Modelos Mistos - Exemplo: Orthodont

Vanderly Janeiro

Dep. de Estatística - UEM

Contents

Modelo Linear de Efeitos Mistos (LME)

Exemplo: Dados do estudo ortodôntico de Potthoff e Roy (1964)

Os pesquisadores estavam interessados no desenvolvimento das crianças ao longo do tempo.[1]. Os dados foram coletados por ortodontistas a partir de radiografias de crânios das crianças. A hipófise e a fissura pterigomaxilar são pontos facilmente localizados nas radiografias. Eles coletaram medidas de crescimento dentário, a distância (mm) do centro da glândula pituitária à fissura pterigomaxilar em 27 crianças (11 meninas e 16 meninos) aos 8, 10, 12 e 14 anos.[2]

[1] Source: Potthoff, R. F. and Roy, S. N. (1964), "A generalized multivariate analysis of variance model useful especially for growth curve problems", Biometrika, 51, 313–326.

[2] Ver Seção 1.4.1 de Pinheiro, J.C., and Bates, D.M. (2000)

```
library(nlme)
```

```
psych::headTail(Orthodont)
```

```
##     distance age Subject    Sex
## 1         26   8     M01   Male
## 2         25  10     M01   Male
## 3         29  12     M01   Male
## 4         31  14     M01   Male
## ...      ... ...    <NA>   <NA>
## 105     24.5   8     F11 Female
## 106       25  10     F11 Female
## 107       28  12     F11 Female
## 108       28  14     F11 Female
```

```
names(Orthodont)
```

```
## [1] "distance" "age"      "Subject"  "Sex"
```

```
levels(Orthodont$Sex)
```

```
## [1] "Male"   "Female"
```

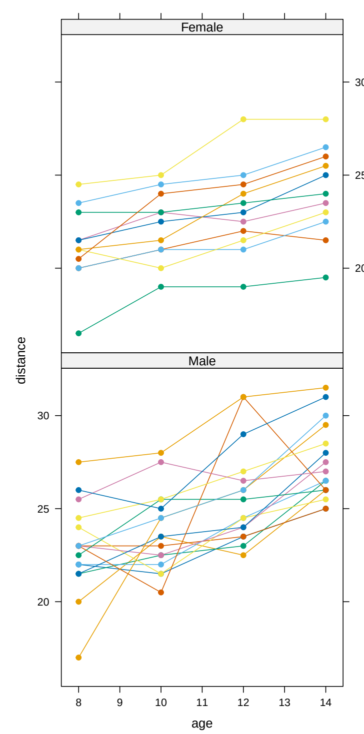


Figure 1: Crescimento dentário (mm) para meninos e meninas

```
Orthodont$id<- as.factor(rep(1:27, each=4))
```

```
library(latticeExtra)
#xyplot(distance ~ age | Sex, data=Orthodont, type="b",
#       groups = Subject,  pch=19)
```

Os perfis por nível de índividuo são mostrados na Figura 1. Parece que *cada índividuo* tem aproximadamente uma *tendência linear* com possivelmente *diferentes interceptos e inclinações*.

```
plot(Orthodont)
```

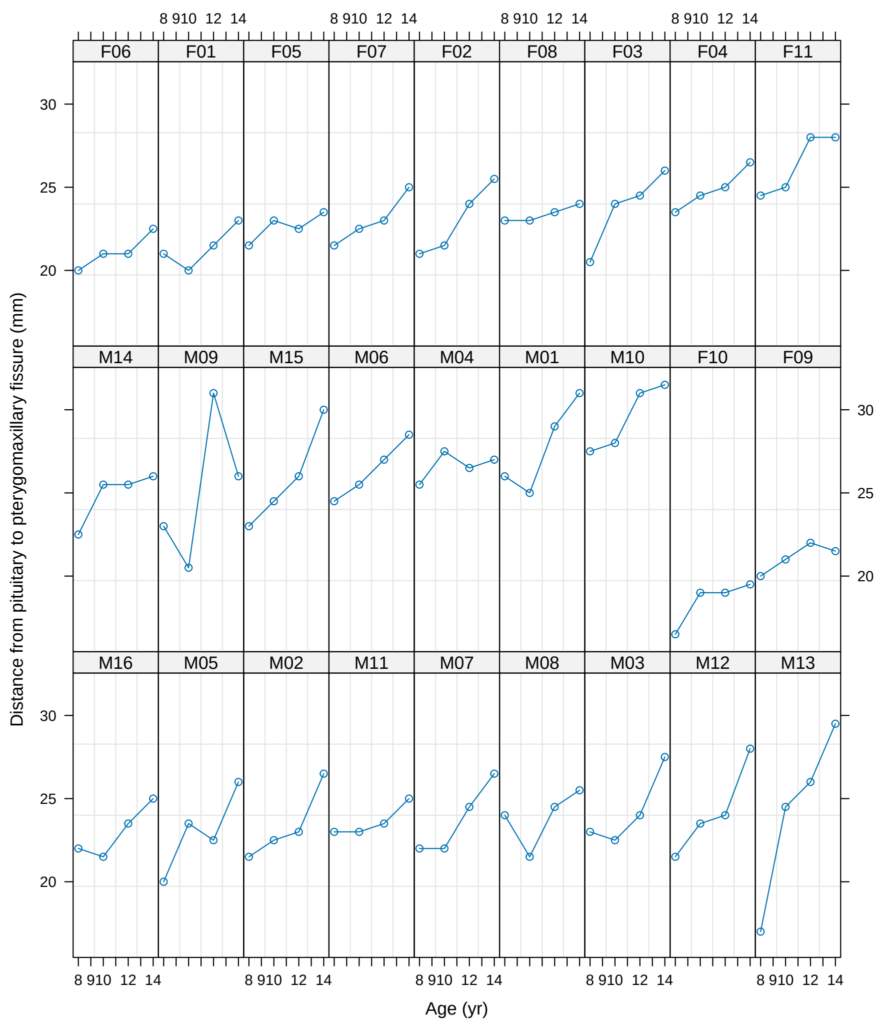


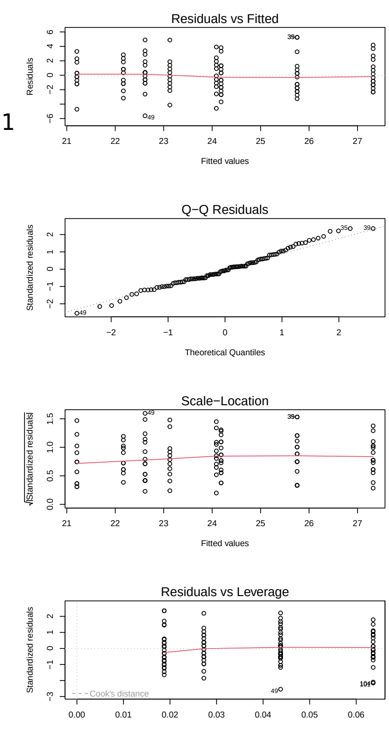
Figure 2: Crescimento dentário (mm) - perfis por indivíduo

Os perfis de nível de sujeito são mostrados na Figura 1. Parece que *cada indivíduo* tem aproximadamente uma *tendência linear* com possivelmente *diferentes intercepto e coeficientes de inclinação*. Vamos agora discutir o modelo (naive):

$$y_{ij} = \beta_0 + \beta_1 age + \beta_2 Sex_F + \beta_3 age : Sex_F + \varepsilon_{ij}$$

```
lm.Ort <- lm(distance ~ age + Sex + age:Sex,
data = Orthodont)
summary(lm.Ort)
```

