Modelos Matemáticos e Aplicações

Módulo 2: Modelos lineares mistos

Linear Mixed Models

Some particular cases and respective application - I

2017-2018

Elsa Gonçalves
ISA/UL

Case 1

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

($G = \sigma_u^2 \, I_q$, $R = \sigma_e^2 \, I_n$)

One factor of random effects, balanced

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

for $i = 1, \ldots, a, j = 1, \ldots, b\ , n = ab.$

Y_{ij} is the jth observation in the ith level of factor A;

μ is a general mean (population);

u_i is the effect of the level i of the factor A (random effects);

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

- $u_i, i.i.d., \mathcal{N}\left(0, \sigma^2{}_u\right), \forall i$
- $e_{ij}, i.i.d., \mathcal{N}\left(0, \sigma^2{}_e\right), \forall ij$

The sums of squares are defined as in the case of fixed effects:

$$SQT = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij} - \bar{Y}_{..}\right)^2$$

$$SQA = \sum_{i=1}^{a} b\,(\bar{Y}_{i.} - \bar{Y}_{..})^2$$

$$SQRE = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij} - \bar{Y}_{i.}\right)^2$$

$$SQT = SQA + SQRE$$

- **Estimators for variance components:** procedure based on expected mean squares from the analysis of variance (ANOVA)

$$\mathrm{E}[SQA] = E[\sum_{i=1}^{a} b\,(\bar{Y}_{i.} - \bar{Y}_{..})^2] = (a-1)\left(b\sigma_u^{\,2} + \sigma_e^{\,2}\right)$$

$$\mathrm{E}[QMA] = \frac{E[SQA]}{(a-1)} = b\sigma_u^{\,2} + \sigma_e^{\,2}$$

$$E[SQRE] = E\left[\sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij} - \bar{Y}_{i.}\right)^2\right] = a(b-1)\sigma_e^{\,2}$$

$$E[QMRE] = \frac{E[SQRE]}{a(b-1)} = \sigma_e^{\,2}$$

Equating sums of squares in their expected values, gives:

$$SQA = (a-1)\left(b\hat{\sigma}_u^{\,2} + \hat{\sigma}_e^{\,2}\right)$$

$$SQRE = a(b-1)\hat{\sigma}_e^{\,2}$$

The estimators are:

$$\hat{\sigma}_e^{\,2} = \frac{SQRE}{a(b-1)} = QMRE \qquad \hat{\sigma}_u^{\,2} = \left(\frac{SQA}{a-1} - \hat{\sigma}_e^{\,2}\right)\Big/ b = \frac{QMA - QMRE}{b}$$

- **Maximum likelihood estimation**

For a model with one factor of random effects and balanced, the log-likelihood is given by:

$$l=\ln L=-\frac{1}{2}n\ln 2\pi-\frac{1}{2}a\left[\ln\left(\sigma_e^2+b\sigma_u^2\right)\right]-\frac{1}{2}a(b-1)\ln\sigma_e^2-\frac{\sum_i\sum_j\left(y_{ij}-\mu\right)^2}{2\sigma_e^2}+\frac{b^2\sigma_u^2\sum_i\left(\bar{y}_{i.}-\mu\right)^2}{2\sigma_e^2\left(\sigma_e^2+b\sigma_u^2\right)}.$$

With some manipulation and rearranged so as to display SQA e SQRE (the sums of squares of ANOVA) and equating to zero the partial derivatives of lnL with respect to μ, ${\sigma_e}^2$ and ${\sigma_u}^2$, the following solutions are obtained:

$$\dot{\mu}=\bar{y}_{..}$$

$$\dot{\sigma}_e{}^2=QMRE$$

$$\dot{\sigma}_u{}^2=\frac{\left(1-\frac{1}{a}\right)QMA-QMRE}{b}$$

These are the solutions to the maximum likelihood equations. But they are not necessarily the maximum likelihood estimators. It is necessary to verify if the matrix of second derivatives (Hessian matrix) is definite negative when the parameters in the Hessian are replaced by the solutions used. And ML estimators must be in the parameter space:

