

MODELOS MATEMÁTICOS E APLICAÇÕES 2023/2024

An explanation of the exercises

LINEAR MIXED MODELS

EXERCISE 1#

```
touriga<-read.table("touriga.txt", header=T)
touriga
```

a) Random model (one factor of random effects, genotype), balanced, with **G** and **R** diagonal matrices

$$Y_{ij} = \mu + u_i + e_{ij}$$

for $i = 1, \dots, 196, j = 1, \dots, 5, n = 196 \times 5$.

Y_{ij} is the j th observation of yield in the i th genotype;

μ is a general mean (population);

u_i is the effect of the genotype i (random effect);

e_{ij} is the random error associated to the observation Y_{ij} .

$u_i, i. i. d., \mathcal{N}(0, \sigma_u^2), \forall i; e_{ij}, i. i. d., \mathcal{N}(0, \sigma_e^2), \forall ij;$

u_i and e_{ij} are mutually independent, $\forall ij$.

b) #ajustar o modelo usando o método REML# MODEL FITTED WITH REML METHOD

```
#alinea b)#
```

```
#lme do package nlme#
```

```
tourigalme1<-lme(rend~1, random=~1|clone, data=touriga)
```

```
summary(tourigalme1)
```

```
#para obter as estimativas das componentes de variância#
```

```
VarCorr(tourigalme1)
```

```
#lmer do package lme4#
```

```
tourigalmer1<-lmer(rend~1+(1|clone), data=touriga)
```

```
summary(tourigalmer1)
```

i) REML estimates for the variance components.

$\hat{\sigma}_u^2=0.18982; \hat{\sigma}_e^2=0.2245$

ii) For a random model with one factor of random effects, balanced (factor with a levels, b observations per level), the empirical best linear unbiased predictor of u_i (for the level i) is:

$$EBLUP(u_i) = \frac{b\hat{\sigma}_u^2}{b\hat{\sigma}_u^2 + \hat{\sigma}_e^2} (\bar{Y}_i - \bar{Y}_{..})$$

Knowing that $\bar{Y}_{..} = 1.196$ kg/plant and $\bar{Y}_{c0101} = 1.6044$ kg/plant, the value of the empirical best linear unbiased predictor of the yield genotypic effect of the genotype c0101 is:

$$EBLUP(c0101) = \frac{0.18982}{0.18982 + 0.2245/5} (1.6044 - 1.196) = 0.3303 \text{ kg/plant}$$

Due to the effects of the factor (genotype) it is predicted that this genotype has a yield 0.3303 kg/plant higher than the mean yield of the population.

(iii) The yield fitted value for genotype c0101 in repetition 2 is given by:

$$\hat{Y}_{c0101,2} = \bar{Y}_{..} + EBLUP(c0101) = 1.196 + 0.3303 = 1.5263 \text{ kg/plant}$$

For the other repetitions of this genotype the yield fitted value will be the same.

(iv) Commands *ranef* and *fitted* of packages “nlme” and “lme4”.

#EBLUPS dos efeitos aleatórios#

ranef(tourigalme1)

#valores ajustados#

fitted(tourigalme1)

#alínea d)#

#ajustamento do modelo com dois factores de efeitos aleatórios: clone e bloco# fitting a model with two factors of random effects: clone and Block#

#com o pacote lme4#

tourigalmer2<-lmer(rend~1+(1|clone)+(1|bloco), data=touriga)

summary(tourigalmer2)

Random model (two factors of random effects, genotype and block), balanced, with **G** and **R** diagonal matrices

$$Y_{ij} = \mu + u_{1i} + u_{2j} + e_{ij}$$

for $i = 1, \dots, 196, j = 1, \dots, 5, n = 196 \times 5$.

Y_{ij} is the observation of yield in the j th block of i th genotype;

μ is a general mean (population);

u_{1i} is the effect of the genotype i (random effect);

u_{2j} is the effect of the block j (random effect);

e_{ij} is the random error associated to the observation Y_{ij} .

$u_{1i}, i. i. d., \mathcal{N}(0, \sigma_{u1}^2), \forall i; u_{2j}, i. i. d., \mathcal{N}(0, \sigma_{u2}^2), \forall j; e_{ij}, i. i. d., \mathcal{N}(0, \sigma_e^2), \forall ij;$

$u_{1i}, u_{2j},$ and e_{ij} are mutually independent, $\forall ij$.

In this model we have 3 variance components: $\sigma_{u1}^2, \sigma_{u2}^2,$ and σ_e^2 .

(ii) Carrying out hypothesis tests for the variance components of the model.

The general procedure to test a variance component is following described.

The likelihood ratio test for a variance component σ_u^2

- Hypotheses: $H_0: \sigma_u^2 = 0$ vs $H_1: \sigma_u^2 > 0$
- The REML likelihood ratio statistic:

$$\Lambda = 2(l_{R_1} - l_{R_0}) \sim \chi_v^2$$

being l_{R_1} the REML log-likelihood of the more general model (full model) and l_{R_0} the REML log-likelihood of the reduced model (that is, the REML log-likelihood under the null hypothesis). Under regularity conditions and under the null hypothesis, the likelihood ratio statistic, has an approximate χ_v^2 distribution with the degrees of freedom (v) equal to the difference in the number of parameters between the two models. However, when we test a variance component, under the null hypothesis

the parameter falls on the boundary of the parameter space. Theoretically it can be shown that for a single variance component, the asymptotic distribution of the REMLRT is a mixture of χ^2 variables, where the mixing probabilities are 0.5, one with 0 degrees of freedom and the other with one degree of freedom. As a consequence we can perform the likelihood ratio test as if the standard conditions apply, and divide the resulting p-value by two.

The REML likelihood ratio test is only valid if the fixed effects are the same for both model.

- Significance level: $\alpha=0.05$
- Rejection region: upper (right-hand) tail
Reject H_0 if $\Lambda_{calc} > \chi^2_{\alpha(v)}$

In this exercise, two tests should be performed: test the genotypic variance component (σ_{u1}^2) and test the block variance component (σ_{u2}^2)

```
#procedimento para realizar de testes de razão de verosimilhanças à componente de variância
genotípica e à componente de variância do bloco; procedure to perform the test the variance
components#
#REML log-verosimilhança do modelo 2; REML of model with both factors#
logLik(tourigalmer2)
#ajustamento do modelo só com o factor de efeitos aleatórios (bloco); fitting a model only with
factor block#
tourigalmer3<-lmer(rend~1+(1|bloco), data=touriga)
#REML log-verosimilhança do modelo 3#
logLik(tourigalmer3)
#ajustamento do modelo só com o factor de efeitos aleatórios (clone); fitting a model only with
factor clone#
tourigalmer4<-lmer(rend~1+(1|clone), data=touriga)
#REML log-verosimilhança do modelo 4#
logLik(tourigalmer4)
#cálculo do p-value do teste; calculate the p-values of the tests#
1-pchisq(206.18,1)
1-pchisq(386.06,1)
#Testes às componentes de variância e comparação de modelos; tests to the variance components
and model comparison#
ranova(tourigalmer2)
ANOVA-like table for random-effects: Single term deletions
Model:
rend ~ (1 | clone) + (1 | bloco)
      npar logLik  AIC  LRT Df Pr(>Chisq)
<none>      4 -719.47 1447.0
(1 | clone)  3 -912.51 1831.0 386.06  1 < 2.2e-16 ***
(1 | bloco)  3 -822.57 1651.1 206.19  1 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To test genotypic variance component (σ_{u1}^2)

- Hypotheses: $H_0: \sigma_{u1}^2 = 0$ vs $H_1: \sigma_{u1}^2 > 0$

- The REML likelihood ratio statistic:

$$\Lambda = 2(l_{R_1} - l_{R_0}) \sim \chi^2_v$$

being l_{R_1} the REML log-likelihood of the more general model (full model, with two factors of random effects) and l_{R_0} the REML log-likelihood of the reduced model (that is, the REML log-likelihood under the null hypothesis, without the factor genotype).

- Significance level: $\alpha=0.05$
- Rejection region: upper (right-hand) tail

$$\text{Reject } H_0 \text{ if } \Lambda_{calc} > \chi^2_{0.05(1)}$$

$$\Lambda_{calc} = 2(l_{R_1} - l_{R_0}) = 2 \times (-719.974 + 912.5058) = 386.06$$

$$\chi^2_{0.05(1)} = 3.84$$

Therefore, $386.06 > 3.84$, we reject H_0 for $\alpha=0.05$. There is genotypic variability of the yield in the variety (or, the estimate of the genotypic variance is significantly different from zero).

To test block variance component (σ_{u2}^2)

- Hypotheses: $H_0: \sigma_{u2}^2 = 0$ vs $H_1: \sigma_{u2}^2 > 0$
- The REML likelihood ratio statistic:

$$\Lambda = 2(l_{R_1} - l_{R_0}) \sim \chi^2_v$$

being l_{R_1} the REML log-likelihood of the more general model (full model, with two factors of random effects) and l_{R_0} the REML log-likelihood of the reduced model (that is, the REML log-likelihood under the null hypothesis, without the factor block).

- Significance level: $\alpha=0.05$
- Rejection region: upper (right-hand) tail

$$\text{Reject } H_0 \text{ if } \Lambda_{calc} > \chi^2_{0.05(1)}$$

$$\Lambda_{calc} = 2(l_{R_1} - l_{R_0}) = 2 \times (-719.974 + 822.5673) = 206.1866$$

$$\chi^2_{0.05(1)} = 3.84$$

Therefore, $206.1866 > 3.84$, we reject H_0 for $\alpha=0.05$. There is block variability of the yield (or, the estimate of the block variance is significantly different from zero).

Note: the code to fit a model with two random effects factor using package nlme