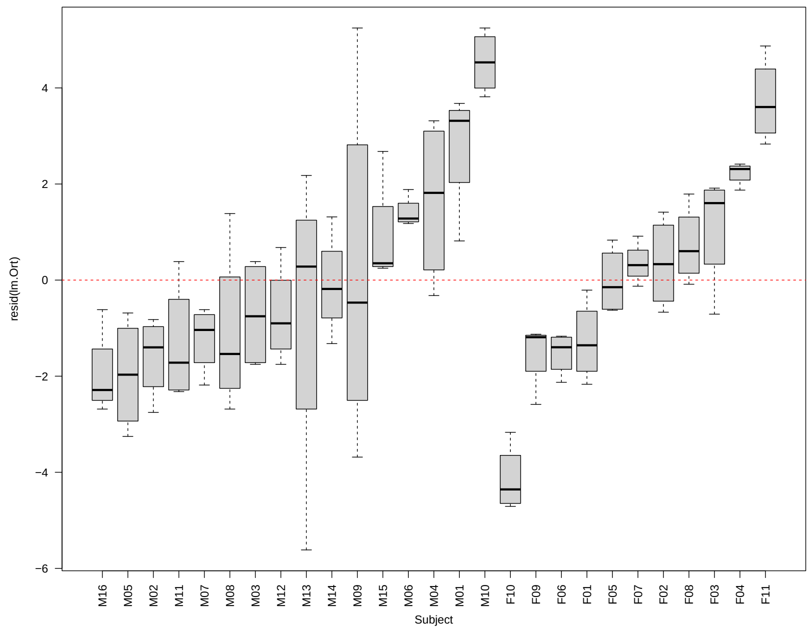
```
##
## Call:
## lm(formula = distance ~ age + Sex + age:Sex, data = Orthodont)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.6156 -1.3219 -0.1682  1.3299  5.2469
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    16.3406     1.4162  11.538  < 2e-16 ***
## age             0.7844     0.1262   6.217 1.07e-08 ***
## SexFemale       1.0321     2.2188   0.465    0.643
## age:SexFemale  -0.3048     0.1977  -1.542    0.126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.257 on 104 degrees of freedom
## Multiple R-squared:  0.4227, Adjusted R-squared:  0.4061
## F-statistic: 25.39 on 3 and 104 DF,  p-value: 2.108e-12
```



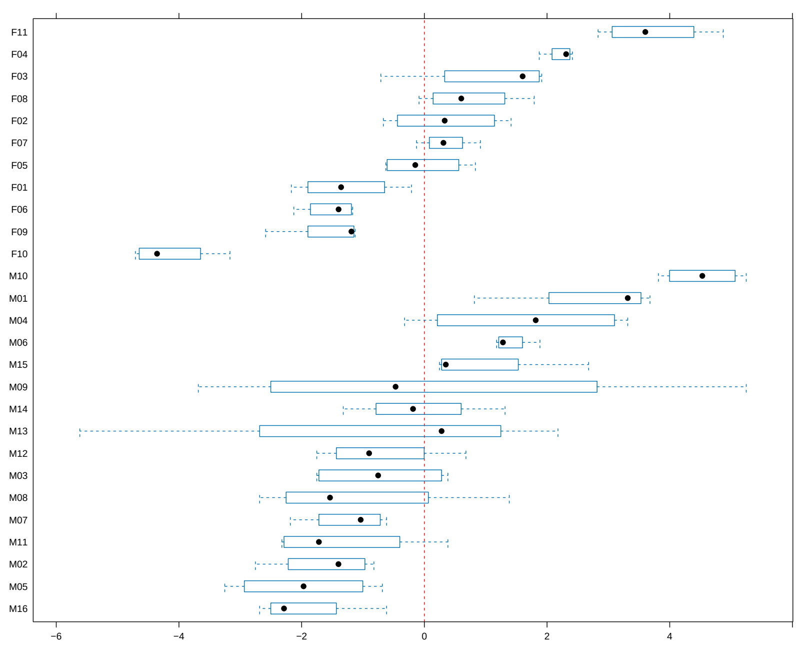
Podemos utilizar este modelo para fazer inferências? Como estão as suposições do modelo?

Observando estes gráficos de resíduo não detectamos nada grave em relação as suposições para o modelo usado. Vamos investigar os resíduos em relação a cada indivíduo.

```
boxplot(resid(lm.Ort) ~ Subject,data = Orthodont,las = 2)
abline(h = 0, col="red", lty="dashed")
```



```r
bwplot(getGroups(Orthodont) ~ resid(lm.Ort),
       panel=function(...) {
         panel.abline(v=0, col="red", lty="dashed")
         panel.bwplot(...)})
```



Pode ser observado que, para muitos indíviduos, os resíduos não estão em torno de zero. Em alguns casos o modelo subestimou as observações do indivíduo e em outros superestimou. Será que incluindo indivíduo no modelo o problema estara resolvido?

```
lm.Ort.2 <- lm(distance ~ id + age * Sex,
data = Orthodont)

summary(lm.Ort.2)

##
## Call:
## lm(formula = distance ~ id + age * Sex, data = Orthodont)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.8969 -0.6151 -0.0212  0.6141  5.0906
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   19.1219     1.0988  17.403  < 2e-16 ***
## id2           -4.3750     0.9803  -4.463 2.65e-05 ***
## id3           -3.5000     0.9803  -3.570 0.000611 ***
## id4           -1.1250     0.9803  -1.148 0.254604
## id5           -4.7500     0.9803  -4.845 6.19e-06 ***
## id6           -1.3750     0.9803  -1.403 0.164653
## id7           -4.0000     0.9803  -4.080 0.000107 ***
## id8           -3.8750     0.9803  -3.953 0.000167 ***
## id9           -2.6250     0.9803  -2.678 0.009014 **
## id10           1.7500     0.9803   1.785 0.078078 .
## id11          -4.1250     0.9803  -4.208 6.76e-05 ***
## id12          -3.5000     0.9803  -3.570 0.000611 ***
## id13          -3.5000     0.9803  -3.570 0.000611 ***
## id14          -2.8750     0.9803  -2.933 0.004395 **
## id15          -1.8750     0.9803  -1.913 0.059418 .
## id16          -4.7500     0.9803  -4.845 6.19e-06 ***
## id17          -3.0219     1.6568  -1.824 0.071943 .
## id18          -1.3969     1.6568  -0.843 0.401705
## id19          -0.6469     1.6568  -0.390 0.697262
## id20           0.4781     1.6568   0.289 0.773654
## id21          -1.7719     1.6568  -1.069 0.288116
## id22          -3.2719     1.6568  -1.975 0.051781 .
## id23          -1.3969     1.6568  -0.843 0.401705
## id24          -1.0219     1.6568  -0.617 0.539152
## id25          -3.2719     1.6568  -1.975 0.051781 .
## id26          -5.8969     1.6568  -3.559 0.000633 ***
```

```
## id27            1.9781     1.6568   1.194 0.236069
## age             0.7844     0.0775  10.121 6.44e-16 ***
## SexFemale           NA         NA      NA       NA
## age:SexFemale  -0.3048     0.1214  -2.511 0.014097 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.386 on 79 degrees of freedom
## Multiple R-squared:  0.8345, Adjusted R-squared:  0.7759
## F-statistic: 14.23 on 28 and 79 DF,  p-value: < 2.2e-16
```

Usando este modelo houve uma melhora no $R^2_{ajustado}$ passando de 0.4061 para 0.7759, mas podemos confiar neste modelo? Observe que o modelo não pode estimar o efeito de *SexFemale*. Parece o caso de um modelo não identificável.

