Alunos: Álvaro Kothe, Giovanni Piccirilli, Larissa Eleutério, Marcos Andrade
Docente: Viviana Giampaoli
Data: 23/09/2021

Modelos Mistos - Lista 2

```r
## importando pacote e dados necessários

library(nlme)
data("Orthodont")
data = Orthodont
```

Exercício 1

Cosidere a base Orthodont, ajuste o modelos misto da distância em função da idade e do sexo, com o efeito aleatório sujeito, considerando o método ML e REML. Observe se existe diferenças.

Resposta:

Primeiramente ajustou-se o modelo pelo método da máxima verossimilhança.

```r
## ajustando um modelo de efeitos fixos pelo método da máxima verossimilhança
lmeML1=lme(fixed= distance~age + Sex, random=~ 1 |Subject, method="ML",data=data)

## resultados
lmeML1
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: data
##   Log-likelihood: -217.4282
##   Fixed: distance ~ age + Sex
## (Intercept)         age   SexFemale
##  17.7067130   0.6601852  -2.3210227
##
## Random effects:
##  Formula: ~1 | Subject
##         (Intercept) Residual
## StdDev:    1.730079 1.422728
##
## Number of Observations: 108
## Number of Groups: 27
```

As estimativas dos parâmetros são:

- intercepto: 17.71
- idade: 0.66
- sexo: -2.32
- desvio padrão dos resíduos: 1.42
- desvio padrão do efeito aleatório: 1.73

Ajustando agora o modelo com o método da máxima verossimilhança restrita é possível ver que as estimativas dos coeficientes de efeito fixo se mantiveram. As estimativas dos desvios tiveram uma pequena alteração.

```r
## ajustando um modelo de efeitos fixos pelo método da máxima verossimilhança restrita
lmeREML1=lme(fixed= distance~age + Sex, random=~ 1 |Subject, method="REML",
             data=data)

## resultados
lmeREML1
```

```
## Linear mixed-effects model fit by REML
##   Data: data
##   Log-restricted-likelihood: -218.7563
##   Fixed: distance ~ age + Sex
## (Intercept)         age   SexFemale
##  17.7067130   0.6601852  -2.3210227
##
## Random effects:
##  Formula: ~1 | Subject
##         (Intercept) Residual
## StdDev:    1.807425 1.431592
##
## Number of Observations: 108
## Number of Groups: 27
```

- desvio padrão dos resíduos: 1.43
- desvio padrão do efeito aleatório: 1.81

```r
comparando_modelos <- rbind(lmeML1$coefficients$fixed,
                            lmeREML1$coefficients$fixed) %>%
  cbind(rbind(as.numeric(VarCorr(lmeML1)[,2]),
              as.numeric(VarCorr(lmeREML1)[,2]))) %>%
  as.data.frame()

names(comparando_modelos) <- c("Intercepto", "Idade", "Sexo",
                               "$\\hat{\\sigma_d}$", "$\\hat{\\sigma}$")

row.names(comparando_modelos) <- c("lmeML1", "lmeREML1")

kable(comparando_modelos, escape = FALSE)
```

	Intercepto	Idade	Sexo	$\hat{\sigma_d}$	$\hat{\sigma}$
lmeML1	17.70671	0.6601852	-2.321023	1.730079	1.422728
lmeREML1	17.70671	0.6601852	-2.321023	1.807425	1.431592

Exercício 2

Refaça o item anterior considerando outra parametrização.

Resposta:

A parametrização padrão para variáveis categóricas no pacote nlme é a casela de referência, onde crianças do sexo masculino foram marcadas com '0' e crianças do sexo femenino como '1'.

```r
lmeML1$contrasts
```

```
## $Sex
##        Female
## Male        0
## Female      1
```

Agora a parametricação será contraste da soma, onde o sexo masculino aparece como '1' e o sexo feminino como '-1'.

```r
lmeML2=lme(fixed= distance~age + Sex, random=~ 1 |Subject, method="ML",data=data,
           contrasts = list(Sex = "contr.sum"))

lmeML2$contrasts
```

```
## $Sex
##        [,1]
## Male      1
## Female   -1
```