```
#com o pacote nlme,ajustamento de um modelo com dois factores de efeitos aleatórios##
```

```
touriga$Dummy <- factor(1)
```

```
touriga2 <- groupedData(rend ~ 1 | Dummy, touriga)
```

```
lme(rend~ 1, data=touriga2,random=pdBlocked(list(pdIdent(~ 0 + bloco), pdIdent(~ 0 + clone))))
```

(iii) Computing AIC and BIC for both fitted models and select the best one according to the criteria.

Two models were fitted. In item a) a model with one factor (genotype) was fitted and in item d) a model with two factors of random effects (genotype and block) was fitted. As these models have same fixed effects they can be compared using AIC and BIC criteria.

Akaike information criterion (AIC_R) is defined as:

$$AIC_R = -2l_R + 2n_{par} ,$$

l_R is the REML log-likelihood of the model and n_{par} is the number of covariance parameters in the model. The model with the smallest value is chosen as the preferred model.

For the model with one factor of random effects (genotype):

$$AIC_1 = -2 \times (-822.8673) + 2 \times 2 = 1649.1346$$

For the model with two factors of random effects (genotype and block):

$$AIC_2 = -2 \times (-719.474) + 2 \times 3 = 1444.95$$

Therefore, according to this criterion the best model is the model with two factors of random effects (genotype and block).

Bayesian information criterion (BIC_R) is defined as,

$$BIC_R = -2l_R + n_{par} \ln(n - r_X) ,$$

l_R is the REML log-likelihood of the model and n_{par} is the number of covariance parameters in the model, n is the number of observations and r_X is the rank of matrix X . The model with the smallest value is chosen as the preferred model.

For the model with one factor of random effects (genotype):

$$BIC_1 = -2 \times (-822.8673) + 2 \times \ln(980 - 1) = 1658.91$$

For the model with two factors of random effects (genotype and block):

$$BIC_2 = -2 \times (-719.474) + 3 \times \ln(980 - 1) = 1459.61$$

Therefore, according to this criterion the best model is the model with two factors of random effects (genotype and block).

#EXERCISE 2#

```
tomate<-read.table("tomate.txt", header=T)
```

```
#No package lme4, ajustamento do modelo a dois factores de efeitos aleatórios, sem interacção:
```

```
factor variedade de tomate, com 41 níveis, e factor bloco, com 4 níveis#
```

```
tomate.lmer1<-lmer(pesofruto~1+(1|variedade)+(1|bloco), data=tomate)
```

```
summary(tomate.lmer1)
```

```
#o teste de hipóteses adequado para responder ao objectivo do estudo: trata-se de um teste de hipóteses à componente de variância associada ao factor variedade (ver aulas para descrição completa do teste de razão de verosimilhanças)#
```

```
#Para obter o valor da log-verosimilhança restrita do modelo ajustado#
```

```
loglikH1<-logLik(tomate.lmer1)
```

```
loglikH1
```

```
#Para obter o valor da log-verosimilhança restrita do modelo sem o factor variedade#
```

```
tomate.lmer0<-lmer(pesofruto~1+(1|bloco), data=tomate)
loglikH0<-logLik(tomate.lmer0)
loglikH0
#cálculo da estatística do teste#
estcalc<-2*(loglikH1-loglikH0)
estcalc
#p-value do teste#
1-pchisq(estcalc,1)
```

#EXERCISE 3#

```
terrenos<-read.table("terrenos.txt", header=T)
head(terrenos)
```

a) Description of the appropriate model for this study.

Linear mixed model: one factor of fixed effects (factor A, variety with 4 levels), one factor of random effects (factor B, soil with 13 levels), balanced, without interaction, with **G** and **R** diagonal matrices.

$$Y_{ijk} = \mu_1 + \beta_i + u_j + e_{ijk}$$

for $i = 1, \dots, 4, j = 1, \dots, 13, k = 1, n = 4 \times 13$, with $\beta_1 = 0$.

Y_{ijk} is the observation of the yield in the i th variety and j th soil;

μ_1 is a general yield mean (population) in the level 1 of factor A (variety);

β_i is the effect of the variety i (the increased concerning to μ_1), fixed;

u_j is the effect of the soil j , random;

e_{ijk} is the random error associated to the observation Y_{ijk} .

$u_j, i. i. d., \mathcal{N}(0, \sigma_u^2), \forall j; e_{ijk}, i. i. d., \mathcal{N}(0, \sigma_e^2), \forall ijk$; u_j and e_{ijk} are mutually independent, $\forall ijk$.

b) Fitting the adequate model for this study.

#ajustamento do modelo com 1 factor de efeitos fixos (variedade) e 1 factor de efeitos aleatórios (terreno ou local), sem interação, método REML; fitting the model with 1 fixed effects factor (variety) and 1 random effects factor (sites), without interaction, REML method#

```
terrenoslme1<-lme(rend~variedade, random=~1|terreno, data=terrenos)
```

```
summary(terrenoslme1)
```

Linear mixed-effects model fit by REML

Data: terrenos

AIC BIC logLik

55.42708 66.65429 -21.71354

Random effects:

Formula: ~1 | terreno

(Intercept) Residual

StdDev: 0.1604919 0.3123811

Fixed effects: rend ~ variedade

	Value	Std.Error	DF	t-value	p-value
(Intercept)	1.5560000	0.09740463	36	15.974600	0.0000
variedadeB	-0.0238462	0.12252595	36	-0.194621	0.8468
variedadeC	-0.3890769	0.12252595	36	-3.175465	0.0031
variedadeD	-0.3778462	0.12252595	36	-3.083805	0.0039

b)

i) Carrying out the hypothesis test for fixed effects of the model

Tests of hypotheses for linear combinations of the fixed effects of the mixed model ($L^T[\beta]$), when L is a matrix (rank of L greater than 1)

- Hypotheses: $H_0: L^T[\beta] = \mathbf{0}$ vs $H_1: \text{not all } L^T[\beta] = \mathbf{0}$;

In this particular case:

Hypotheses: $H_0: \beta_i = 0, \forall i=2,\dots,4$ vs $H_1: \exists i=2,\dots,4 : \beta_i \neq 0$

- Test statistic: $F = \frac{[\hat{\beta}]^T L(L^T C_{11} L)^{-1} L^T [\hat{\beta}]}{\text{rank}(L)} \sim \mathcal{F}_{v_1, v_2}$, sob H_0

F in general has an approximate F-distribution, with $v_1 = \text{rank}(L)$ and v_2 must be estimated (for example, using *Satterthwaite approximation*). This not happen only for particular cases for data exhibiting certain types of balance and for some special unbalanced cases with the elements of the vectors \mathbf{u} e \mathbf{e} being *i.i.d.* random variables. In these cases, $v_2 = n - r(\mathbf{W})$, where $r(\mathbf{W})$ is the rank of the matrix \mathbf{W} which contains the columns of matrices \mathbf{X} and \mathbf{Z} .

In this particular case, the test for fixed effects is identical to what was described in the context of fixed effects ANOVA: $F = \frac{QMA}{QMRE} \cap F_{(a-1, n-(a+b-1))}$, sob H_0 .

- Significance level : $\alpha=0.05$
- Rejection region: upper (right-hand) tail, Reject H_0 if $F_{calc} > f_{\alpha(v_1, v_2)}$

In this case, Reject H_0 se $F_{calc} > f_{\alpha(a-1, n-(a+b-1))}$

```
#cálculo da estatística do teste aos efeitos fixos#
#matriz de covariâncias estimadas dos estimadores dos efeitos fixos#
vcov(terrenoslme1)
c<-vcov(terrenoslme1)
#Construção da matriz L#
l<-matrix(c(0,1,0,0,0,0,1,0,0,0,0,1),4,3)
ll
  [,1] [,2] [,3]
[1,]  0  0  0
[2,]  1  0  0
[3,]  0  1  0
[4,]  0  0  1
#vector com as estimativas dos efeitos fixos#
b<-c(1.556,-0.0238462, -0.3890769,-0.3778462)
#cálculo do numerador da estatística F do teste aos efeitos fixos#
numerador<-t(b)%*%l%*%solve(t(l)%*%c%*%l)%*%t(l)%*%b
#informação sobre a matriz L, nomeadamente sobre a característica da matriz#
```



```

qr(l)
#cálculo da estatística do teste aos efeitos fixos#
F<-numerador/3
> Fcal
      [,1]
[1,] 6.145421
#valor que numa distribuição F com 3 e 36 graus de liberdade deixa à direita uma região de
probabilidade 0.05; value that in an F distribution with 3 and 36 degrees of freedom leaves to the
right a probability region 0.05#
qf(1-0.05,3,36)
[1] 2.866266
#p-value#
1-pf(6.1464,3,36)
[1] 0.00174488

```

$F_{calc} = 6.145421$; $f_{0.05(3,36)} = 2.866$. Therefore, $F_{calc} > 2.866$, we reject H_0 for $\alpha=0.05$. There are significant effects of the variety in the yield of the wheat.