Este modelo contém parâmetros em excesso, e não estamos interessados em todos esses efeitos, por exemplo: qual o interesse no efeido do indivíduo *id*27? Porém, o modelo, pode nos mostrar que cada indívíduo possue coeficiente de inclinação diferente e a inclusão do indivíduo causou diminuição no erro (de $2.257/104 = 0.02170192$ para $1.386/79 = 0.0175443$)

Seria melhor propor um modelo que:[3]

- controle cada "indivíduo" sem estimar todos esses parâmetros, ou seja, remover o ruído devido a indíviduo com um *baixo custo*;
- evitar o problema de rank-deficiência do modelo quando fatores aninhados estão presentes.

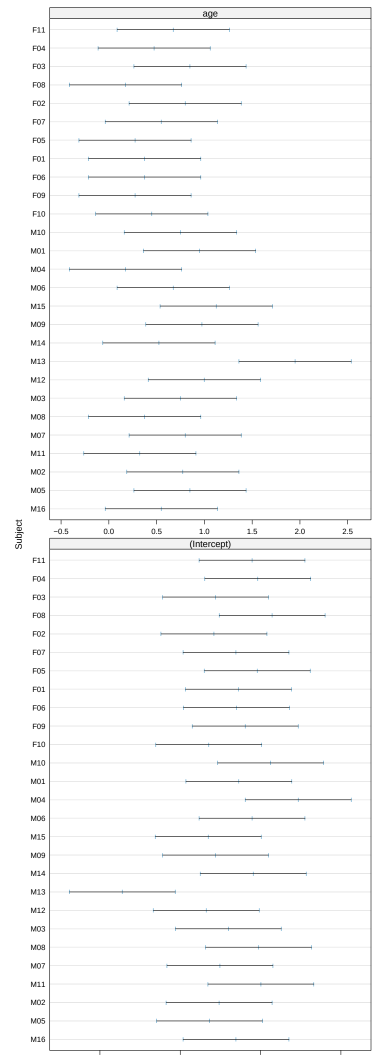
Uma boa solução é ajustar um modelo linear de efeitos mistos aos dados. Vantagens de usar LMMs:

- você pode controlar o efeito de "indivíduo" às custas de estimar um único parâmetro (ou seja, variância de indivíduo)
- o modelo não é mais *rank-deficiente*.

[3] rank-deficiência: neste contexto indica que não há informações suficientes contidas nos dados para estimar o modelo desejado

Antes de implementar um modelo misto, vamos obter um modelo linear independente para cada indivíduo.[4]

[4] Notamos que os intervalos para as interceptações têm a mesma largura, assim como os intervalos para a inclinação em relação à idade. Esta é uma consequência de ter dados balanceados; ou seja, todos os sujeitos foram observados o mesmo número de vezes e nas mesmas idades.




```
fm1Orth.lis <- lmList(distance ~ age, data = Orthodont)
fm1Orth.lis
```

```
## Call:
##   Model: distance ~ age | Subject
##    Data: Orthodont
##
## Coefficients:
##     (Intercept)   age
## M16       16.95 0.550
## M05       13.65 0.850
## M02       14.85 0.775
## M11       20.05 0.325
## M07       14.95 0.800
## M08       19.75 0.375
## M03       16.00 0.750
## M12       13.25 1.000
## M13        2.80 1.950
## M14       19.10 0.525
## M09       14.40 0.975
## M15       13.50 1.125
## M06       18.95 0.675
## M04       24.70 0.175
## M01       17.30 0.950
## M10       21.25 0.750
## F10       13.55 0.450
## F09       18.10 0.275
## F06       17.00 0.375
## F01       17.25 0.375
## F05       19.60 0.275
## F07       16.95 0.550
## F02       14.20 0.800
## F08       21.45 0.175
## F03       14.40 0.850
## F04       19.65 0.475
## F11       18.95 0.675
##
## Degrees of freedom: 108 total; 54 residual
## Residual standard error: 1.31004
```

```
summary(fm1Orth.lis)
```

```
## Call:
```

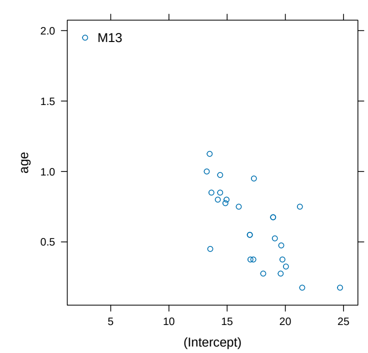
```
##   Model: distance ~ age | Subject
##    Data: Orthodont
##
## Coefficients:
##    (Intercept)
##     Estimate Std. Error   t value     Pr(>|t|)
## M16    16.95   3.288173 5.1548379 3.695247e-06
## M05    13.65   3.288173 4.1512411 1.181678e-04
## M02    14.85   3.288173 4.5161854 3.458934e-05
## M11    20.05   3.288173 6.0976106 1.188838e-07
## M07    14.95   3.288173 4.5465974 3.116705e-05
## M08    19.75   3.288173 6.0063745 1.665712e-07
## M03    16.00   3.288173 4.8659237 1.028488e-05
## M12    13.25   3.288173 4.0295930 1.762580e-04
## M13     2.80   3.288173 0.8515366 3.982319e-01
## M14    19.10   3.288173 5.8086964 3.449588e-07
## M09    14.40   3.288173 4.3793313 5.509579e-05
## M15    13.50   3.288173 4.1056231 1.373664e-04
## M06    18.95   3.288173 5.7630783 4.078189e-07
## M04    24.70   3.288173 7.5117696 6.081644e-10
## M01    17.30   3.288173 5.2612799 2.523621e-06
## M10    21.25   3.288173 6.4625549 3.065505e-08
## F10    13.55   3.288173 4.1208291 1.306536e-04
## F09    18.10   3.288173 5.5045761 1.047769e-06
## F06    17.00   3.288173 5.1700439 3.499774e-06
## F01    17.25   3.288173 5.2460739 2.665260e-06
## F05    19.60   3.288173 5.9607565 1.971127e-07
## F07    16.95   3.288173 5.1548379 3.695247e-06
## F02    14.20   3.288173 4.3185072 6.763806e-05
## F08    21.45   3.288173 6.5233789 2.443813e-08
## F03    14.40   3.288173 4.3793313 5.509579e-05
## F04    19.65   3.288173 5.9759625 1.863600e-07
## F11    18.95   3.288173 5.7630783 4.078189e-07
##    age
##     Estimate Std. Error   t value     Pr(>|t|)
## M16    0.550  0.2929338 1.8775576 6.584707e-02
## M05    0.850  0.2929338 2.9016799 5.361639e-03
## M02    0.775  0.2929338 2.6456493 1.065760e-02
## M11    0.325  0.2929338 1.1094659 2.721458e-01
## M07    0.800  0.2929338 2.7309929 8.511442e-03
## M08    0.375  0.2929338 1.2801529 2.059634e-01
## M03    0.750  0.2929338 2.5603058 1.328807e-02
## M12    1.000  0.2929338 3.4137411 1.222240e-03
## M13    1.950  0.2929338 6.6567951 1.485652e-08
```

