

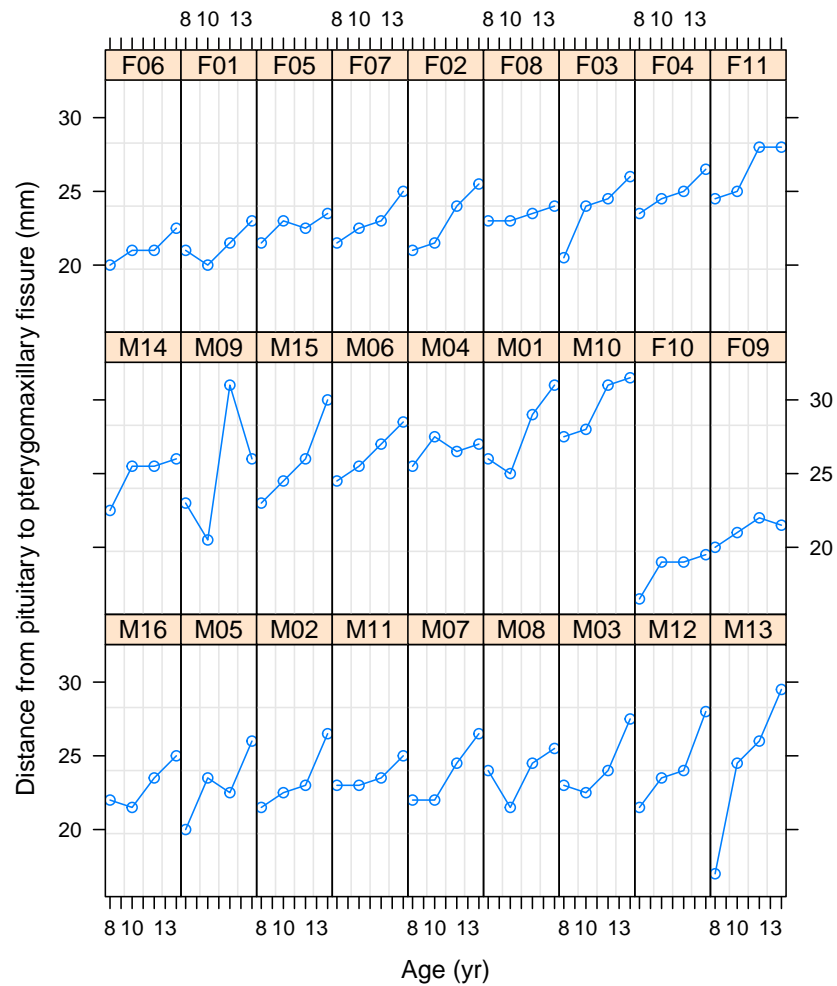
1 Analysis of Longitudinal Data

Q1.1: To study the growth of children, the distance from the pituitary gland to the pterygomaxillary fissure is measured every two years from 8 years of age until 14 years of age. A sample of 27 children — 16 males and 11 females was obtained by orthodontists from x-rays of the children's skulls. (The dataset is denoted by *Orthodont*.) We use this example to illustrate statistical modeling procedures using linear mixed-effects models. A detailed description of the example can be found in the book by Pinheiro and Bates.

- (a) Download the data and fit a simple linear regression model: distance vs age. Interpret the estimates for intercept and slope, comment on the model fit. Is it a good way to analyze such data? Which data aspect was ignored?
- (b) Fit separate linear regression models for each child. To check the variability of estimates across individuals, plot confidence intervals from all models. How well do the confidence intervals overlap? Based on this, we would decide whether we need random effects for the corresponding parameters.
- (c) Using the R package *nlme*, fit a linear mixed effects model with random effects for both intercept and slope. Compare with individual naive *lm* fit and notice that the estimates are more efficient (SE are reduced).
- (d) Using the R package *lmerTest*, fit a linear mixed effects model with random effects for both intercept and slope. Compare the fit results with those obtained from *nlme*.
- (e) Using *anova()* command compare the LME model, fitted by either of the two ways above, with the *lm* model without random effects. Check AIC and BIC values to conclude which model fits better. What is the p-value for the likelihood ratio test?
- (f) To test the difference between boys and girls, fit a model with the *sex*age* interaction. Use the *anova()* command to see whether the interaction is significant.