```

> #teste aos efeitos fixos directamente com o comando anova; testing fixed effects directly with
the anova command#
> anova(terrenoslme1)
      numDF denDF F-value p-value
(Intercept)  1  36 478.2352 <.0001
variedade    3  36  6.1454 0.0017

```

```

#com o package lmerTest, que utiliza objectos do lme4; with the lmerTest package, which uses
objects from lme4#
library(lmerTest)
terrenos<-read.table("terrenos.txt", header=T)
terrenoslmer1<-lmer(rend~variedade+(1|terreno), data=terrenos)
class(terrenoslmer1)
#Teste à componente de variância; Testing the variance component#
ranova(terrenoslmer1)
#Teste aos efeitos fixos; Testing fixed effects"
#por omissão, Satterthwaite's method; by default, Satterthwaite's method#
anova(terrenoslmer1)
#Use the Kenward-Roger method#
if(requireNamespace("pbkrtest", quietly = TRUE))
  summary(terrenoslmer1, ddf="Kenward-Roger")
anova(terrenoslmer1)

```

ii) Is the mean yield of variety B equal to the mean yield of variety A (for $\alpha = 0.05$)?

Tests of hypotheses for linear combinations of the fixed effects of the mixed model ($L^T[\beta]$), being L a non random vector

- Hypotheses : $H_0: L^T[\beta] = \mathbf{0}$ vs $H_1: L^T[\beta] \neq \mathbf{0}$

- Test statistic: $T = \frac{L^T[\hat{\beta}]}{\sqrt{(L^T\hat{C}_{11}L)}} \sim t_{v_2}$, under H_0
- Under the assumed normality of \mathbf{u} and \mathbf{e} , T has an exact t -distribution only for data exhibiting certain types of balance and for some special unbalanced cases. In general, it is only approximately t -distributed, and its degrees of freedom must be estimated (for example, using *Satterthwaite approximation*). This not happen only for particular cases for data exhibiting certain types of balance and for some special unbalanced cases with the elements of the vectors \mathbf{u} e \mathbf{e} being *i.i.d.* random variables. In these cases, $v_2 = n - r(\mathbf{W})$, where $r(\mathbf{W})$ is the rank of the matrix \mathbf{W} which contains the columns of matrices \mathbf{X} and \mathbf{Z} .
- $\sqrt{(L^T\hat{C}_{11}L)}$ is a scalar, is the standard error of the estimator of the parameter being tested, matrix $\hat{C}_{11} = (\mathbf{X}^T\hat{V}^{-1}\mathbf{X})^{-1}$ was defined in theoretical part (slides 61 and 62).
- Significance level: α
- Rejection region: two-tailed; Reject H_0 if $|T_{calc}| > t_{\alpha/2}(v_2)$

In this example, $\beta_2 = \mu_2 - \mu_1$, therefore, the test is:

Hypotheses: $H_0: \beta_2 = 0$ vs $H_1: \beta_2 \neq 0$

$$L^T[\boldsymbol{\beta}] = [0 \quad 1 \quad 0 \quad 0] \begin{bmatrix} \mu_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \beta_2$$

Test statistic: $T = \frac{\hat{\beta}_2}{\hat{\sigma}_{\hat{\beta}_2}} \sim t_{n-(a+b-1)}$, under H_0

$\alpha=0.05$

The output:

```
terrenoslme1 <- lme(rend~variedade, random=~1|terreno, data=terrenos)
summary(terrenoslme1)
```

...

Fixed effects: rend ~ variedade

	Value	Std.Error	DF	t-value	p-value
(Intercept)	1.5560000	0.09740463	36	15.974600	0.0000
variedadeB	-0.0238462	0.12252595	36	-0.194621	0.8468
variedadeC	-0.3890769	0.12252595	36	-3.175465	0.0031
variedadeD	-0.3778462	0.12252595	36	-3.083805	0.0039

$|T_{calc}| = 0.194621$

p-value of the test=0.8468

For any usual significance level, H_0 is not rejected. Therefore, the mean yield of variety B is equal to the mean yield of variety A (or, the sample means of the yield of the two varieties are not significantly different).

#EXERCISE 4#

```
preliminar#  
plot.design(Machines)  
attach(Machines)  
interaction.plot(Machine, Worker, score)  
detach()
```

a) The appropriate model for this study.

Linear mixed model: one factor of fixed effects (factor A, machine with 3 levels), one factor with random effects (factor B, worker with 6 levels), balanced, with interaction (3 observations for each experimental situation), with G and R diagonal matrices.

$$Y_{ijk} = \mu_1 + \beta_i + u_j + (\beta u)_{ij} + e_{ijk}$$

for $i = 1, 2, 3, j = 1, \dots, 6, k = 1, 2, 3, n = 3 \times 6 \times 3$, with $\beta_1 = 0$.

Y_{ijk} is the k th observation of the productivity score in the i th level of factor A (machine) and j th level of factor B (worker);

μ_1 is a general mean (population) in the level 1 of factor A (machine);

β_i is the effect of the level i of the factor A (the increased concerning to μ_1), fixed (**machine**);

u_j is the effect of the level j of the factor B, random (**worker**);

$(\beta u)_{ij}$ is the interaction effect of the i th level of factor A with the j th level of factor B, **random (interaction worker $i \times$ machine j)**;

e_{ijk} is the random error associated to the observation Y_{ijk} .

$u_j, i. i. d., \mathcal{N}(0, \sigma^2_u), \forall j; (\beta u)_{ij}, i. i. d., \mathcal{N}(0, \sigma^2_{\beta u}), \forall ij; e_{ijk}, i. i. d., \mathcal{N}(0, \sigma^2_e), \forall ijk$

$u_j, (\beta u)_{ij},$ and e_{ijk} are mutually independent, $\forall ijk$.

#Ajustamento do modelo descrito na alínea a): modelo linear misto, 1 factor de efeitos fixos (máquina, com 3 níveis, 1 factor de efeitos aleatórios (trabalhador, com 6 níveis), com interacção#
machines1r<-lmer(score~Machine+(1|Worker)+(1|Worker:Machine), data=Machines)

#para ver os resultados do ajustamento#

```
summary(machines1r)
```

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

Formula: score ~ Machine + (1 | Worker) + (1 | Worker:Machine)

Random effects:

Groups	Name	Variance	Std.Dev.
	Worker:Machine (Intercept)	13.9095	3.7295
	Worker (Intercept)	22.8584	4.7811
	Residual	0.9246	0.9616

Number of obs: 54, groups: Worker:Machine, 18; Worker, 6

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	52.356	2.486	8.522	21.062	1.20e-08 ***
MachineB	7.967	2.177	10.000	3.660	0.00439 **
MachineC	13.917	2.177	10.000	6.393	7.91e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

b) The restricted maximum likelihood estimates for the variance components of the model:
 $\hat{\sigma}_u^2 = 22.8584$ (worker); $\hat{\sigma}_{\beta_{u}}^2 = 13.9095$ (interaction worker $_i$ \times machine); $\hat{\sigma}_e^2 = 0.9246$

c) maximum likelihood estimates for the variance components
 #Para ver as estimativas de máxima verosimilhança; To see the maximum likelihood estimates#
 summary(lmer(score~Machine+(1|Worker)+(1|Worker:Machine), REML=FALSE,
 data=Machines))

Linear mixed model fit by maximum likelihood . t-tests use

...

Random effects:

Groups	Name	Variance	Std.Dev.
Worker:Machine	(Intercept)	11.5398	3.3970
Worker	(Intercept)	19.0487	4.3645
Residual		0.9246	0.9616

Number of obs: 54, groups: Worker:Machine, 18; Worker, 6

Maximum likelihood estimates for the variance components are biased downward (whereas REML are based on taking into account the degrees of freedom for the fixed effects in the model, losing 1 degree of freedom for each).

a) Carry out the hypothesis test for worker \times machine interaction. Use a significance level of 0.01

- Hypotheses: $H_0: \sigma_{\beta_{u}}^2 = 0$ vs $H_1: \sigma_{\beta_{u}}^2 > 0$
- The REML likelihood ratio statistic:

$$\Lambda = 2(l_{R_1} - l_{R_0}) \sim \chi_v^2$$

being l_{R_1} the REML log-likelihood of the more general model (full model, with two factors of random effects) and l_{R_0} the REML log-likelihood of the reduced model (that is, the REML log-likelihood under the null hypothesis, without the interaction).

- Significance level: $\alpha=0.05$
- Rejection region: upper (right-hand) tail

$$\text{Reject } H_0 \text{ if } \Lambda_{calc} > \chi_{0.01(1)}^2$$

> #log-verosimilhança restrita do modelo com interação; restricted log-likelihood of the model with interaction"

> lr1<-logLik(machines1r)

> lr1

'log Lik.' -107.8438

> #log-verosimilhança restrita do modelo sem interação; restricted log-likelihood of the model without interaction"

> lr0<-logLik(machines2r)

> lr0

```
'log Lik.' -143.4391
> #cálculo da estatística do teste de razão de verosimilhanças; calculation of the likelihood ratio
test statistic#
> estcal<-2*(lr1-lr0)
> estcal
'log Lik.' 71.19063  $A_{calc}$ 
> qchisq(1-0.01,1)
[1] 6.634897
#conclusão: existe variabilidade significativa da interacção máquinaxtrabalhador, para qualquer
nível de significância usual; conclusion: there is significant variability in the machine-worker
interaction, for any usual level of singificance#
```

b) Carry out the hypothesis test for the variability associated to worker. Use a significance level of 0.01.

Hypotheses: $H_0: \sigma_u^2 = 0$ vs $H_1: \sigma_u^2 > 0$

The same procedure.