$$-\infty<\mu<+\infty,0<{\sigma_e}^2<\infty,0\leq{\sigma_u}^2<\infty$$

The maximum likelihood estimators for variance components are:

$$\begin{cases} \hat{\sigma}_u{}^2 = \dfrac{\left(1-\frac{1}{a}\right)QMA - QMRE}{b}, & \text{if } \left(1-\frac{1}{a}\right)QMA \geq QMRE, \\ \hat{\sigma}_u{}^2 = 0, & \text{otherwise} \end{cases}$$

$$\begin{cases} \hat{\sigma}_e{}^2 = QMRE, & \text{if } \left(1-\frac{1}{a}\right)QMA \geq QMRE, \\ \hat{\sigma}_e{}^2 = \dfrac{SQT}{ab}, & \text{otherwise} \end{cases}$$

- **Restrited maximum likelihood estimation for variance components**

For a model with one factor of random effects and balanced, the restrited log-likelihood (l_R) is given by:

$$l_R = -\frac{1}{2}(ab-1)\ln 2\pi - \frac{1}{2}\ln ab - \frac{1}{2}a(b-1)\ln \sigma_e^2 - \frac{1}{2}(a-1)\ln \lambda - \frac{SQRE}{2\sigma_e^2} - \frac{SQA}{2\lambda}.$$

with $\lambda = \sigma_e^2 + b\sigma_u^2$

Equating to zero the partial derivatives of l_R with respect to $\sigma_e{}^2$ and $\sigma_u{}^2$, the following solutions are obtained:

$$\dot{\sigma}_e{}^2 = \frac{SQRE}{a(b-1)} = QMRE$$

$$\dot{\sigma}_u{}^2 = \frac{QMA - QMRE}{b}$$

The restricted maximum likelihood estimators for variance components are:

$$\begin{cases} \hat{\sigma}_u{}^2 = \frac{QMA - QMRE}{b}, & \text{se } QMA \geq QMRE, \\ \hat{\sigma}_u{}^2 = 0, & \text{caso contrário} \end{cases}$$

$$\begin{cases} \hat{\sigma}_e{}^2 = QMRE, & \text{se } QMA \geq QMRE, \\ \hat{\sigma}_e{}^2 = \frac{SQT}{ab-1}, & \text{caso contrário} \end{cases}$$

Asymptotic covariance matrix for REML estimators

$$\operatorname{var}\begin{bmatrix}\hat{\sigma}_e{}^2\\ \hat{\sigma}_u{}^2\end{bmatrix} \approx \begin{bmatrix}\dfrac{2\sigma_e{}^4}{a(b-1)} & \dfrac{-2\sigma_e{}^4}{ab(b-1)}\\ & \dfrac{2\sigma_e{}^4}{b^2}\left[\dfrac{(\sigma_e{}^2+b\sigma_u{}^2)^2/\sigma_e{}^4}{a}+\dfrac{1}{a(b-1)}\right]\end{bmatrix}$$

ANOVA TABLE: random model with one factor of random effects (Factor A), balanced with $G=\sigma_u^2\, I_q$, $R=\sigma_e^2\, I_n$

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

for $i = 1, \dots, a$, $j = 1, \dots, b$, $n = ab$.