Hint: The following code may be useful:

```
library(nlme) # get the NLME library
data(Orthodont)
#head(Orthodont)
plot(Orthodont)
```



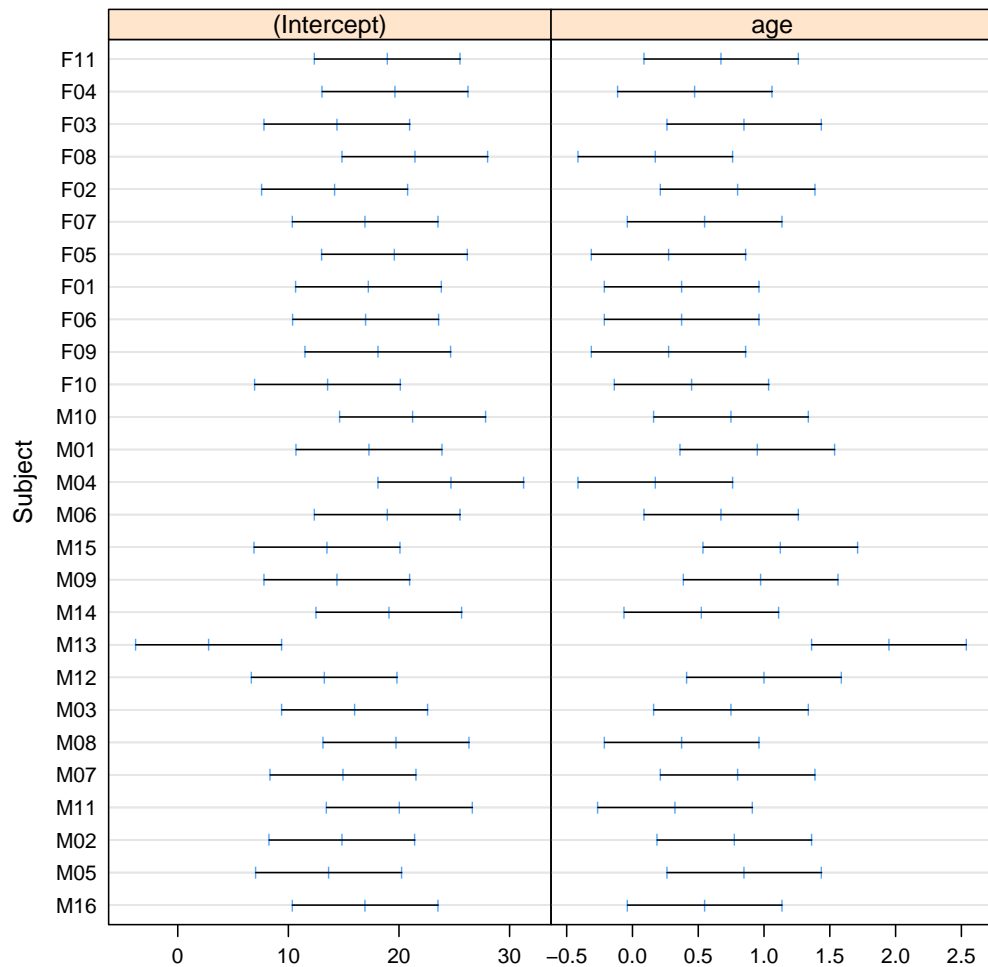
```
# A simple regression model:
fm1Orth.lm <- lm(distance ~ age, Orthodont)
summary(fm1Orth.lm)

##
## Call:
## lm(formula = distance ~ age, data = Orthodont)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.5037 -1.5778 -0.1833  1.3519  6.3167
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  16.7611     1.2256  13.676 < 2e-16 ***
## age           0.6602     0.1092   6.047 2.25e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2.537 on 106 degrees of freedom
## Multiple R-squared:  0.2565, Adjusted R-squared:  0.2495
## F-statistic: 36.56 on 1 and 106 DF,  p-value: 2.248e-08

## This model explains only about 26 percent variability in the data
## The model intercept = 16.76 is the average distance at the age of
## zero years, which does not have a meaningful interpretation
## in the context.
## The slope shows that as a child gets one year older,
## the distance increases, on average, by 0.66 units.

## The simple linear regression model ignores the fact that measurements
## taken for the same child are not independent and may be highly
## correlated with each other.
## That is why, it is better to use a model which accounts
## for that.
## For longitudinal data, one can start with fitting a linear regression
## model for each subject separately:
fm1Orth.lis <- lmList(distance ~ age | Subject, Orthodont)
## To plot confidence intervals:
plot(intervals(fm1Orth.lis))
```