```
#comandos para efectuar o teste de hipóteses à componente de variância associada factor trabalhador;
commands to perform the hypothesis test to the variance component associated worker factor#
#ajustamento do modelo sem factor trabalhador; fitting the model without worker factor#
machines3r<-lmer(score~Machine+(1|Worker:Machine), data=Machines)
#log-verosimilhança restrita do modelo sem factor trabalhador; restricted log-likelihood of the model
without worker factor"
lr30<-logLik(machines3r)
lr30
'log Lik.' -110.6278
#cálculo da estatística do teste de razão de verosimilhanças; calculation of the likelihood ratio test
statistic#
estcal2<-2*(lr1-lr30)
estcal2
'log Lik.' 5.568022
#p-value#
1-pchisq(estcal2,1)
0.01829141
```

However, when we test a variance component, under the null hypothesis the parameter falls on the boundary of the parameter space. Theoretically it can be shown that for a single variance component, the asymptotic distribution of the REMLRT is a mixture of χ^2 variables, where the mixing probabilities are 0.5, one with 0 degrees of freedom and the other with one degree of freedom. As a consequence we can perform the likelihood ratio test as if the standard conditions apply, and divide the resulting p-value by two.

```
#p-value corrigido de acordo com a verdadeira distribuição da estatística do teste; corrected according to
the true distribution of the test statistic#
0.5*(1-pchisq(estcal2,1))
'0.009145706
```

#conclusão:: there is significant variability in workers, for a level of significance of 0.01#

```
> ranova(machines1r)
ANOVA-like table for random-effects: Single term deletions
```

Model:

```
score ~ Machine + (1 | Worker) + (1 | Worker:Machine)
npar logLik AIC LRT Df Pr(>Chisq)
<none>          6 -107.84 227.69
(1 | Worker)    5 -110.63 231.26 5.568 1 0.01829 *
(1 | Worker:Machine) 5 -143.44 296.88 71.191 1 < 2e-16 ***
```

f) The ANOVA table for a linear mixed model considering one factor of fixed effects (factor A) and one factor of random effects (factor B), balanced, with interaction, and **G** and **R** diagonal matrices, is described as follows:

	G.L.	S.Q.	QM	E[QM]
Factor A	$a - 1$	SQA	QMA	$\frac{bc}{a-1} \sum_{i=1}^a (\beta_i - \bar{\beta})^2 + \sigma_e^2 + c\sigma_{\beta u}^2$
Factor B	$b - 1$	SQB	QMB	$\sigma_e^2 + c\sigma_{\beta u}^2 + ca\sigma_u^2$
Interaction	$(a - 1)(b - 1)$	$SQAB$	$QMAB$	$\sigma_e^2 + c\sigma_{\beta u}^2$
Residuals	$ab(c - 1)$	$SQRE$	$QMRE$	σ_e^2
TOTAL	$n - 1$	SQT		

The appropriate F-statistic is a quotient of QM that is chosen such that the expected value of the numerator and the expected value of the denominator differ only in the fixed effects of the factor being tested. Thus, for this example, the F-statistic for the test of fixed effects of factor A (Machine) will be obtained with the following procedure.

$$\frac{(a - 1)QMA}{E[QMA]} \cap \chi_{(a-1)}^2$$

$$\frac{(a - 1)(b - 1)QMAB}{E[QMAB]} \cap \chi_{(a-1)(b-1)}^2$$

These are independent random variables, therefore,

$$\frac{\frac{(a-1)QMA}{E[QMA]}}{(a-1)} = \frac{\frac{QMA}{\frac{bc}{a-1} \sum_{i=1}^a (\beta_i - \bar{\beta})^2 + \sigma_e^2 + c\sigma_{\beta u}^2}}{\frac{QMA}{\sigma_e^2 + c\sigma_{\beta u}^2}}$$

$$\frac{(a-1)(b-1)QMAB}{E[QMAB]} = \frac{QMAB}{\sigma_e^2 + c\sigma_{\beta u}^2}$$

Under H_0 ($\beta_i = 0, \forall_i$),

$$\frac{\frac{QMA}{\sigma_e^2 + c\sigma_{\beta u}^2}}{\frac{QMAB}{\sigma_e^2 + c\sigma_{\beta u}^2}} = \frac{QMA}{QMAB} \cap F_{(a-1, (a-1)(b-1))}$$

g) Hypothesis test to assess if there are any major effects associated with machine brands.

> #Para efectuar o teste de hipóteses aos efeitos fixos do modelo, consultar o valor da estatística do teste calculada; To perform the hypothesis test to the fixed effects of the model, see the value of the calculated test statistic#

> anova(machines1r)

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Machine	38.051	19.025	2	10	20.576	0.0002855 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> #valor que numa distribuição F com 2 e 10 graus de liberdade deixa à direita uma região de probabilidade 0.05; value that in an F distribution with 2 and 10 degrees of freedom leaves to the right a probability region 0.05#

> qf(1-0.01,2,10)

[1] 7.559432

> #p-value#

> 1-pf(20.576,2,10)

[1] 0.000285531

In this particular case, test of hypotheses for fixed effects of factor Machine

- Hypotheses: $H_0: \beta_i = 0, \forall_i$ vs $H_1: \exists_i$ tal que $\beta_i \neq 0$
- Test statistic: $F = \frac{QMA}{QMAB} \cap F_{(a-1, (a-1)(b-1))}$, sob H_0
- Significance level: $\alpha=0.01$
- Rejection region: upper (right-hand) tail: unilateral direita
Rejeitar H_0 se $F_{calc} > f_{\alpha(a-1, (a-1)(b-1))}$

Therefore, $F_{calc} = 20.576$, $f_{0.01(2,10)} = 7.559432$. We reject H_0 for $\alpha=0.01$. There are main effects of the machines in the productivity score.

We can confirm also with the following commands:

```
> vcov(machines1r)
3 x 3 Matrix of class "dpoMatrix"
      (Intercept) MachineB MachineC
(Intercept)  6.179352 -2.369611 -2.369611
MachineB     -2.369611  4.739222  2.369611
MachineC     -2.369611  2.369611  4.739222
> c<-vcov(machines1r)
> #Construção da matriz L; Construction of matrix L#
> l<-matrix(c(0,1,0,0,0,1),3,2)
> l
      [,1] [,2]
[1,]  0  0
[2,]  1  0
[3,]  0  1
> #vector com as estimativas dos efeitos fixos; vector with the fixed effects estimates#
> b<-c(52.356,7.967,13.917)
> #cálculo do numerador da estatística F do teste aos efeitos fixos; calculation of the numerator of
the F test statistic of the fixed effects #
> numerador<-t(b)%*%l%*%solve(t(l)%*%c%*%l)%*%t(l)%*%b
> #informação sobre a matriz L, nomeadamente sobre a característica da matriz; information about
the matrix L, namely the rank of the matrix#
$rank
[1] 2
#cálculo da estatística do teste aos efeitos fixos; calculation of the fixed effects test statistic#
> Fcal<-numerador/2
> Fcal
[1,] 20.57711
> #valor que numa distribuição F com 3 e 36 graus de liberdade deixa à direita uma região de
probabilidade 0.05; value that in an F distribution with 3 and 36 degrees of freedom leaves to the
right a probability region 0.05#
> qf(1-0.05,2,10)
[1] 4.102821
> #p-value#
> 1-pf(20.57711,2,10)
[1] 0.0002854912
```

h) Use the commands and comment the results.