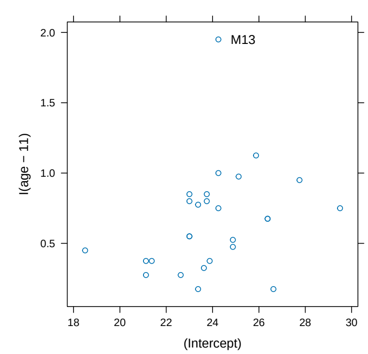
```
## M14    0.525  0.2929338 1.7922141 7.870160e-02
## M09    0.975  0.2929338 3.3283976 1.577941e-03
## M15    1.125  0.2929338 3.8404587 3.247135e-04
## M06    0.675  0.2929338 2.3042752 2.508117e-02
## M04    0.175  0.2929338 0.5974047 5.527342e-01
## M01    0.950  0.2929338 3.2430540 2.030113e-03
## M10    0.750  0.2929338 2.5603058 1.328807e-02
## F10    0.450  0.2929338 1.5361835 1.303325e-01
## F09    0.275  0.2929338 0.9387788 3.520246e-01
## F06    0.375  0.2929338 1.2801529 2.059634e-01
## F01    0.375  0.2929338 1.2801529 2.059634e-01
## F05    0.275  0.2929338 0.9387788 3.520246e-01
## F07    0.550  0.2929338 1.8775576 6.584707e-02
## F02    0.800  0.2929338 2.7309929 8.511442e-03
## F08    0.175  0.2929338 0.5974047 5.527342e-01
## F03    0.850  0.2929338 2.9016799 5.361639e-03
## F04    0.475  0.2929338 1.6215270 1.107298e-01
## F11    0.675  0.2929338 2.3042752 2.508117e-02
##
## Residual standard error: 1.31004 on 54 degrees of freedom
```

```
par(mfrow=c(1,2))
pairs(fm1Orth.lis, id = 0.01, adj = -0.5)


pairs(fm2Orth.lis, id = 0.01, adj = -0.5)


layout(1)
```





Finalmente, vemos que o padrão nos indivíduos, os intervalos dos interceptos é quase um reflexo do padrão dos intervalos dos coeficientes de inclinação. Segundo Pinheiro e Bates (2000)[5], isso ocorre porque todos os dados foram coletados entre as idades de 8 e 14 anos, mas o intercepto representa uma distância na idade de 0 (zero). A extrapolação de volta para a idade de 0 resultará em uma alta correlação negativa (cerca de -0.98) entre as estimativas dos coeficentes de inclinação e sua estimativa de intercepto correspondente.

Para removermos isso ajustamos a distância como uma função linear da $(age - 11)$, de modo que os dois coeficientes sendo estimados são a distância aos 11 anos.

[5] Pinheiro e Bates (2000), Seção 4.1.1 - pág. 141

```
fm2Orth.lis <- update(fm1Orth.lis, distance ~ I(age - 11))
plot(intervals(fm2Orth.lis))
```

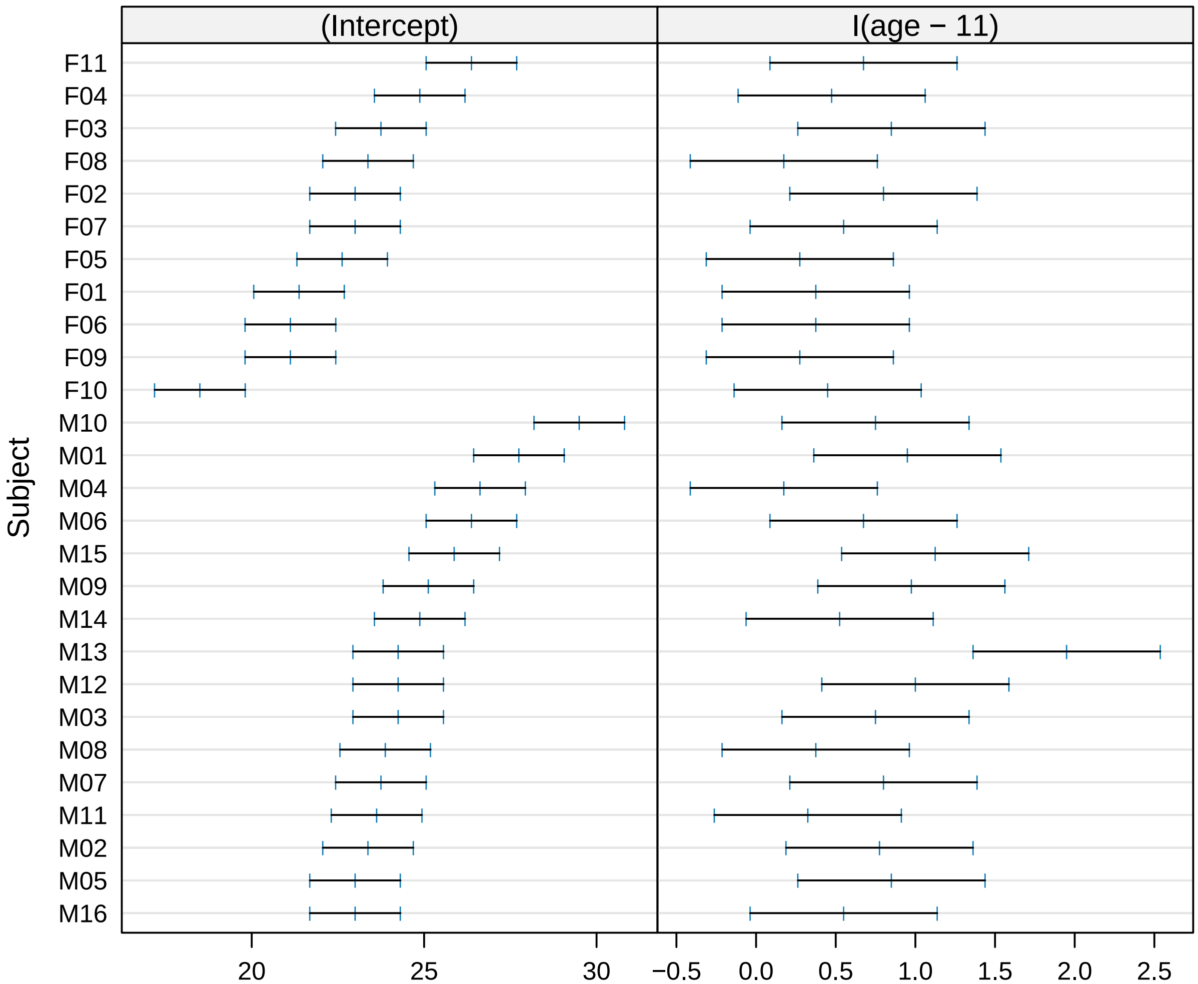


Figure 3: Estimativas para intercepto e coeficiente angular considerando (age - 11)

Para continuar com a análise desses dados, poderíamos ajustar um modelo de regressão aos dados centralizados com uma taxa de crescimento comum, mas interceptos diferentes para cada indivíduo.

Especificação de um modelo geral

Um modelo geral é dado por:

$$\begin{aligned} y_{ti} = \beta_0 + \beta_1(age_{ti} - 11) + \beta_2 Sex_{Fi} + \beta_3(age_{ti} - 11) \times Sex_{Fi} + & \Big\} \text{ Fixo} \\ u_{0i} + u_{1i} + \varepsilon_{ti} & \Big\} \text{ aleatório} \end{aligned} \tag{1}$$

com a distância (y_{ti}) sendo o resultado na idade $6 + 2t (t = 1, \ldots, 4)$ para a i-ésima criança $(i = 1, \ldots, 27)$ e sex_i é o sexo da criança.

Usando esta especificação de modelo, notamos:

- O termo u_{0i} representa a intercepto aleatória.
- O termo u_{1i} representa a inclinação aleatória (efeito aleatório associado com a inclinação para a criança i).

Assumimos que a distribuição dos efeitos aleatórios associados à criança i, u_{0i} e u_{1i} é bivariada normal:

$$\boldsymbol{u}_i = \begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \overset{iid}{\sim} N_2(\boldsymbol{0}, \boldsymbol{D}) \qquad \boldsymbol{D} = \begin{pmatrix} \sigma^2_{int} & \sigma_{int,inc} \\ \sigma_{int,inc} & \sigma^2_{inc} \end{pmatrix}$$

e

$$\boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \end{pmatrix} \overset{iid}{\sim} N_4(\boldsymbol{0}, \boldsymbol{R}_i)$$

e podemos considerar duas diferentes estruturas de covariância dentro da unidade:

- $\boldsymbol{R}_i = cov(\boldsymbol{e}_i) = \sigma^2 \boldsymbol{I}_4$ for both groups
- $\boldsymbol{R}_i = cov(\boldsymbol{e}_i) = \sigma^2_{male} \boldsymbol{I}_4$ para garotas e $\boldsymbol{R}_i = cov(\boldsymbol{e}_i) = \sigma^2_{female} \boldsymbol{I}_4$ para garotos.