Ajustando o modelo com estimativa de máxima verossimilhança para os parâmetros, apenas as estimativas do intercepto e do coeficiente associado ao fator fixo de idade se alteram. As demais estimativas se mantêm iguais ao modelo ajustado pela máxima verissimilhança com a parametrização casela de referência.

```r
lmeML2
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: data
##   Log-likelihood: -217.4282
##   Fixed: distance ~ age + Sex
## (Intercept)         age        Sex1
##  16.5462016   0.6601852   1.1605114
##
## Random effects:
##  Formula: ~1 | Subject
##         (Intercept) Residual
## StdDev:    1.730079 1.422728
##
## Number of Observations: 108
## Number of Groups: 27
```

Similar a situação anterior, quando se comparam as estimativas do modelo estimado com máxima verossimilhança restrita, e diferença na parametrização de sexo, apenas intercepto e estimativa de coeficiente associado ao sexo são diferentes. Além disso, as variâncias de ambos os modelos permaneceram iguais.

```r
lmeREML2=lme(fixed= distance~age + Sex, random=~ 1 |Subject, method="REML",
             data=data, contrasts = list(Sex = "contr.sum"))

lmeREML2
```

```
## Linear mixed-effects model fit by REML
##   Data: data
##   Log-restricted-likelihood: -219.4494
##   Fixed: distance ~ age + Sex
## (Intercept)         age        Sex1
##  16.5462016   0.6601852   1.1605114
##
## Random effects:
##  Formula: ~1 | Subject
##         (Intercept) Residual
## StdDev:    1.807425 1.431592
##
## Number of Observations: 108
## Number of Groups: 27
```

É importante destacar que, independente da parametrização, a relação entre os fatores da variável sexo se mantém.

Quando se usa casela de referência, o fator referência é o masculino e o coeficiente traz informação que, o fato de ser menina diminui a estimativa de tamanho da glândula em -2.32.

Na utilização do contraste da soma, multiplicamos o coeficiente pela representação do sexo. No caso masculino a representação é '1' e no caso feminino a representação é '-1'. O valor do coeficiente é 1.16 (positivo). Isto significa que a estimativa de tamanho da glândula para meninos é 1*1.16 e para meninas é -1*1.16.

Portanto, em ambas as parametrizações, se interpreta que o sexo masculino influencia positivamente no tamanho da glândula.

```r
comparando_modelos2 <- rbind(lmeML1$coefficients$fixed,
                             lmeML2$coefficients$fixed,
                             lmeREML1$coefficients$fixed,
                             lmeREML2$coefficients$fixed) %>%
  cbind(rbind(as.numeric(VarCorr(lmeML1)[,2]), as.numeric(VarCorr(lmeML2)[,2]),
              as.numeric(VarCorr(lmeREML1)[,2]),
              as.numeric(VarCorr(lmeREML2)[,2]))) %>%
  as.data.frame()

names(comparando_modelos2) <- c("Intercepto", "Idade", "Sexo",
                                "$\\hat{\\sigma_d}$", "$\\hat{\\sigma}$")

row.names(comparando_modelos2) <- c("lmeML1", "lmeML2", "lmeREML1", "lmeREML2")


kable(comparando_modelos2, escape = FALSE)
```

	Intercepto	Idade	Sexo	$\hat{\sigma_d}$	$\hat{\sigma}$
lmeML1	17.70671	0.6601852	-2.321023	1.730079	1.422728
lmeML2	16.54620	0.6601852	1.160511	1.730079	1.422728
lmeREML1	17.70671	0.6601852	-2.321023	1.807425	1.431592
lmeREML2	16.54620	0.6601852	1.160511	1.807425	1.431592

Exercício 3

Considerando valores próximos as estimativas de ML para os parâmetros fixos e de REML para as componentes de variância, simule duas novas bases de dados, uma em que o número de meninos é 32 e o de meninas é 22, e outra em que o número de meninos é 96 e o de meninas 33.

Resposta:

Para gerar a variável resposta simulada definimos os seguintes parâmetros:

- Intercepto: 17
- Sexo: -2.5
- Idade: 0.6
- desvio padrão dos resíduos: 1.45
- desvio padrão do efeito aleatório: 1,85

