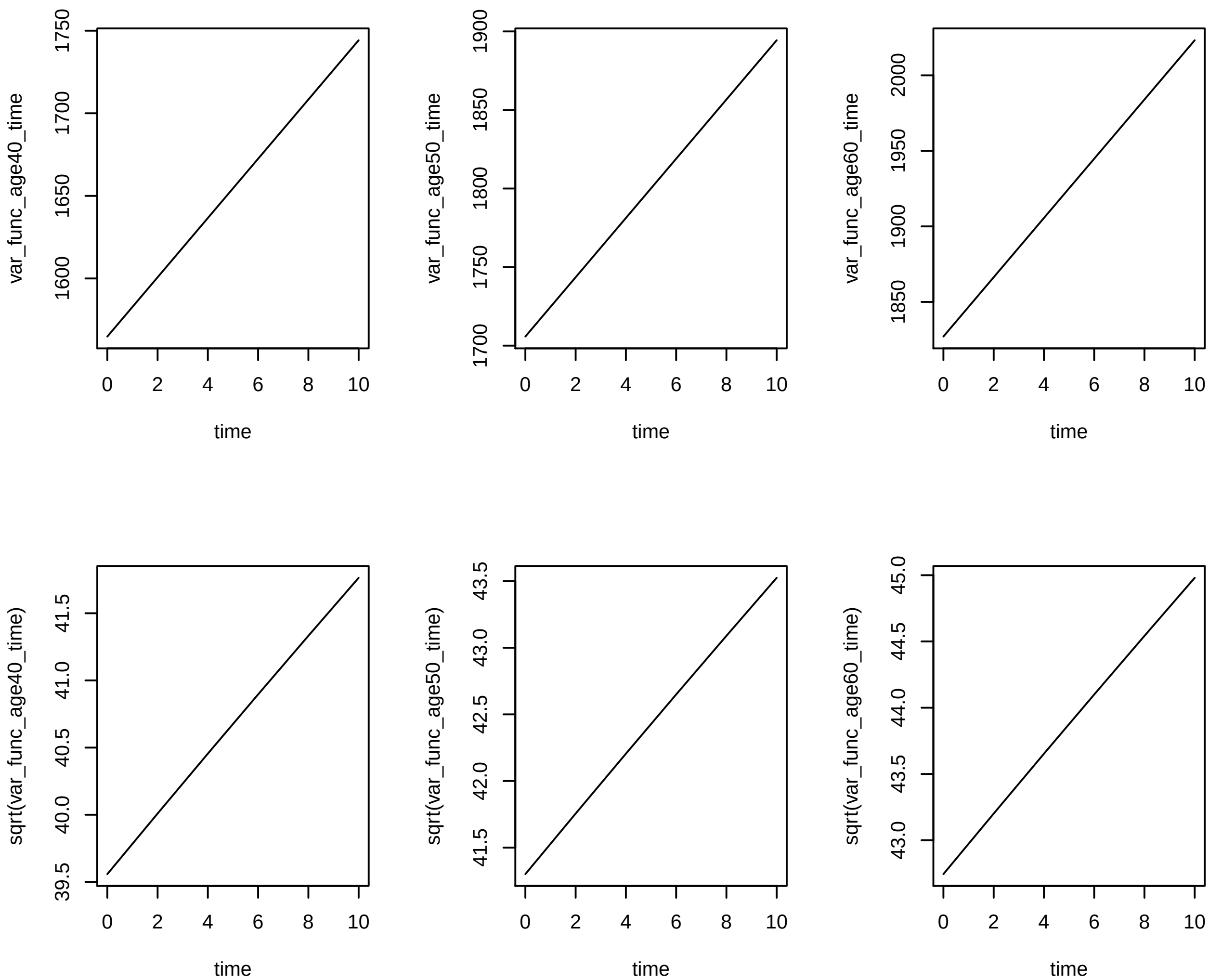
grp	var1	var2	vcov	sdcor
ID	time	NA	2.700901	1.6434419
ID	log(age_at_meas)	NA	83.100829	9.1159656
ID	time	log(age_at_meas)	2.065037	0.1378386
Residual	NA	NA	434.029092	20.8333649

```
vcov_est_timeage <- vcov_est_timeage$vcov

time <- 0
log_age <- seq(3.434, 4.277, length.out = 200)

var_func_age_time0 <- vcov_est_timeage[2]*log_age^2 +
  2*vcov_est_timeage[3]*log_age*time +
  vcov_est_timeage[1]*time + vcov_est_timeage[4]

time <- 5
var_func_age_time5 <- vcov_est_timeage[2]*log_age^2 +
  2*vcov_est_timeage[3]*log_age*time +
  vcov_est_timeage[1]*time + vcov_est_timeage[4]

time <- 10
var_func_age_time10 <- vcov_est_timeage[2]*log_age^2 +
  2*vcov_est_timeage[3]*log_age*time +
  vcov_est_timeage[1]*time + vcov_est_timeage[4]

par(mfrow = c(2,3))
plot( exp( log_age) , var_func_age_time0, type="l")
plot( exp( log_age) , var_func_age_time5, type="l")
plot( exp( log_age) , var_func_age_time10, type="l")

plot( exp( log_age) , sqrt(var_func_age_time0), type="l")
plot( exp( log_age) , sqrt(var_func_age_time5), type="l")
plot( exp( log_age) , sqrt(var_func_age_time10), type="l")
```




```
### Now we'll look at the variance by time.

time <- seq(0,10,2)
log_age <- log(40)

var_func_age40_time <- vcov_est_timeage[2]*log_age^2 +
  2*vcov_est_timeage[3]*log_age*time +
  vcov_est_timeage[1]*time + vcov_est_timeage[4]

log_age <- log(50)
var_func_age50_time <- vcov_est_timeage[2]*log_age^2 +
  2*vcov_est_timeage[3]*log_age*time +
  vcov_est_timeage[1]*time + vcov_est_timeage[4]

log_age <- log(60)
var_func_age60_time <- vcov_est_timeage[2]*log_age^2 +
  2*vcov_est_timeage[3]*log_age*time +
  vcov_est_timeage[1]*time + vcov_est_timeage[4]

par(mfrow = c(2,3))
plot( time , var_func_age40_time, type="l")