Forma matricial do modelo para as observações da i-ésima criança:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{u}_i + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots, 27$$

sendo

$$\boldsymbol{Y}_i = \begin{pmatrix} Y_{1i} \\ Y_{2i} \\ Y_{3i} \\ Y_{4i} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}$$

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & -3 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 \end{pmatrix} \quad \text{ou} \quad \boldsymbol{X}_i = \begin{pmatrix} 1 & -3 & 1 & -3 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 3 & 1 & 3 \end{pmatrix}$$

O primeiro se refere ao masculino.
Os fatores estao a favor do sexo femino.
Impacta nas colunas 3 e 4 dos X

se a i-ésima criança é masculina ou feminina, respectivamante. Para os termos aleatórios temos:

$$\boldsymbol{Z}_i = \begin{pmatrix} 1 & -3 \\ 1 & -1 \\ 1 & 1 \\ 1 & 3 \end{pmatrix}, \quad \boldsymbol{u}_i = \begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \quad \text{e} \quad \boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \end{pmatrix}$$

em que $\boldsymbol{u}_i \overset{iid}{\sim} N_2(\boldsymbol{0}, \boldsymbol{D})$ e $\boldsymbol{\varepsilon}_i \overset{iid}{\sim} N_4(\boldsymbol{0}, \boldsymbol{R}_i)$, com $\boldsymbol{u}_i$ e $\boldsymbol{\varepsilon}_i$ independentes entre si.

Reescrevendo o modelo na forma matricial para dodas crianças:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\varepsilon}$$

em que (com $n = 27 \times 4 = 108$)

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_{27} \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_{27} \end{pmatrix}, \quad \boldsymbol{Z} = \begin{pmatrix} \boldsymbol{Z}_1 & & & \\ & \boldsymbol{Z}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{Z}_{27} \end{pmatrix}$$

$$\boldsymbol{u} = \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_{27} \end{pmatrix}, \quad \text{e} \quad \boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \end{pmatrix}$$

Temos então que, $\boldsymbol{Y}$ é um vetor $(108 \times 1$, $\boldsymbol{X}$ uma matrix $(108 \times 4$, $\boldsymbol{Z})$ uma matriz (108×54), $\boldsymbol{u}$ um vetor (54×1) e $\boldsymbol{\varepsilon}$ um vetor (108×1). Dessa forma:

$$\boldsymbol{u} \overset{iid}{\sim} N_2(\boldsymbol{0}, \boldsymbol{G}) \quad \text{e} \quad \boldsymbol{\varepsilon} \overset{iid}{\sim} N(\boldsymbol{0}, \boldsymbol{R})$$

em que

$$G = \begin{pmatrix} D & & & \\ & D & & \\ & & \ddots & \\ & & & D \end{pmatrix}, \quad \text{e} \quad R = \begin{pmatrix} R_1 & & & \\ & R_2 & & \\ & & \ddots & \\ & & & R_{27} \end{pmatrix}. \quad (2)$$

Tal como Pinheiro e Bates (pág. 142), nós ajustamos o modelo

```
fm2Orth.lis <- update(fm1Orth.lis, distance ~ I(age - 11))
```

O ajuste de um modelo linear misto simples para os dados `Othodont`:[6]

[6] Apresentado por Pinheiro e Bates na seção 4.2.1 (pág. 146)

```
fm1Orth.lme <- lme( distance ~ I(age-11), data = Orthodont,
               random = ~ I(age-11) | Subject )
summary(fm1Orth.lme)
```

yij = B0 + B1(Age-11) [Parte fixa] +
U0i + U1i + Eij [Parte aleatoria]

random = ~ 1 | Subject -> Efeito apenas intercepto
random = ~ I(Age - 11) | Subject -> Efeito no intercepto e efeito de inclinacao

```
## Linear mixed-effects model fit by REML
##   Data: Orthodont
##        AIC      BIC    logLik
##   454.6367 470.6173 -221.3183
##
## Random effects:
##  Formula: ~I(age - 11) | Subject
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 2.1343289 (Intr)
## I(age - 11) 0.2264278 0.503
## Residual    1.3100402
##
## Fixed effects:  distance ~ I(age - 11)
##                 Value Std.Error DF  t-value p-value
## (Intercept) 24.023148 0.4296601 80 55.91198       0
## I(age - 11)  0.660185 0.0712533 80  9.26533       0
##  Correlation:
##             (Intr)
## I(age - 11) 0.294
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -3.223106868 -0.493760901  0.007316482  0.472151218  3.916031759
##
## Number of Observations: 108
## Number of Groups: 27
```

```
fm1Orth.lme <- lme( distance ~ I(age-11), data = Orthodont )
summary(fm1Orth.lme)
```

```
## Linear mixed-effects model fit by REML
##   Data: Orthodont
##        AIC      BIC    logLik
##   454.6367 470.6173 -221.3183
##
## Random effects:
##  Formula: ~I(age - 11) | Subject
##  Structure: General positive-definite
##             StdDev    Corr
## (Intercept) 2.1343328 (Intr)
## I(age - 11) 0.2264275 0.503
## Residual    1.3100394
##
## Fixed effects:  distance ~ I(age - 11)
##                 Value Std.Error DF  t-value p-value
## (Intercept) 24.023148 0.4296608 80 55.91189       0
## I(age - 11)  0.660185 0.0712532 80  9.26534       0
##  Correlation:
##             (Intr)
## I(age - 11) 0.294
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -3.223106430 -0.493761203  0.007316812  0.472151146  3.916034259
##
## Number of Observations: 108
## Number of Groups: 27
```

```
fm1Orth.lme <- lme( fm2Orth.lis )
summary(fm1Orth.lme)
```

```
## Linear mixed-effects model fit by REML
##   Data: Orthodont
##        AIC      BIC    logLik
##   454.6367 470.6173 -221.3183
##
## Random effects:
##  Formula: ~I(age - 11) | Subject
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 2.1343294 (Intr)
## I(age - 11) 0.2264278 0.503
## Residual    1.3100400
##
## Fixed effects:  distance ~ I(age - 11)
```

```
##                  Value Std.Error DF  t-value p-value
## (Intercept) 24.023148 0.4296602 80 55.91197       0
## I(age - 11)  0.660185 0.0712533 80  9.26533       0
##  Correlation:
##             (Intr)
## I(age - 11) 0.294
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.22310613 -0.49376096  0.00731653  0.47215112  3.91603230
##
## Number of Observations: 108
## Number of Groups: 27
```

Uma vantagem de utilizar a ultima forma de implementar o modelo é que as estimativas iniciais para os parâmetros na verossimilhança perfilada (restrita), do modelo de efeitos mistos, são calculadas automaticamente a partir do objeto `lmList`.