Not all confidence intervals overlap, and there is
 ## a lot of variability in the individual estimates of both intercept
 ## and slope. This indicates, that fitting a model
 ## with random intercept and, possibly, random slope
 ## may give a better description of the data.

```
## Using R package nlme
library(nlme)
fm1Orth.lme <- lme(distance~age, data=Orthodont,
                  random = ~age | Subject, method = "ML")
summary(fm1Orth.lme)
```

```
## Linear mixed-effects model fit by maximum likelihood
## Data: Orthodont
##      AIC      BIC    logLik
## 451.2116 467.3044 -219.6058
##
## Random effects:
```

```

## Formula: ~age | Subject
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.1940996 (Intr)
## age         0.2149244 -0.581
## Residual    1.3100399
##
## Fixed effects: distance ~ age
##           Value Std.Error DF   t-value p-value
## (Intercept) 16.761111 0.7678975 80 21.827278      0
## age         0.660185 0.0705779 80  9.353997      0
## Correlation:
##   (Intr)
## age -0.848
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -3.305969305 -0.487429621  0.007597971  0.482237071  3.922789749
##
## Number of Observations: 108
## Number of Groups: 27

## Using R package lmerTest

library(lmerTest)

## Lade nötiges Paket: lme4
## Lade nötiges Paket: Matrix
##
## Attache Paket: 'lme4'
## Das folgende Objekt ist maskiert 'package:nlme':
##
##   lmList
##
## Attache Paket: 'lmerTest'
## Das folgende Objekt ist maskiert 'package:lme4':
##
##   lmer
## Das folgende Objekt ist maskiert 'package:stats':
##
##   step

fm1Orth.lme.1 <- lmer(distance~age + (1+age|Subject), data=Orthodont)
summary(fm1Orth.lme.1)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: distance ~ age + (1 + age | Subject)
## Data: Orthodont
##
## REML criterion at convergence: 442.6
##
## Scaled residuals:

```

```

##      Min      1Q  Median      3Q      Max
## -3.2229 -0.4938  0.0073  0.4721  3.9161
##
## Random effects:
## Groups   Name                Variance Std.Dev. Corr
## Subject (Intercept)  5.41660  2.3274
##          age          0.05128  0.2264  -0.61
## Residual                1.71616  1.3100
## Number of obs: 108, groups: Subject, 27
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept) 16.76111      0.77528 25.99907  21.620 < 2e-16 ***
## age         0.66019      0.07126 25.99794   9.265 1.01e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## age -0.848

AIC(fm1Orth.lme); AIC(fm1Orth.lm)

## [1] 451.2116
## [1] 511.577

BIC(fm1Orth.lme); BIC(fm1Orth.lm)

## [1] 467.3044
## [1] 519.6234

# To compare the LME model with the lm model without random effects
anova(fm1Orth.lme, fm1Orth.lm)

##              Model df      AIC      BIC    logLik    Test  L.Ratio p-value
## fm1Orth.lme      1  6 451.2116 467.3044 -219.6058
## fm1Orth.lm       2  3 511.5770 519.6234 -252.7885 1 vs 2 66.36537 <.0001

# The LME model fits much better than the LM model
# since it has smaller AIC and BIC values
# and the LRT shows significant difference!

# test for diff between boys and girls
fm2Orth.lme <- update(fm1Orth.lme, fixed=distance~Sex*age)
summary(fm2Orth.lme)

## Linear mixed-effects model fit by maximum likelihood
##   Data: Orthodont
##      AIC      BIC    logLik
##  443.806 465.263 -213.903
##
## Random effects:
## Formula: ~age | Subject
## Structure: General positive-definite, Log-Cholesky parametrization

```

```

##              StdDev   Corr
## (Intercept)  2.134688 (Intr)
## age          0.154139 -0.603
## Residual     1.310040
##
## Fixed effects: distance ~ Sex + age + Sex:age
##              Value Std.Error DF   t-value p-value
## (Intercept)  16.340625 0.9987521 79 16.361042  0.0000
## SexFemale    1.032102 1.5647438 25  0.659598  0.5155
## age          0.784375 0.0843294 79  9.301322  0.0000
## SexFemale:age -0.304830 0.1321188 79 -2.307238  0.0237
## Correlation:
##              (Intr) SexFml age
## SexFemale    -0.638
## age          -0.880  0.562
## SexFemale:age  0.562 -0.880 -0.638
##
## Standardized Within-Group Residuals:
##              Min           Q1           Med           Q3           Max
## -3.33603363 -0.41539842  0.01039175  0.49169520  3.85819292
##
## Number of Observations: 108
## Number of Groups: 27

# There seems to be a significant difference between boys and girls

# Let's do ANOVA test for comparing models
anova(fm1Orth.lme, fm2Orth.lme)

##              Model df          AIC          BIC      logLik   Test  L.Ratio p-value
## fm1Orth.lme      1  6 451.2116 467.3044 -219.6058
## fm2Orth.lme      2  8 443.8060 465.2630 -213.9030 1 vs 2 11.40565  0.0033

```