```
#gráficos de diagnóstico para validação de alguns dos pressupostos do modelo; diagnostic graphs
to validate some of the model's assumptions#
#plot of residuals vs fitted values"
plot(machines1r)
#Residuals#
residuos<-resid(machines1r)
```



```

qqnorm(residuos)
#qqnorm dos preditores dos efeitos dos trabalhadores; qqnorm of the predictors of workers
effects#
eblupsworker<-ranef(machines1r)$Worker
qqnorm(eblupsworker[,1])
#qqnorm of the predictors of the interaction effects#
eblupsinteracao<-ranef(machines1r)$`Worker:Machine`
qqnorm(eblupsinteracao[,1])

```

EXERCISE 5#

```

antaovaz<-read.table("antaovaz.txt", header=T)
#no package lme4, ajustamento do modelo linear misto com 1 factor de efeitos fixos (clone, com 6
níveis), 1 factor de efeitos aleatórios (ambiente, com 11 níveis), com interacção (por cada
combinação clonexambiente fizeram-se 8 observações), método REML#
antaovazlmer1<-lmer(rend~clone+(1|ambiente)+(1|ambiente:clone), data=antaovaz)
summary(antaovazlmer1)
#Proceda de forma adequada para avaliar se existe variabilidade da interacção: teste de razão de
verosimilhanças à componente de variância associada à interacção cloneXambiente#
#nota: descrever o teste detalhadamente por escrito, hipóteses, estatística do teste, distribuição da
estatística do teste e região crítica##
loglikH1<-logLik(antaovazlmer1)
#ajustamento do modelo sem interacção#
antaovazlmer0<-lmer(rend~clone+(1|ambiente), data=antaovaz)
loglikH0<-logLik(antaovazlmer0)
#cálculo da estatística do teste#
estcalc<-2*(loglikH1-loglikH0)
estcalc
#p-value do teste#
1-pchisq(estcalc,1)

#Proceda de forma adequada para avaliar se existe variabilidade entre ambientes: teste de razão de
verosimilhanças à componente de variância associada ao factor ambiente# #
#nota: descrever o teste detalhadamente por escrito, hipóteses, estatística do teste, distribuição da
estatística do teste e região crítica##
#ajustamento do modelo sem factor ambiente#
antaovazlmer2<-lmer(rend~clone+(1|ambiente:clone), data=antaovaz)
loglikH02<-logLik(antaovazlmer2)
#cálculo da estatística do teste#
estcalc2<-2*(loglikH1-loglikH02)
estcalc2
#p-value do teste#
1-pchisq(estcalc2,1)

#Proceda de forma adequada para avaliar se existem efeitos principais associados aos clones: teste
de hipóteses aos efeitos fixos do modelo#

```

```
#nota: descrever o teste detalhadamente por escrito, hipóteses, estatística do teste, distribuição da estatística do teste e região crítica##  
#consultar o valor da estatística do teste calculada, estatística  $F=QMA/QMAB$ #  
anova(antaoavazlmer1)  
#valor que numa distribuição F com 5 e 50 graus de liberdade deixa à direita uma região de probabilidade 0.05#  
qf(1-0.05,5,50)  
#p-value#  
1-pf(0.5659,5,50)
```

EXERCISE 6#

```
#alínea b)#  
plot.design(Alfalfa)  
attach(Alfalfa)  
interaction.plot(Date, Variety, Yield)  
detach(Alfalfa)
```

```
#alínea c), ajustamento do modelo descrito na alínea a)#  
alfalfalmer1<-  
lmer(Yield~Variety+Date+Variety*Date+(1|Block)+(1|Block:Variety)+(1|Block:Date),  
data=Alfalfa)  
summary(alfalfalmer1)  
logLik(alfalfalmer1)  
#ajustamento do modelo descrito na alínea a), mas sem a interacção Block*Date#  
alfalfalmer2<-lmer(Yield~Variety+Date+Variety*Date+(1|Block)+(1|Block:Variety),  
data=Alfalfa)  
summary(alfalfalmer2)  
logLik(alfalfalmer2)
```

```
#comparação dos dois modelos#  
estcal1<-2*(-4.4728+4.978184)  
estcal1  
1-pchisq(estcal1,1)  
0.5*(1-pchisq(estcal1,1))
```

```
#Foi escolhido o modelo 2, o que se segue é com base neste modelo#  
summary(alfalfalmer2)
```

```
#alínea d), teste aos efeitos fixos#  
anova(alfalfalmer2)  
#p-value do teste aos efeitos da interacção#  
1-pf(1.2548,6,45)  
#p-value do teste aos efeitos principais da variedade#  
1-pf(0.6534,2,10)  
#p-value do teste aos efeitos principais da data de corte#
```

1-pf(23.3897,3,45)

```
#o mesmo modelo ajustado com a função lme do package nlme#
alfafalme<-lme(Yield~Variety*Date, random=~1|Block/Variety, data=Alfalfa)
summary(alfafalme)
anova(alfafalme)
```

```
#alínea e)#
aov(Yield~Variety*Date+Error(Block*Variety), data=Alfalfa)
summary(aov(Yield~Variety*Date+Error(Block*Variety), data=Alfalfa))
```

EXERCISE 6, explanation

6. a) The model:

$$Y_{ijk} = \mu_{11} + \alpha_i + u_j + (\alpha u)_{ij} + \beta_k + (\alpha\beta)_{ik} + (\beta u)_{kj} + e_{ijk}$$

com $i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, c, n = abc$,
e com $\alpha_1 = 0, \beta_1 = 0, (\alpha\beta)_{1k} = 0, \forall k, (\alpha\beta)_{i1} = 0, \forall i$.

Sendo:

Y_{ijk} , rendimento na variedade i (atribuído à grande parcela, *whole-plot*), no bloco j e na data de corte k (atribuído à parcela dividida, *sub-plot* ou *split-plot*); (yield on variety i (assigned to the whole plot), on block j and on cutting date k (assigned to the split plot));

μ_{11} , a média populacional de rendimento na célula (1,1), ou seja, na combinação variedade 1(Cossak) sem corte; (overall yield mean in cell (1,1), i.e., in combination variety 1(Cossak) without cutting);

α_i , o efeito da variedade i (acrécimo), atribuída à grande parcela (*fixo*); (the effect of variety i (increment), assigned to the large plot (fixed));

u_j , o efeito do bloco j (*aleatório*); (the effect of block j (random))

$(\alpha u)_{ij}$, o efeito da interação da variedade i com o bloco j , designado correntemente como *whole-plot error* (*aleatório*); the effect of the interaction of variety i with block j , commonly referred to as whole-plot error (random);

β_k , o efeito da data de corte k (acrécimo), atribuído à parcela dividida (*fixo*); the effect of the date of cutting k (increment), assigned to the split plot (fixed);

$(\alpha\beta)_{ik}$, o efeito da interação da variedade i com a data de corte k (acrécimo) (*fixo*); the effect of the interaction of variety i and cutting date k (addition) (fixed);

$(\beta u)_{kj}$, o efeito da interação da data de corte k com o bloco j (*aleatório*); the effect of the interaction of cutting date k with block j (random);

e_{ijk} , o erro aleatório associado à observação Y_{ijk} . the random error associated to observation Y_{ijk} .

$u_j, i. i. d., \mathcal{N}(0, \sigma^2_u), \forall j$; $(\alpha u)_{ij}, i. i. d., \mathcal{N}(0, \sigma^2_{\alpha u}), \forall ij$; $(\beta u)_{jk} i. i. d., \mathcal{N}(0, \sigma^2_{\beta u}), \forall ij$;

$e_{ijk}, i. i. d., \mathcal{N}(0, \sigma^2_e), \forall ijk$;

$u_j, (\alpha u)_{ij}, (\beta u)_{jk}, e_{ijk}$, mutuamente independentes.

Ajuste o modelo descrito no R recorrendo ao comando *lme* do pacote “*nlme*”.

A análise mais comum incorpora o efeito $(\beta u)_{jk}$ no erro aleatório e_{ijk} , tendo-se no e_{ijk} a parte $(\beta u)_{jk}$ e a parte $(\alpha\beta)_{ijk}$ (designada correntemente como *within plot error*). Mas, antes de optar por esta opção vamos

comparar os dois modelos, fazendo um teste de razão de verossimilhanças restritas ($H_0: \sigma_{\beta u}^2 = 0$ vs $H_1: \sigma_{\beta u}^2 > 0$).

The most common analysis incorporates the effect $(\beta u)_{jk}$ in the random error e_{ijk} , having in the e_{ijk} the part $(\beta u)_{jk}$ and the part $(\alpha\beta u)_{ijk}$ (commonly known as within plot error). But, before choosing this option we will compare the two models, doing a restricted likelihood ratio test.

Modelo descrito na alínea a)

```
alfalfalmer1<-lmer(Yield~Variety+Date+Variety*Date+(1|Block)+(1|Block:Variety)+(1|Block:Date),
data=Alfalfa)
```

```
summary(alfalfalmer1)
```

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

```
lmerModLmerTest]
```

```
Formula: Yield ~ Variety + Date + Variety * Date + (1 | Block) + (1 | Block:Variety) + (1 | Block:Date)
```

```
Data: Alfalfa
```

```
REML criterion at convergence: 8.9
```

```
Scaled residuals:
```

```
  Min    1Q  Median    3Q   Max
-1.70134 -0.52534  0.00645  0.37926  1.71698
```

```
Random effects:
```

```
Groups   Name      Variance Std.Dev.
Block:Date (Intercept) 0.004354 0.06599
Block:Variety (Intercept) 0.028155 0.16780
Block    (Intercept) 0.056722 0.23816
Residual                0.023614 0.15367
```

```
Number of obs: 72, groups: Block:Date, 24; Block:Variety, 18; Block, 6
```

```
Fixed effects:
```

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	1.76500	0.13714	12.50236	12.870	1.40e-08 ***
VarietyLadak	0.11000	0.13136	21.19356	0.837	0.4117
VarietyRanger	-0.06167	0.13136	21.19356	-0.469	0.6436
DateS1	-0.46333	0.09655	42.91952	-4.799	1.96e-05 ***
DateS20	-0.18833	0.09655	42.91952	-1.951	0.0577 .
DateO7	-0.12167	0.09655	42.91952	-1.260	0.2144
VarietyLadak:DateS1	-0.10500	0.12547	30.00000	-0.837	0.4093
VarietyRanger:DateS1	0.17333	0.12547	30.00000	1.381	0.1773
VarietyLadak:DateS20	-0.02333	0.12547	30.00000	-0.186	0.8537
VarietyRanger:DateS20	-0.03167	0.12547	30.00000	-0.252	0.8025
VarietyLadak:DateO7	0.06667	0.12547	30.00000	0.531	0.5991
VarietyRanger:DateO7	0.02833	0.12547	30.00000	0.226	0.8229

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
logLik(alfalfalmer1)
```

```
'log Lik.' -4.4728 -lr1: log-verossimilhança restrita do modelo completo ( $H_1: \sigma_{\beta u}^2 > 0$ )
```

```
alfalfalmer2<-lmer(Yield~Variety+Date+Variety*Date+(1|Block)+(1|Block:Variety), data=Alfalfa)
```

```
- Ajustamento do modelo sem (1 | Block:Date)
```

```
> summary(alfalfalmer2)
```