	G.L.	S.Q.	QM	E[QM]	F
Factor A	$a-1$	$SQA = \sum_{i=1}^{a} b\,(\bar{y}_{i.} - \bar{y}_{..})^2$	$QMA = \frac{SQA}{a-1}$	$b\sigma_u^2 + \sigma_e^2$	$\frac{QMA}{QMRE}$
Resíduals	$a(b-1)$	$SQRE = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{i.})^2$	$QMRE = \frac{SQRE}{a(b-1)}$	σ_e^2	
TOTAL	$ab-1$	$SQT = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{..})^2$			

Hypothesis test for variance component associated to factor A

- Hypotheses: $H_0: \sigma_u^{\ 2} = 0 \quad vs \quad H_1: \sigma_u^{\ 2} > 0$
- Test statistic: $F = \frac{QMA}{QMRE} \cap \mathcal{F}_{(a-1,a(b-1))}$, under H_0
- Significance level: α
- Rejection region: upper (right-hand) tail

Reject H_0 if $F_{calc} > f_{\alpha(a-1,a(b-1))}$

Note:

$$\frac{SQRE}{\sigma_e^{\ 2}} \sim \chi^2{}_{a(b-1)}$$

$$\frac{SQA}{b\sigma_u^{\ 2} + \sigma_e^{\ 2}} \sim \chi^2{}_{a-1}$$

Independent random variables,

$$\frac{\frac{QMA}{b\sigma_u^{\ 2} + \sigma_e^{\ 2}}}{\frac{QMRE}{\sigma_e^{\ 2}}} \sim \mathcal{F}_{a-1,ab-a}$$

Likelihood ratio tests for variance component ${\sigma_u}^2$

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

$$\Lambda = 2\left(l_{R_1} - l_{R_0}\right) \sim \chi_\nu^2$$

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_ν^2 distribution with the degrees of freedom (ν) 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 variates , 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: α
- Rejection region: upper (right-hand) tail

Reject H_0 if $\Lambda_{calc} > {\chi^2}_{\alpha(\nu)}$

Example: 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:

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

Exercise 1

In a grapevine selection study to evaluate the genetic variability of the yield of the Touriga Nacional variety, a field trial was installed in Vila Nova de Fozcoa, with a random sample of genotypes (196 genotypes) of the variety. In the field, each genotype was randomly assigned in 5 plots (trial with 5 replicates). The yield (kg/plant) data obtained in 1994 are available in *data.frame* touriga.

a) Describe the adequate model to study the yield genetic variability of the variety.

b) Use the command *aov* of R to obtain the ANOVA table of the model previously described.

(i) What are the variance components estimates involved in the model described in item a?

(ii) With the available information, carry out an hypotesis test for yield genetic variance of the variety (use a significance level of 0.05).

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

Exercise 1 (cont.)

c) Fit the model previously described, with the restricted maximum likelihood (REML) method.

(i) Use *lme* of the package“nlme”, and *lmer* of the package “lme4”;

(ii) Apply command *summary* to the two objects above created and identify the REML estimates for variance componentes. Compare the results with those obtained in item b(i).

(iii) What is the yield fitted value for clone c0101 in repetition 2?

(iv) Explore commands *ranef* and *fitted* of packages “nmle” and “lme4”.

Exercise 1 (cont.)

d) In fact, the Touriga Nacional field trial above described was planted according to a randomized complete block design (5 blocks).

(i) Fit a new model considering the block effect (assuming a random effects factor). Use package *lme4*.

(ii) Carry out hypotesis tests for variance componets of the model.

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

Case 2

Linear mixed model: one factor of fixed effects, one factor of random effects, balanced, without interaction and with interaction, with G and R diagonal matrices ($G=\sigma_u{}^2\ I_q$, $R=\sigma_e{}^2\ I_n$)

Linear mixed model: one factor of fixed effects (factor A), one factor of random effects (factor B), balanced, without interaction

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

for $i = 1, \ldots, a, j = 1, \ldots, b, k = 1, \ldots, c, n = abc$, with $\beta_1 = 0$.

Y_{ijk} is the observation in the ith level of factor A and jth level of factor B;
μ_1 is a general mean (population) in the level 1 of factor A;
β_i is the effect of the level i of the factor A ((the increased concerning to μ_1), fixed;
u_j is the effect of the level j of the factor B, random;
e_{ijk} is the random error associated to the observation Y_{ijk} .