plot( time , var_func_age50_time, type="l")
```

```r
plot( time , var_func_age60_time, type="l")

plot( time , sqrt(var_func_age40_time), type="l")
plot( time , sqrt(var_func_age50_time), type="l")
plot( time , sqrt(var_func_age60_time), type="l")
```




```r
data.frame( fixef(LMM_age_log), fixef(LMM_age_log_rand_timeage) )
```

	fixef.LMM_age_log.	fixef.LMM_age_log_rand_timeage.
(Intercept)	58.718554	34.9100917
time	1.083544	0.9538988
log(age_at_meas)	44.338982	50.7562017
sexM	-285.276065	-278.4110593
log(age_at_meas):sexM	73.977934	72.0923021

```r
summary(LMM_age_log_rand_timeage)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
```

```
## Formula: LMM_formula
##    Data: Cholst
##
## REML criterion at convergence: 9887.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.3351 -0.5005 -0.0204  0.5267  3.8899
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  ID       time               2.701   1.643
##           log(age_at_meas)  83.101   9.116   0.14
##  Residual                  434.029  20.833
## Number of obs: 1044, groups:  ID, 200
##
## Fixed effects:
##                        Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)             34.9101    65.0400  281.8480   0.537  0.59187
## time                     0.9539     0.3946  317.8071   2.417  0.01620 *
## log(age_at_meas)        50.7562    17.4697  285.5899   2.905  0.00395 **
## sexM                  -278.4111    64.7146  316.6421  -4.302 2.26e-05 ***
## log(age_at_meas):sexM   72.0923    17.1525  325.7443   4.203 3.41e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) time   lg(__) sexM
## time         0.710
## log(g_t_ms) -0.998 -0.716
## sexM        -0.538 -0.055  0.532
## lg(g_t_m):M  0.542  0.061 -0.539 -0.997
```