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

```
lmerModLmerTest]
```

```
Formula: Yield ~ Variety + Date + Variety * Date + (1 | Block) + (1 |
```

```
Block:Variety)
```

```
Data: Alfalfa
```

```
REML criterion at convergence: 10
```

```
Scaled residuals:
```

Min 1Q Median 3Q Max
 -1.87049 -0.56237 0.01084 0.34992 1.79699

Random effects:

Groups	Name	Variance	Std.Dev.
Block:Variety	(Intercept)	0.02707	0.1645
Block	(Intercept)	0.05781	0.2404
Residual		0.02797	0.1672

Number of obs: 72, groups: Block:Variety, 18; Block, 6

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	1.76500	0.13714	12.50820	12.870	1.39e-08 ***
VarietyLadak	0.11000	0.13544	24.08068	0.812	0.4247
VarietyRanger	-0.06167	0.13544	24.08068	-0.455	0.6530
DateS1	-0.46333	0.09655	45.00000	-4.799	1.79e-05 ***
DateS20	-0.18833	0.09655	45.00000	-1.951	0.0574 .
DateO7	-0.12167	0.09655	45.00000	-1.260	0.2141
VarietyLadak:DateS1	-0.10500	0.13655	45.00000	-0.769	0.4459
VarietyRanger:DateS1	0.17333	0.13655	45.00000	1.269	0.2108
VarietyLadak:DateS20	-0.02333	0.13655	45.00000	-0.171	0.8651
VarietyRanger:DateS20	-0.03167	0.13655	45.00000	-0.232	0.8177
VarietyLadak:DateO7	0.06667	0.13655	45.00000	0.488	0.6278
VarietyRanger:DateO7	0.02833	0.13655	45.00000	0.207	0.8366

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

logLik(alfalfalmer2)

'log Lik.' -4.978184 (lr0: log-verosimilhança restrita do modelo sem (1 | Block:Date) ($H_0: \sigma_{\beta u}^2 = 0$))

#comparação dos dois modelos#

estcall<-2*(-4.4728+4.978184) – Cálculo da estatística do teste para a nossa amostra, $\Lambda_{calc} = 2(l_{R_1} - l_{R_0})$

estcall

[1] 1.010768 Λ_{calc}

1-pchisq(estcall,1) – p-value do teste

[1] 0.3147189

Os dois modelos não diferem significativamente para qualquer nível de significância usual, optando-se pelo modelo mais simples, isto é, sem o efeito $(\beta u)_{jk}$.

The two models do not differ significantly for any usual significance level, opting for the simpler model, i.e. without the effect $(\beta u)_{jk}$.

d) Efectue os testes de hipóteses que acha adequados para responder aos objectivos do estudo.

alfalfalmer2<-lmer(Yield~Variety+Date+Variety*Date+(1|Block)+(1|Block:Variety), data=Alfalfa)

Teste de razão de verosimilhanças restritas às componentes do modelo

> ranova(alfalfalmer2)

ANOVA-like table for random-effects: Single term deletions

Model:

Yield ~ Variety + Date + (1 | Block) + (1 | Block:Variety) +

Variety:Date

	npar	logLik	AIC	LRT (Λ_{calc})	Df	Pr(>Chisq)

```

<none>          15  -4.9782  39.956
(1 | Block)      14  -7.9028  43.806  5.8492      1    0.0155837 * (H0: σu2 = 0 vs H1: σu2 > 0).
(1 | Block:Variety) 14  -11.7159 51.432 13.4755      1    0.0002417 ***(H0: σαu2 = 0 vs H1: σαu2 > 0).

```

 Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Rejeita-se a hipótese $\sigma_u^2 = 0$ (variabilidade associada ao factor bloco) para um nível de significância 0.05; Rejeita-se a hipótese $\sigma_{\alpha u}^2 = 0$ (variabilidade associada à interação variedade×bloco) para qualquer nível de significância usual.

We reject the hypothesis $\sigma_u^2 = 0$ (variability associated with the block factor) for a significance level 0.05; We reject the hypothesis $\sigma_{\alpha u}^2 = 0$ (variability associated with the variety×block interaction) for any usual significance level.

Teste aos factores de efeitos fixos do modelo (Testing the fixed-effects factors)

```
> anova(alfalfalmer2)
```

Type III Analysis of Variance Table with **Satterthwaite's method**

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)	
Variety	0.03655	0.01827	2	10	0.6534	0.5412	$H_0: \alpha_i = 0, \forall_i$ vs $H_1: \exists_i: \alpha_i \neq 0$
Date	1.96247	0.65416	3	45	23.3897	2.826e-09	*** $H_0: \beta_k = 0, \forall_k$ vs $H_1: \exists_k: \beta_k \neq 0$
Variety:Date	0.21056	0.03509	6	45	1.2548	0.2973	$H_0: (\alpha\beta)_{ik} = 0, \forall_{i,k}$ vs $H_1: \exists_{i,k}: (\alpha\beta)_{ik} \neq 0$

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

A interação Variedade×Data de Corte não é significativa para qualquer nível de significância usual. Os efeitos principais associados à variedade não são significativos para qualquer nível de significância usual.

Os efeitos principais associados à data de corte são significativos para qualquer nível de significância usual, isto é, a data de corte afecta o rendimento da luzerna.

The interaction Variety×Cutting Date is not significant for any usual significance level.

The main effects associated with variety are not significant for any usual level of significance.

The main effects associated with cutting date are significant for any usual significance level, i.e., cutting date affects the yield of alfalfa.

Nos testes aos efeitos fixos, a estatística do teste e os graus de liberdade estão descritos no seguinte quadro:

In fixed effects tests, the test statistic and degrees of freedom are described as follows:

	G.L.	S.Q.	QM	E[QM]	F
Factor A, variedade	$a - 1$	SQA	QMA	$c\sigma_{au}^2 + \sigma_e^2 + bc \frac{\sum_{i=1}^a (\alpha_i - \bar{\alpha})^2}{a - 1}$	$\frac{QMA}{QMError}$
Bloco	$b - 1$	SQBL	QMBL	$ac\sigma_u^2 + c\sigma_{au}^2 + \sigma_e^2$	
Interação Variedade×Bloco (Whole-plot error)	$(a - 1)(b - 1)$	SQWErr or	QMWError	$c\sigma_{au}^2 + \sigma_e^2$	
Factor B, data corte	$c - 1$	SQB	QMB	$\sigma_e^2 + ab \frac{\sum_{k=1}^c (\beta_k - \bar{\beta})^2}{c - 1}$	$\frac{QMB}{QMRE}$
Interação variedade×data de corte	$(a - 1)(c - 1)$	SQAB	QMAB	$\sigma_e^2 + b \frac{\sum_{i=1}^a \sum_{k=1}^c (\alpha\beta_{ik} - \bar{\alpha}\bar{\beta}_{.})^2}{(a - 1)(c - 1)}$	$\frac{QMAB}{QMRE}$
Resíduos	$a(b - 1)(c - 1)$	SQRE	QMRE	σ_e^2	

EXERCISE 7#

```
#No package "nlme" do R, existe um conjunto de dados (designado por "Oats")#  
library(nlme)  
Oats  
#utilizar o comando factor(nitro) para que esta variável seja considerada como factor#  
nitrof<-as.factor(Oats$nitro)  
#Ajustamento do modelo no R com a função lmer: análise de um split-plot, factor variedade de  
efeitos fixos, factor nitro como de efeitos fixos, interacção variedade*nitro de efeitos fixos, factor  
bloco de efeitos aleatórios, interacção bloco*variedade de efeitos aleatórios, interacção  
bloco*nitro como efeitos aleatórios#  
oatslmer1<-  
lmer(yield~Variety+nitrof+nitrof*Variety+(1|Block)+(1|Block:Variety)+(1|Block:nitrof),  
data=Oats)  
summary(oatslmer1)  
  
#Pela análise do resultado do ajustamento, conclui-se que se pode retirar a interacção  
bloco*nitrof#  
  
#ajustamento do modelo descrito anteriormente, mas sem a interacção Block*nitrof#  
oatslmer2<-lmer(yield~Variety+nitrof+nitrof*Variety+(1|Block)+(1|Block:Variety), data=Oats)  
summary(oatslmer2)  
#Conclusão sobre a influência da variedade e da concentração de nitrogénio sobre o rendimento#  
#testes aos efeitos fixos da interacção, aos efeitos fixos da variedade e aos efeitos fixos da  
concentração de nitrogénio#  
#valor calculado da estatística F para cada caso)  
anova(oatslmer2)  
#p-value do teste aos efeitos da interacção#  
1-pf( 0.3028,6,45)  
#p-value do teste aos efeitos principais da variedade#  
1-pf(1.4853,2,10)  
#p-value do teste aos efeitos principais da concentração de nitrogénio#  
1-pf(37.6856,3,45)
```

Segue-se um exemplo de aplicação do modelo linear misto com variáveis predictoras categóricas e numéricas (análise de covariância) e em que as observações são feitas no mesmo individuo ao longo do tempo*

*As matrizes de correlação utilizadas para este tipo de análise são inspiradas nas séries temporais e na estatística espacial. Para o seu aprofundamento seriam necessárias umas bases teóricas sobre séries temporais e estatística espacial, o que não faz parte desta UC. Assim sendo, vamos apenas

exemplificar a sua aplicação, para que fique registado que são instrumentos actualmente muito utilizados no âmbito dos modelos mistos.