Simulação para 54 indivívuos Vemos que as variáveis y (tamanho da glândula), intercepto e idade são numéricas, entquanto que sexo é uma variável categórica com dois fatores e sujeito outra variável categórica com 54 fatores. O número total de linhas é 216 que representam os 54 indivíduos em cada uma das 4 idades observadas.

```
age = 0.6
sex = -2.5
intercepto = 17
sigmab = 1.85^2 #o summary solta o desvio padrão
sigma = 1.45^2

set.seed(2109)

#Para 32 meninos e 22 meninas
#head(data)
X = model.matrix(~age + Sex, data)
X2 = rbind(X,X) #Dobrei a matrix X
beta1 = c(intercepto, age, sex)
muv = X2%*%beta1
Zi = matrix(1,4,1) #matriz dos efeitos aletorios de cada individuo i (são iguais
#para todo i)
Vi = sigma*(diag(4)) + sigmab*(Zi%*%(t(Zi))) # Matriz de variância para indiv i
V = kronecker(diag(27*2), Vi) #matriz de variancia geral, essa multiplicação

#cria uma matriz quadrada em que cada elemento da diagonal é Vi
library(MASS)
y = mvrnorm(1, muv, V) #simulando da normal multivariada
individuo = rep(seq(1:54),each = 4) #gerando o indice dos individuos
datasim2 = data.frame(cbind(y, X2, individuo))
names(datasim2) = c("y","Intercept", "age", "Sex","Subject")

datasim2$Sex <- as.factor(datasim2$Sex)
datasim2$Subject <- as.factor(datasim2$Subject)

str(datasim2)
```

```
## 'data.frame':    216 obs. of  5 variables:
##  $ y        : num  21.6 25.2 25.5 25.9 24.6 ...
##  $ Intercept: num  1 1 1 1 1 1 1 1 1 1 ...
##  $ age      : num  8 10 12 14 8 10 12 14 8 10 ...
##  $ Sex      : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Subject  : Factor w/ 54 levels "1","2","3","4",..: 1 1 1 1 2 2 2 2 3 3 ...
```

```r
# table(datasim2$Sex)/4 # 4 coletas por indivíduo
```

Simulação para 129 indivívuos Todas as catacterísticas observadas anteriormente se mantêm nesta simulação. O número total de linhas é 512 que representam os 129 indivíduos em cada uma das 4 idades observadas.

```r
#Para 96 meninos e 32 meninas

X = model.matrix(~age + Sex, data)
Xmale = X[X[,3] == 0,]
Xmale = matrix(rep(t(Xmale), 6), ncol = ncol(Xmale), byrow = TRUE)
Xfemale = X[X[,3] == 1,]
Xfemale = matrix(rep(t(Xfemale), 3), ncol = ncol(Xfemale), byrow = TRUE)

X3 = rbind(Xmale, Xfemale)
beta1 = c(intercepto, age, sex)
muv = X3%*%beta1
Zi = matrix(1,4,1)
Z = kronecker(diag(129), Zi)
Vi = sigma*(diag(4)) + sigmab*(Zi%*%(t(Zi)))
V = kronecker(diag(129), Vi)
y = mvrnorm(1, muv, V)
individuo = rep(seq(1:129),each = 4)
datasim3 = data.frame(cbind(y, X3, individuo))
names(datasim3) = c("y","Intercept", "age", "Sex","Subject")

datasim3$Sex <- as.factor(datasim3$Sex)
datasim3$Subject <- as.factor(datasim3$Subject)

str(datasim3)
```

```
## 'data.frame':    516 obs. of  5 variables:
##  $ y        : num  22.6 21.5 21.2 24.5 19.2 ...
##  $ Intercept: num  1 1 1 1 1 1 1 1 1 1 ...
##  $ age      : num  8 10 12 14 8 10 12 14 8 10 ...
##  $ Sex      : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Subject  : Factor w/ 129 levels "1","2","3","4",..: 1 1 1 1 2 2 2 2 3 3 ...
```

Exercício 4

Realize o solicitado nos itens anteriores para a nova base.

Resposta:

Base simulada 1: 32 meninos e 22 meninas. Ajuste do modelo pelo método da máxima verossimilhança com dados de 54 crianças.

```r
lmeML_sim1=lme(fixed= y~age + Sex, random=~ 1 |Subject, method="ML",data=datasim2)
lmeML_sim1
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: datasim2
##   Log-likelihood: -428.5442
##   Fixed: y ~ age + Sex
## (Intercept)         age        Sex1
##   17.052539    0.605882   -3.426249
##
## Random effects:
##  Formula: ~1 | Subject
##         (Intercept) Residual
## StdDev:    1.924368 1.330241
##
## Number of Observations: 216
## Number of Groups: 54
```

Ajuste do modelo pelo método da máxima verossimilhança restrita com dados de 54 crianças.

```
lmeREML_sim1=lme(fixed= y~age + Sex, random=~ 1 |Subject, method="REML",data=datasim2)
lmeREML_sim1
```

```
## Linear mixed-effects model fit by REML
##   Data: datasim2
##   Log-restricted-likelihood: -430.8304
##   Fixed: y ~ age + Sex
## (Intercept)         age        Sex1
##   17.052539    0.605882   -3.426249
##
## Random effects:
##  Formula: ~1 | Subject
##         (Intercept) Residual
## StdDev:     1.96466 1.334366
##
## Number of Observations: 216
## Number of Groups: 54
```

Base simulada 2: 96 meninos e 33 meninas.