Uma das questões de interesse relativa aos dados de crescimento ortodôntico é se meninos e meninas têm padrões de crescimento diferentes.[7] Esta questão pode ser avaliada ajustando o modelo

[7] Observe que o `lmList` não pode ser usado para testar diferenças de gênero nos dados de crescimento ortodôntico, pois estima coeficientes individuais para cada sujeito.

```
fm2Orth.lme <- update(fm1Orth.lme,fixed = distance~Sex*I(age-11)-1)
summary(fm2Orth.lme)
```

```
## Linear mixed-effects model fit by REML
##   Data: Orthodont
##        AIC      BIC    logLik
##   448.5817 469.7368 -216.2908
##
## Random effects:
##  Formula: ~I(age - 11) | Subject
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 1.8303271 (Intr)
## I(age - 11) 0.1803454 0.206
## Residual    1.3100396
##
## Fixed effects:  distance ~ Sex + I(age - 11) + Sex:I(age - 11) - 1
##                           Value Std.Error DF  t-value p-value
## SexMale               24.968750 0.4860008 25 51.37595  0.0000
## SexFemale             22.647727 0.5861390 25 38.63883  0.0000
## I(age - 11)            0.784375 0.0859995 80  9.12069  0.0000
## SexFemale:I(age - 11) -0.304830 0.1347353 80 -2.26243  0.0264
##  Correlation:
```

```
##                       SexMal SexFml I(-11)
## SexFemale              0.000
## I(age - 11)            0.102  0.000
## SexFemale:I(age - 11) -0.065  0.078 -0.638
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -3.168078588 -0.385939131  0.007103648  0.445154526  3.849463491
##
## Number of Observations: 108
## Number of Groups: 27
```

Os pequenos valores de p associados a Sex e $Sex : I(age - 11)$ no resultado do `summary` indicam que meninos e meninas têm padrões de crescimento ortodôntico significativamente diferentes.

A função `fitted` é usado para extrair os valores ajustados do objeto `lme`. Por padrão, os valores ajustados dentro do grupo, ou seja, os valores ajustados correspondentes às estimativas dos coeficientes individuais, são produzidos. Os valores preditos à população, baseados apenas nas estimativas de efeitos fixos, são obtidos definindo o argumento `level=0`. Ambos os tipos de valores ajustados podem ser obtidos simultaneamente com:

```
fm2Orth.lme$coefficients$fixed
```

```
##               SexMale             SexFemale           I(age - 11)
##            24.9687500            22.6477273             0.7843750
## SexFemale:I(age - 11)
##            -0.3048295
```

```
fm2Orth.lme$coefficients$random$Subject[c(1:5,17), ]
```

```
##     (Intercept) I(age - 11)
## M16   -1.758374 -0.08864715
## M05   -1.738555 -0.00847416
## M02   -1.411531 -0.02356274
## M11   -1.219939 -0.14051911
## M07   -1.077900 -0.01192699
## F10   -3.673842 -0.06269744
```

Vamos considerar duas crianças M16 e F10, avaliados aos 8 anos

```
coefficients(fm2Orth.lme)[c(1,17),]
```

```
##      SexMale SexFemale I(age - 11) SexFemale:I(age - 11) (Intercept)
## M16 24.96875  22.64773   0.6957278            -0.3048295   -1.758374
## F10 24.96875  22.64773   0.7216776            -0.3048295   -3.673842
```

$$
\begin{aligned}
\widehat{y_{t,M16}} =&(24.9687500 - 1.758374) - 2.3210227 SexFemale + (0.7843750 - 0.088647147) I(age - 11) \\
&- 0.3048295 SexFemale : I(age - 11) \\
=&23.21038 + 0.6957279 I(age - 11)
\end{aligned}
$$

para 8 anos

$$
\begin{aligned}
y_{\widehat{t=-3,M16}} =&23.21038 + 0.6957279(-3) \\
y_{\widehat{t=-3,M16}} =&21.1232
\end{aligned}
$$

e

$$
\begin{aligned}
\widehat{y_{t,F10}} =&(24.9687500 - 3.673842) - 2.3210227 SexFemale + (0.7843750 - 0.062697440) I(age - 11) \\
&- 0.3048295 SexFemale : I(age - 11) \\
=&21.29491 - 2.3210227 SexFemale + 0.7216776 I(age - 11) - 0.3048295 SexFemale : I(age - 11) \\
=&18.97389 + 0.4168481 I(age - 11)
\end{aligned}
$$

para 8 anos

$$
\begin{aligned}
y_{\widehat{t=-3,F10}} =&18.97389 + 0.4168481(-3) \\
y_{\widehat{t=-3,F10}} =&17.72335
\end{aligned}
$$

Conferindo as estimativas obtidas com a saída do R:

```
predict(fm2Orth.lme)[c(61, 101)]
```

```
##      M16      F10
## 21.12319 17.72334
```

```
fitted( fm2Orth.lme, level = 0 )[c(61, 101)]  # sem efeito aleatório
```

```
##      M16      F10
## 22.61562 21.20909
```

```
fitted( fm2Orth.lme, level = 1 )[c(61, 101)]  # com efeito aleatório
```

```
##      M16      F10
## 21.12319 17.72334
```

Obtendo somente os efeitos aleatórios para as crianças M16 e F10:

```
ef<- ranef(fm2Orth.lme)
ef[c(1, 17), ]
```

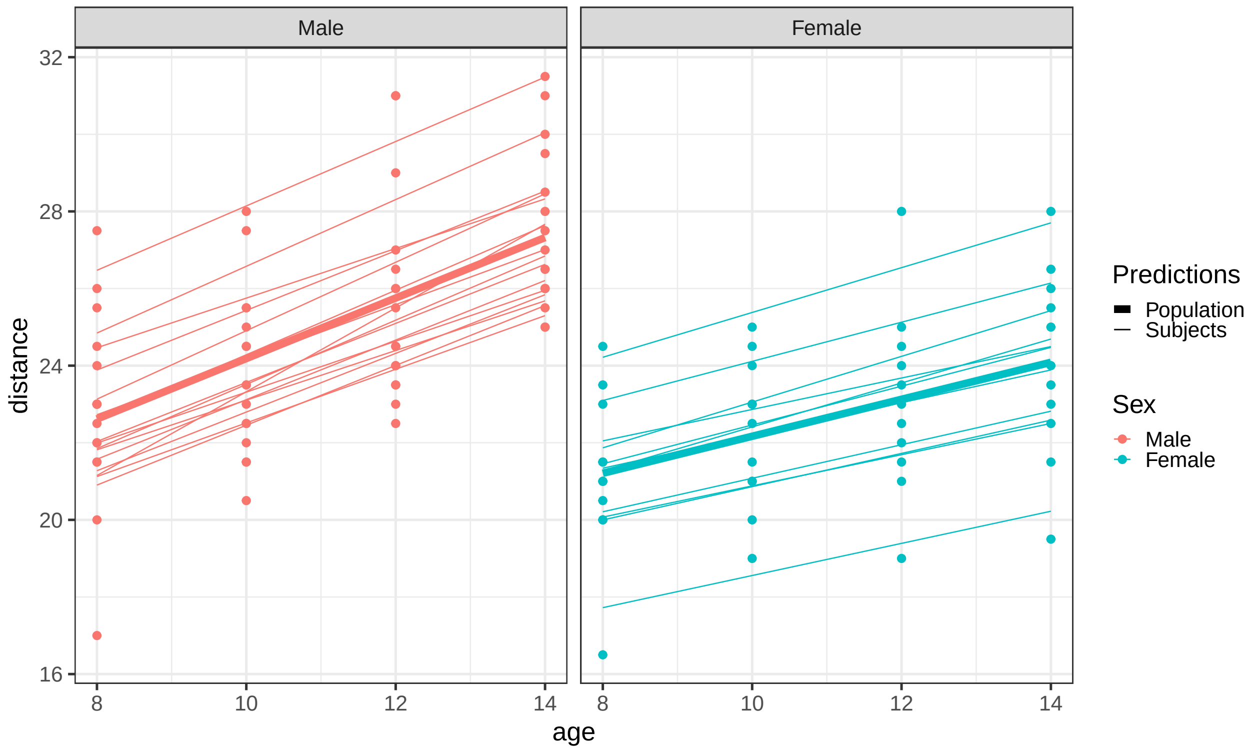


Figure 4: Crescimento dentário (mm) ajustado por fm2Orth.lme, efeito aleatório para intercepto e coeficiente angular

```
##     (Intercept) I(age - 11)
## M16   -1.758374 -0.08864715
## F10   -3.673842 -0.06269744
```