EXERCISE 8#

#alínea a)#

```
plot(BodyWeight)
```

#alínea b), ajustamento do modelo linear misto com a função lme#

```
bodyw1.lme<-(lme(weight~Time*Diet,random=~Time|Rat, data=BodyWeight))
```

```
summary(bodyw1.lme)
```

```
anova(bodyw1.lme)
```

#eblups dos efeitos aleatórios#

```
ranef(bodyw1.lme)
```

#valores ajustados#

```
fitted(bodyw1.lme)
```

#alínea d)#

```
bodyw2.lme<-update(bodyw1.lme, corr=corExp(form=~Time))
```

```
bodyw3.lme<-update(bodyw1.lme, corr=corGaus(form=~Time))
```

```
bodyw4.lme<-update(bodyw1.lme, corr=corSpher(form=~Time))
```

#tratam-se de modelos não encaixados, só podem ser comparados com base nos critérios de informação#

```
anova(bodyw2.lme,bodyw3.lme, bodyw4.lme)
```

```
summary(bodyw2.lme)
```

#alínea e) - são modelos encaixados, pode-se fazer um teste de razão de verosimilhanças##

```
anova(bodyw1.lme,bodyw2.lme)
```

EXERCÍCIO 8, explicação

No R está disponível o conjunto de dados *BodyWeight* (Pinheiro e Bates, 2000), referente ao peso do corpo de ratos (em gramas), ao longo de 64 dias. Mais precisamente, cada rato é pesado no dia 1 e a cada sete dias até ao dia 64, com uma pesagem adicional no dia 44. Há três grupos de ratos, cada grupo sujeito a uma dieta diferente.

```
> head(BodyWeight)
```

```
Grouped Data: weight ~ Time | Rat
```

```
  weight Time Rat Diet
```

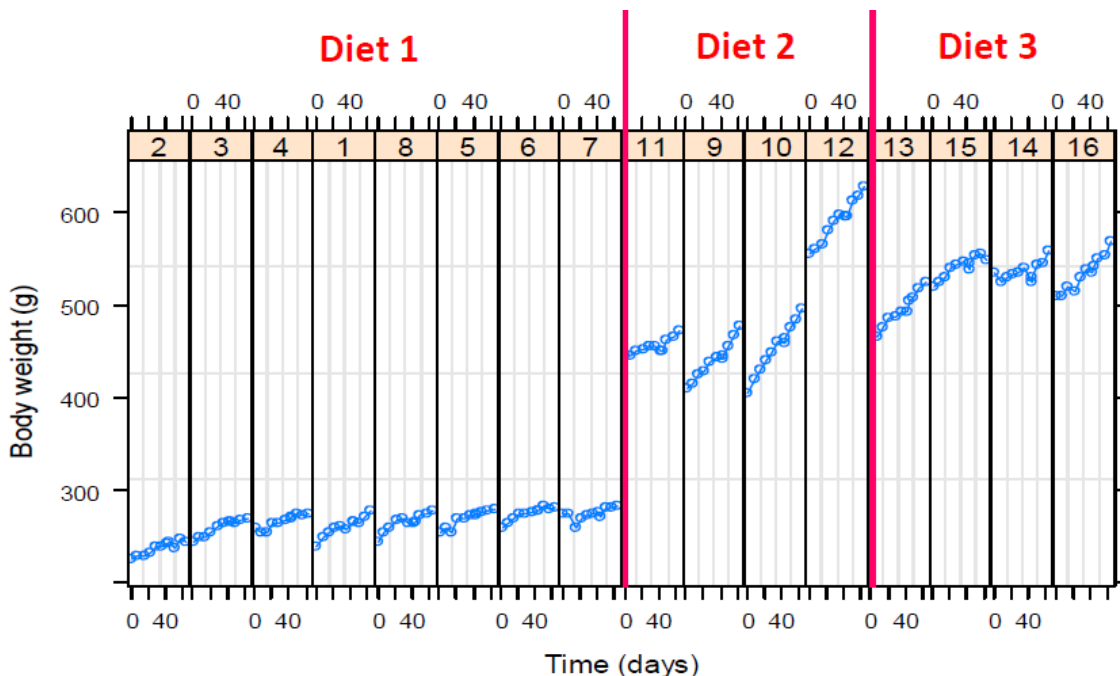
```
1  240   1  1  1
```

```
2  250   8  1  1
```

```
3  255  15  1  1
```

```
4  260  22  1  1
```

a) Execute o comando `plot(BodyWeight)` e comente.



Pode ser observado:

- diferenças entre os três grupos de dieta (dieta, 1 factor, com 3 níveis, de efeitos fixos);
- há evidência de um rato no grupo 2 com um peso corporal inicialmente elevado;
- os pesos corporais parecem crescer linearmente com o tempo (tempo preditora quantitativa, possivelmente com diferentes ordenadas na origem e declives para cada dieta);
- para tomar em conta a variação dos ratos, ordenada na origem e declive para cada rato como efeitos aleatórios. Uma vez que no mesmo indivíduo são realizadas várias pesagens ao longo do tempo, as observações feitas no mesmo rato ao longo do tempo não são independentes (partilham algo em comum).

- differences among the three diet groups can be observed;
- there is evidence of a rat in diet group 2 with an unusually high initial body weight;
- the body weights appear to grow linearly with time, possibly with different intercepts and slopes for each diet.

b) Ajuste o modelo que lhe parecer adequado com a função *lme* do package *nlme*. Execute os comandos *summary*, *anova*, *ranef* e *fitted*. Explique como se obtém cada valor ajustado.

O modelo a ajustar:

$$Y_{ijk} = (\beta_0 + \alpha_{0:i} + b_{0ik(i)}) + (\beta_1 + \alpha_{1:i} + b_{1ik(i)})x_{ij} + e_{ijk}$$

Y_{ijk} – peso do corpo do rato k na dieta i com o tempo j ;

β_0 – ordenada na origem na dieta 1 (fixo);

$\alpha_{0:i}$ – acréscimo na ordenada na origem da dieta i relativamente à dieta 1 (fixo);

$b_{0ik(i)}$ - acréscimo na ordenada na origem do rato k na dieta i (aleatório);

β_1 – declive da dieta 1 (fixo);

$\alpha_{1:i}$ – acréscimo de declive da dieta i relativamente à dieta 1 (fixo);

$b_{1ik(i)}$ – declive do rato k na dieta i (aleatório);

x_{ij} – tempo j na dieta i ;

e_{ijk} – erro aleatório associado à observação Y_{ijk} .

$\mathbf{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{G})$, \mathbf{b}_i é o vector de efeitos aleatórios, admitidos independentes para diferentes i .

$e_{ijk} \sim \mathcal{N}(0, \sigma_e^2)$, $\forall ijk$, admitem-se independentes para diferentes i e j , e independentes dos efeitos aleatórios.

No R, a função `lme` do package `nlme`, será:

```
bodyw1.lme<-(lme(weight~Time*Diet,random=~Time|Rat, data=BodyWeight))
```

```
> summary(bodyw1.lme)
```

Linear mixed-effects model fit by REML

Data: BodyWeight

AIC	BIC	logLik
1171.72	1203.078	-575.8599

Random effects:

Formula: ~Time | Rat

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	36.9390723	(Intr)
Time	0.2484113	-0.149
Residual	4.4436052	

Fixed effects: weight ~ Time * Diet

	Value	Std.Error	DF	t-value	p-value
(Intercept)	251.65165	13.094025	157	19.218816	0.0000
Time	0.35964	0.091140	157	3.946019	0.0001
Diet2	200.66549	22.679516	13	8.847873	0.0000
Diet3	252.07168	22.679516	13	11.114509	0.0000
Time:Diet2	0.60584	0.157859	157	3.837858	0.0002
Time:Diet3	0.29834	0.157859	157	1.889903	0.0606

```
> anova(bodyw1.lme)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	157	1713.1661	<.0001
Time	1	157	82.5924	<.0001
Diet	2	13	85.3505	<.0001
Time:Diet	2	157	7.5743	7e-04