Ajuste do modelo pelo método da máxima verossimilhança com dados de 129 crianças.

```
lmeML_sim2=lme(fixed= y~age + Sex, random=~ 1 |Subject, method="ML",data=datasim3)
lmeML_sim2
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: datasim3
##   Log-likelihood: -1059.951
##   Fixed: y ~ age + Sex
## (Intercept)         age        Sex1
##  17.0602429   0.5663241  -2.1310935
##
## Random effects:
##  Formula: ~1 | Subject
##         (Intercept) Residual
```

```
## StdDev:    1.833312 1.475319
##
## Number of Observations: 516
## Number of Groups: 129
```

Ajuste do modelo pelo método da máxima verossimilhança restrita com dados de 129 crianças.

```
lmeREML_sim2=lme(fixed= y~age + Sex, random=~ 1 |Subject, method="REML",data=datasim3)
lmeREML_sim2
```

```
## Linear mixed-effects model fit by REML
##   Data: datasim3
##   Log-restricted-likelihood: -1063.393
##   Fixed: y ~ age + Sex
## (Intercept)         age        Sex1
##  17.0602429   0.5663241  -2.1310935
##
## Random effects:
##  Formula: ~1 | Subject
##         (Intercept) Residual
## StdDev:    1.849627 1.477229
##
## Number of Observations: 516
## Number of Groups: 129
```

Exercício 5

Compare os resultados.

Resposta:

```
# Extrai informação do modelo
extrai_estimativas = function(modelo) {
    betas = modelo$coefficients$fixed
    desvios = as.numeric(VarCorr(modelo)[,2])
    estimativas = c(betas, desvios)
    names(estimativas) = c("Intercepto", "Idade", "Sexo", "$\\hat{\\sigma_d}$",
                           "$\\hat{\\sigma}$")

    return(estimativas)
}
```

```
library(purrr)

lst_modelos = list(
    "ML" = lmeML1,
    "ML reparametrizado" = lmeML2,
    "ML Simulado 1" = lmeML_sim1,
    "ML Simulado 2" = lmeML_sim2,

    "REML" = lmeREML1,
    "REML reparametrizado" = lmeREML2,
```

```
    "REML Simulado 1" = lmeREML_sim1,
    "REML Simulado 2" = lmeREML_sim2
)

tabela_estimativas = map_dfr(lst_modelos, extrai_estimativas, .id = "Modelo")

kable(tabela_estimativas, escape = FALSE, booktabs = TRUE, linesep = "")
```

Modelo	Intercepto	Idade	Sexo	$\hat{\sigma_d}$	$\hat{\sigma}$
ML	17.70671	0.6601852	-2.321023	1.730079	1.422728
ML reparametrizado	16.54620	0.6601852	1.160511	1.730079	1.422728
ML Simulado 1	17.05254	0.6058820	-3.426249	1.924368	1.330241
ML Simulado 2	17.06024	0.5663241	-2.131093	1.833312	1.475319
REML	17.70671	0.6601852	-2.321023	1.807425	1.431592
REML reparametrizado	16.54620	0.6601852	1.160511	1.807425	1.431592
REML Simulado 1	17.05254	0.6058820	-3.426249	1.964660	1.334366
REML Simulado 2	17.06024	0.5663241	-2.131093	1.849627	1.477229

Pela tabela acima percebe-se que as estimativas dos efeitos fixos permanecem iguais quando ajustado no mesmo banco de dados, independente se foi utilizado o método de Maxíma verossimilhança ou máxima verossimilhança restrita.

A principal diferença entre os métodos se encontra na estimativas dos desvios padrão, em que esta diferença é visível, porém pequena.

Utilizando um banco de dados maior, que são as bases simuladas, vemos que a diferença entre os desvios é menor, o que leva a considerar que não existe diferença entre os métodos para uma amostra suficientemente grande.