Os resíduos são extraídos com a função `resid`, que também usa um argumento de (level).[8]

```
resid( fm2Orth.lme, level = 0 )[c(61, 101)]  # sem efeito aleatório
```

```
##       M16       F10
## -0.615625 -4.709091
```

```
resid( fm2Orth.lme, level = 1 )[c(61, 101)]  # com efeito aleatório
```

```
##        M16        F10
##  0.8768074 -1.2233409
```

[8] Os resíduos padronizados ou de Pearson, correspondentes aos resíduos brutos divididos pelo desvio padrão estimado dentro do grupo, são obtidos usando o argumento `type = "pearson"`

A função `compareFits` pode ser usada para uma comparação dos parâmetros fixos para cada indivíduo. A saída fornecida por essa função é um (array), veja o caso das crianças M16 e F10:

```
compOrth <- compareFits( coef(fm2Orth.lis), coef(fm1Orth.lme) )
compOrth["M16",1:2,1:2]
```

```
##                    (Intercept) I(age - 11)
## coef(fm2Orth.lis)      23.000   0.5500000
## coef(fm1Orth.lme)      23.078   0.5913314
```

```
compOrth["F10",1:2,1:2]
```

```
##                    (Intercept) I(age - 11)
## coef(fm2Orth.lis)    18.50000   0.4500000
## coef(fm1Orth.lme)    18.98527   0.4095941
```

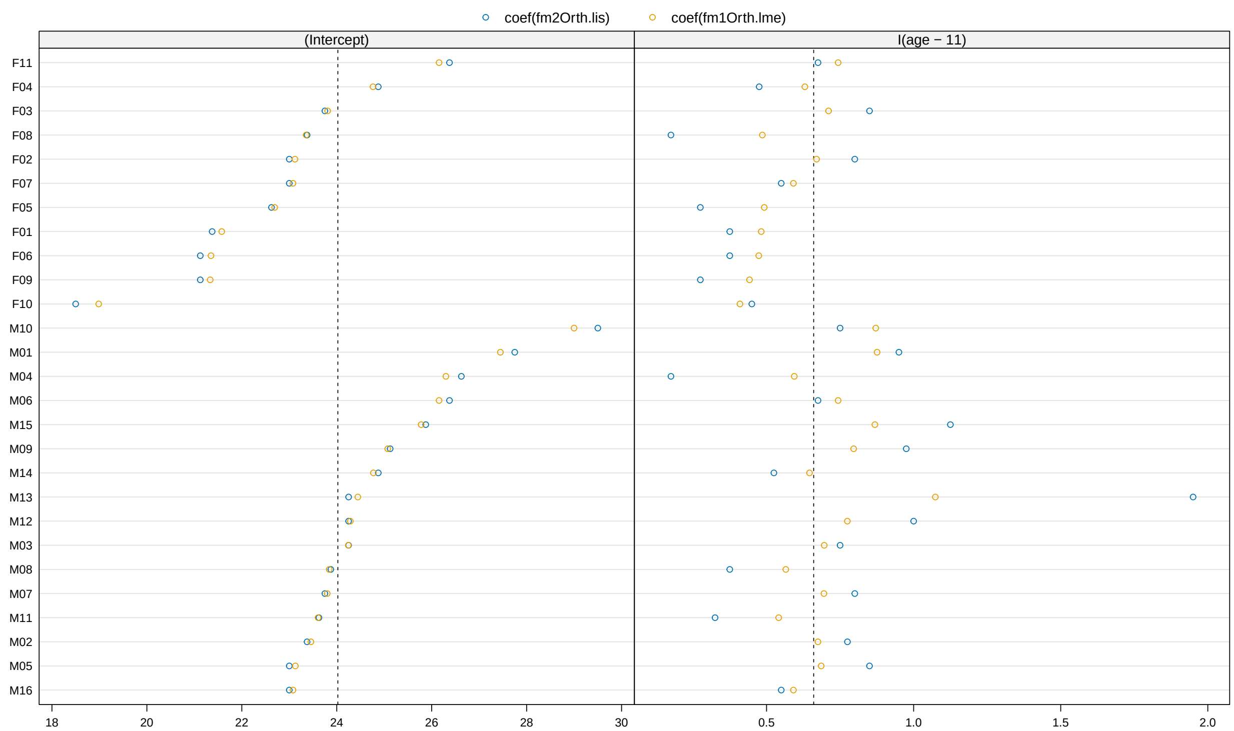


Figure 5: Comparativo entre as estimativas dos parâmetros de efeitos fixos do modelo sem e com efeitos aleatórios

Em seu material, na página 26, Arnab Maity de ajustou o modelo a que ele chamou de {**Model A}, aqui iremos ajustar o mesmo modelo porém com parametrização diferente.**

```
fit.a <- lme(fixed = distance ~ -1 + G + G:age + M + M:age,
             random = ~ age | id, data = dental, method = "ML")

fit.a2 <- lme(fixed = distance ~ -1 + Sex + I(age) + Sex:I(age),
              random = ~ age | id, data = dental, method = "ML")

fit.a3 <- lme(fixed = distance ~ -1 + Sex + I(age - 11) + Sex:I(age - 11),
              random = ~ age | id, data = dental, method = "ML")
```

```
fit.a4 <- lme(fixed = distance ~ Sex + I(age) + Sex:I(age),
             random = ~ age | id, data = dental, method = "ML")

fit.a5 <- lme(fixed = distance ~ Sex + I(age - 11) + Sex:I(age - 11),
             random = ~ age | id, data = dental, method = "ML")
```

Primeiramente podemos verificar que o llogLik e o AIC, são iguais para as três formas de implementar:

```
(LogLik<- c(logLik(fit.a ),logLik(fit.a2),logLik(fit.a3),logLik(fit.a4),logLik(fit.a5)))

## [1] -213.903 -213.903 -213.903 -213.903 -213.903

(AIC<- c(AIC(fit.a ),AIC(fit.a2),AIC(fit.a3),AIC(fit.a4),AIC(fit.a5)))

## [1] 443.806 443.806 443.806 443.806 443.806
```

Os summary dos seis ajustes apresentam os mesmos resultados sobre Random effects:, como era esperado. Com os resultados obtidos podemos escrever as matrizes G e R, dadas em (2). [9]

```
Random effects:
 Formula: ~age | id
 Structure: General positive-definite, Log-Cholesky
            StdDev   Corr
(Intercept) 2.134688 (Intr)
age         0.154139 -0.603
Residual    1.310040
```

Vejamos agora como são apresentrados os efeitos fixos do modelo:

```
# fit.a
fixed.effects(fit.a)

##          G          M      G:age      age:M
## 17.3727273 16.3406250  0.4795455  0.7843750
```

$$\begin{aligned} \textbf{Girls } (G=1) &: \quad \widehat{\mu}(age) \approx 17.373 + 0.48 age \\ \textbf{Boys } (G=0) &: \quad \widehat{\mu}(age) \approx 16.341 + 0.784 age. \end{aligned}$$

```
# fit.a2
fixed.effects(fit.a2)

##         SexMale        SexFemale           I(age) SexFemale:I(age)
##      16.3406250       17.3727273        0.7843750       -0.3048295
```