```
> ranef(bodyw1.lme) – EBLUPS dos efeitos aleatórios,  $\tilde{b}_{0i_k(i)}$  e  $\tilde{b}_{1i_k(i)}$ 
```

	(Intercept)	Time
2	-25.1003212	-0.02815920
3	-4.7198954	0.03550692
4	3.7594065	-0.02784546
1	-6.5013805	0.11571393
8	0.2074242	0.04559703
5	4.2867407	0.04307122
6	12.2645516	-0.03781619
7	15.8034741	-0.14606825
11	-11.0426374	-0.56079560
9	-45.1783445	0.04029053
10	-43.5847615	0.34746188
12	99.8057435	0.17304318
13	-40.9098116	0.24060827
15	19.4784670	-0.15244283
14	21.1313288	-0.31816525
16	0.3000158	0.22999981

```
> #valores ajustados#
```

```
> fitted(bodyw1.lme), valores ajustados
```

1	1	1	1	1	1	1	1	1
---	---	---	---	---	---	---	---	---

245.6256 248.9531 252.2806 255.6080 258.9355 262.2630 265.5905 266.0658
 1 1 1 2 2 2 2 2
 268.9179 272.2454 275.5729 226.8828 229.2032 231.5235 233.8439 236.1642
 2 2 2 2 2 2 3 3
 238.4846 240.8050 241.1364 243.1253 245.4457 247.7660 247.3269 250.0929
 3 3 3 3 3 3 3 3
 252.8589 255.6250 258.3910 261.1570 263.9230 264.3182 266.6891 269.4551
 3 4 4 4 4 4 4 4
 272.2211 255.7429 258.0654 260.3880 262.7105 265.0331 267.3556 269.6782
 4 4 4 4 5 5 5 5
 270.0100 272.0007 274.3233 276.6459 256.3411 259.1601 261.9790 264.7980
 5 5 5 5 5 5 5 6
 267.6170 270.4360 273.2549 273.6576 276.0739 278.8929 281.7119 264.2380
 6 6 6 6 6 6 6 6
 266.4908 268.7435 270.9963 273.2491 275.5018 277.7546 278.0764 280.0073
 6 6 7 7 7 7 7 7
 282.2601 284.5129 267.6687 269.1637 270.6587 272.1537 273.6487 275.1437
 7 7 7 7 7 8 8 8
 276.6387 276.8522 278.1337 279.6287 281.1237 252.2643 255.1010 257.9376
 8 8 8 8 8 8 8 8
 260.7743 263.6109 266.4476 269.2842 269.6895 272.1209 274.9575 277.7942
 9 9 9 9 9 9 9 9
 408.1446 415.1849 422.2253 429.2657 436.3061 443.3465 450.3869 451.3926
 9 9 9 10 10 10 10 10
 457.4272 464.4676 471.5080 410.0453 419.2359 428.4265 437.6171 446.8076
 10 10 10 10 10 10 11 11
 455.9982 465.1888 466.5017 474.3794 483.5700 492.7605 441.6792 444.5120
 11 11 11 11 11 11 11 11
 447.3447 450.1775 453.0103 455.8431 458.6759 459.0805 461.5086 464.3414
 11 12 12 12 12 12 12 12
 467.1742 553.2614 561.2311 569.2007 577.1704 585.1400 593.1097 601.0793
 12 12 12 12 13 13 13 13
 602.2178 609.0490 617.0186 624.9883 463.7121 470.0022 476.2923 482.5824
 13 13 13 13 13 13 13 14
 488.8725 495.1626 501.4527 502.3513 507.7428 514.0329 520.3230 525.1945
 14 14 14 14 14 14 14 14
 527.5731 529.9518 532.3305 534.7092 537.0879 539.4665 539.8064 541.8452
 14 14 15 15 15 15 15 15
 544.2239 546.6026 523.7073 527.2461 530.7848 534.3235 537.8623 541.4010
 15 15 15 15 15 16 16 16
 544.9397 545.4453 548.4785 552.0172 555.5560 504.9113 511.1272 517.3430
 16 16 16 16 16 16 16 16
 523.5588 529.7747 535.9905 542.2063 543.0943 548.4222 554.6380 560.8538
 attr("label")
 [1] "Fitted values (g)"

Cada valor ajustado é obtido da forma seguidamente descrita.

Vejam os exemplos:

- Recta ajustada para o rato 1 na dieta 1

$$y = (251.65165 - 6.5013805) + (0.35964 + 0.11571393)x$$

$$y = 245.1503 + 0.4753539x$$

O valor ajustado do rendimento para o rato 1 na dieta 1 e no dia 1, é:

$$\hat{y} = 245.1503 + 0.4753539 \times 1$$

$$\hat{y} = 245.6257$$

e assim sucessivamente até o tempo 64 dias:

$$\hat{y} = 245.1503 + 0.4753539 \times 64$$

$$\hat{y} = 275.5729$$

- Recta ajustada para o rato 9 na dieta 2

$$y = (251.65165 + 200.66549 - 45.1783445) + (0.35964 + 0.60584 + 0.04029053)x$$

$$y = 407.1388 + 1.005771x$$

O valor ajustado do rendimento para o rato 9 na dieta 2 e no dia 1, é:

$$\hat{y} = 407.1388 + 1.005771 \times 1$$

$$\hat{y} = 408.1446$$

e assim sucessivamente.

- Recta ajustada para o rato 13 na dieta 3

$$y = (251.65165 + 252.07168 - 40.9098116) + (0.35964 + 0.29834 + 0.24060827)x$$

$$y = 462.8135 + 0.8985883x$$

O valor ajustado do rendimento para o rato 13 na dieta 3 e no dia 1, é:

$$\hat{y} = 463.7121$$

e assim sucessivamente.

c) Um vez que as observações não são igualmente espaçadas no tempo (note-se que se fez uma observação adicional no dia 44), é mais adequado considerar uma estrutura de correlação espacial para os erros aleatórios. Faz sentido que observações mais próximas no tempo estejam mais correlacionadas. No package *nlme* existem várias estruturas de correlação que podem ser aplicáveis ao caso em estudo, por exemplo, *corExp*, *corGaus*, *corSpher*. Efectue os comandos:

```
bodyw2.lme<-update(bodyw1.lme, corr=corExp(form=~Time))
```

```
bodyw3.lme<-update(bodyw1.lme, corr=corGaus(form=~Time))
```

```
bodyw4.lme<-update(bodyw1.lme, corr=corSpher(form=~Time)).
```

Com base no AIC e no BIC qual é a estrutura de correlação mais adequada?

```
> bodyw2.lme<-update(bodyw1.lme, corr=corExp(form=~Time))
```

```
> bodyw3.lme<-update(bodyw1.lme, corr=corGaus(form=~Time))
```

```
> bodyw4.lme<-update(bodyw1.lme, corr=corSpher(form=~Time))
```

```
> anova(bodyw2.lme,bodyw3.lme, bodyw4.lme)
```

Model	df	AIC	BIC	logLik
bodyw2.lme	1 11	1154.059	1188.553	-566.0296
bodyw3.lme	2 11	1157.233	1191.727	-567.6167
bodyw4.lme	3 11	1157.233	1191.727	-567.6167

Sabemos que o melhor modelo é que tem menores valores de AIC e BIC, portanto, o modelo com uma estrutura de correlação exponencial é o que tem melhor ajustamento. Vejamos o resultado do seu ajustamento.

```
> summary(bodyw2.lme)
```

Linear mixed-effects model fit by REML

Data: BodyWeight

	AIC	BIC	logLik
	1154.059	1188.553	-566.0296

Random effects:

Formula: ~Time | Rat

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	36.9746832	(Intr)
Time	0.2434623	-0.149
Residual	4.5501115	

Correlation Structure: Exponential spatial correlation

Formula: ~Time | Rat

Parameter estimate(s):

range
3.650438

Fixed effects: weight ~ Time * Diet

	Value	Std.Error	DF	t-value	p-value
(Intercept)	251.59245	13.116740	157	19.181020	0.0000
Time	0.36030	0.090468	157	3.982590	0.0001
Diet2	200.68891	22.718860	13	8.833582	0.0000
Diet3	252.31519	22.718860	13	11.105979	0.0000
Time:Diet2	0.62552	0.156695	157	3.991948	0.0001
Time:Diet3	0.31095	0.156695	157	1.984401	0.0490

Correlation:

	(Intr)	Time	Diet2	Diet3	Tm:Dt2
Time		-0.163			
Diet2		-0.577	0.094		
Diet3		-0.577	0.094	0.333	

Time:Diet2 0.094 -0.577 -0.163 -0.054
Time:Diet3 0.094 -0.577 -0.054 -0.163 0.333

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-3.562580701	-0.499770731	0.004618437	0.507775915	2.164527296

Number of Observations: 176

Number of Groups: 16

d) O modelo selecionado em c) difere significativamente do modelo ajustado em b)?

> anova(bodyw1.lme,bodyw2.lme)

Model	df	AIC	BIC	logLik	Test L.Ratio	p-value
bodyw1.lme	1	10	1171.720	1203.078	-575.8599	
bodyw2.lme	2	11	1154.059	1188.553	-566.0296	1 vs 2 19.66057 <.0001

Pode ser feito um teste de razão de verossimilhanças restritas, os dois modelos são encaixados (diferem apenas em 1 parâmetro, o range, distância a partir da qual as observações não estão correlacionadas). Pelo resultado do teste conclui-se que o segundo modelo (com estrutura de correlação exponencial) tem um ajustamento significativamente melhor que o primeiro modelo, para qualquer nível de significância usual.