[9]
$$cov(\boldsymbol{b}_i) = \boldsymbol{D} = \begin{pmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{pmatrix}$$
$$= \begin{pmatrix} 2.13^2 & -0.603x2.13^2x0.15^2 \\ -0.603x2.13^2x0.15^2 & 0.15^2 \end{pmatrix}$$
$$= \begin{pmatrix} 4.56 & -0.60 \\ -0.60 & 0.024 \end{pmatrix},$$
a matriz $\boldsymbol{G}$ será bloco diagonal. Temos $\sigma = 1.31$, então
$$\boldsymbol{R}_i = \begin{pmatrix} 1.72 & & & \\ & 1.72 & & \\ & & 1.72 & \\ & & & 1.72 \end{pmatrix}$$
.

$$\begin{aligned} \textbf{Girls } (G = 1) &: \quad \widehat{\mu}(age) \approx 17.373 + 0.479 age \\ \textbf{Boys } (G = 0) &: \quad \widehat{\mu}(age) \approx 16.341 + 0.784 age. \end{aligned}$$

```
# fit.a3
fixed.effects(fit.a3)
```

```
##              SexMale            SexFemale          I(age - 11)
##           24.9687500           22.6477273            0.7843750
## SexFemale:I(age - 11)
##           -0.3048295
```

$$\begin{aligned} \textbf{Girls } (SexFemale = 1) &: \quad \widehat{\mu}(age) \approx 17.379 + 0.479 age \\ \textbf{Boys } (SexFemale = 0) &: \quad \widehat{\mu}(age) \approx 16.345 + 0.784 age. \end{aligned}$$

```
# fit.a4
fixed.effects(fit.a4)
```

```
##     (Intercept)       SexFemale          I(age) SexFemale:I(age)
##      16.3406250       1.0321023       0.7843750       -0.3048295
```

$$\begin{aligned} \textbf{Girls } (SexFemale = 1) &: \quad \widehat{\mu}(age) \approx 17.373 + 0.479 age \\ \textbf{Boys } (SexFemale = 0) &: \quad \widehat{\mu}(age) \approx 16.341 + 0.784 age. \end{aligned}$$

```
# fit.a5
fixed.effects(fit.a5)
```

```
##          (Intercept)            SexFemale          I(age - 11)
##           24.9687500           -2.3210227            0.7843750
## SexFemale:I(age - 11)
##           -0.3048295
```

$$\begin{aligned} \textbf{Girls } (SexFemale = 1) &: \quad \widehat{\mu}(age) \approx 17.379 + 0.479 age \\ \textbf{Boys } (SexFemale = 0) &: \quad \widehat{\mu}(age) \approx 16.345 + 0.784 age. \end{aligned}$$

Como pode ser observado todos os cinco ajustes resultam no mesmo modelo, o que hé de diferente é a forma de interpretar as saídas.

```
# fit.a2
Fixed effects: distance ~ -1 + Sex + I(age) + Sex:I(age)
```

```
                    Value Std.Error DF   t-value p-value
SexMale         16.340625 0.9987521 25 16.361042  0.0000
SexFemale       17.372727 1.2045404 25 14.422702  0.0000
I(age)           0.784375 0.0843294 80  9.301321  0.0000
SexFemale:I(age) -0.304830 0.1321188 80 -2.307238  0.0236


# fit.a4
Fixed effects: distance ~ Sex + I(age) + Sex:I(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
I(age)           0.784375 0.0843294 79  9.301322  0.0000
SexFemale:I(age) -0.304830 0.1321188 79 -2.307238  0.0237
```

Note que com o ajuste `fit.a2` podemos testar cada um dos parâmetros de efeito fixo e com `fit.a4` testamos se há diferença para o crescimento médio entre os gêneros. Observe ainda que 17.372727-16.340625=1.032102, que é exatamente a estimativa apresentada para `SexFemale` no ajuste `fit.a4`.

Observe a seguir a estrutura de correlação entre os parâmetros em dois modelos que se diferem por ter a variável independe *age* não centralizada e centralizada. A correlação entre os parâmetros de intercepto e inclinação sendo -0.880 e 0.562 para *age* não centralizada e 0.102 e -0.065 para *age* centralizada.

```
# fit.a4
Fixed effects: distance ~ Sex + I(age) + Sex:I(age)
 Correlation:
                 (Intr) SexFml I(age)
SexFemale        -0.638
I(age)           -0.880  0.562
SexFemale:I(age)  0.562 -0.880 -0.638

# fit.a5
Fixed effects: distance ~ Sex + I(age - 11) + Sex:I(age - 11)
 Correlation:
                      (Intr) SexFml I(-11)
SexFemale             -0.638
I(age - 11)            0.102 -0.065
SexFemale:I(age - 11) -0.065  0.102 -0.638
```

Dado que os modelos são apenas implementados de forma diferente, de acordo com o que deseja-se testar, os resíduos e novas suposições podem seguir do material de Arnab Maity

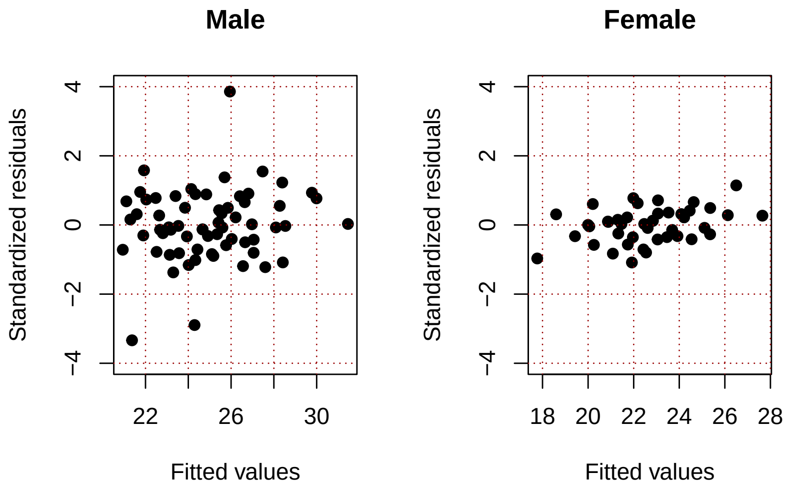


Figure 6: Plot of Pearson residuals vs. subject-level fitted values for both the groups (G = 0: male, and G = 1: female) for 'fit.a'.

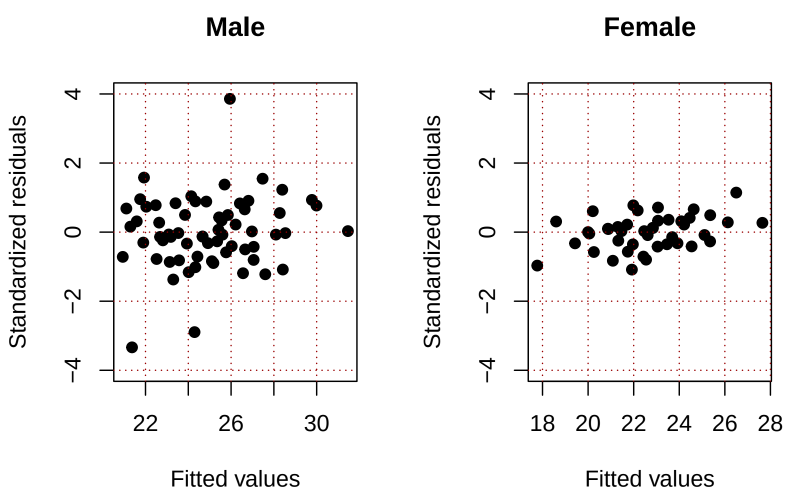


Figure 7: Plot of Pearson residuals vs. subject-level fitted values for both the groups (G = 0: male, and G = 1: female) for 'fit.a2'.

Nas Figuras ?? e ?? são apresentados gráficos de resíduos para os ajustes `fit.a` e `fit.a2`, corroborando com o que já foi discutido.