- $u_j, i.i.d., \mathcal{N}\left(0, \sigma^2{}_u\right), \forall j$
- $e_{ijk}, i.i.d., \mathcal{N}\left(0, \sigma^2{}_e\right), \forall ijk$

The sums of squares are defined as :

$$SQT = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\left(Y_{ijk} - \bar{Y}_{...}\right)^2$$

$$SQA = \sum_{i=1}^{a} bc\left(\bar{Y}_{i..} - \bar{Y}_{..,}\right)^2$$

$$SQB = \sum_{j=1}^{b} ac\left(\bar{Y}_{.j.} - \bar{Y}_{...,}\right)^2$$

$$SQRE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\left(Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...,}\right)^2$$

$$SQT = SQA + SQB + SQRE$$

- **Estimators for variance components:** procedure based on expected mean squares from the analysis of variance (ANOVA)

$$\mathrm{E}[SQB] = (b-1)\left(ac\sigma_u{}^2 + \sigma_e{}^2\right)$$

$$\mathrm{E}[QMB] = \frac{E[SQB]}{(b-1)} = ac\sigma_u{}^2 + \sigma_e{}^2$$

$$E[SQRE] = n-(a+b-1)\sigma_e{}^2$$

$$E[QMRE] = \frac{E[SQRE]}{n-(a+b-1)} = \sigma_e{}^2$$

The estimators are:

$$\hat{\sigma}_e{}^2 = \frac{SQRE}{n-(a+b-1)} = QMRE \qquad \hat{\sigma}_u{}^2 = \frac{QMB-QMRE}{ac}$$

- **The maximum likelihood estimators for variance components are($\hat{\sigma}_u{}^2 \geq 0$)**

$$\hat{\sigma}_e{}^2 = \left[1 - \frac{a-1}{b(ac-1)}\right] QMRE,$$

$$\hat{\sigma}_u{}^2 = \frac{SQB/b - \hat{\sigma}_e{}^2}{ac}$$

- **The restricted maximum likelihood estimators for variance components are ($\hat{\sigma}_u{}^2 \geq 0$):**

$$\hat{\sigma}_e{}^2 = \frac{SQRE}{n-(a+b-1)} = QMRE \qquad \hat{\sigma}_u{}^2 = \frac{QMB - QMRE}{ac}$$

Asymptotic variance matrix for REML estimators

$$\operatorname{var}\begin{bmatrix}\hat{\sigma}_e{}^2\\ \hat{\sigma}_u{}^2\end{bmatrix} \approx \frac{2\sigma_e{}^4}{b(ac-1)}\begin{bmatrix}1 & \dfrac{-1}{ac}\\ \dfrac{-1}{ac} & \left[\dfrac{1+(ac-1)(1+ac\sigma_u{}^2/\sigma_e{}^2)^2}{a^2c^2}\right]\end{bmatrix}$$

ANOVA TABLE: linear mixed model, one factor of fixed effects (Factor A) and one factor of random effects (factor B), balanced, with ***G**=σ_u^2 I_q, **R**=σ_e^2 I_n*

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

for $i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, c, n = abc$, with $\beta_1 = 0$.

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

Hypothesis test for variance component associated to factor B

- Hypotheses : $H_0: {\sigma_u}^2 = 0 \quad vs \quad H_1: {\sigma_u}^2 > 0$

- Test statistic : $F = \frac{QMB}{QMRE} \cap F_{(b-1, n-(a+b-1)\ )}$, under H_0

- Significance level : α

- Rejection region : upper (right-hand) tail

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

Or, a **Likelihood ratio test for the variance component.**

Hypothesis test for fixed effects

- Hypotheses $H_0: \beta_i = 0, \ \forall_{i=2,\ldots a} \ \ vs \ \ H_1: \exists_{i=2,\ldots,a} : \beta_i \neq 0$

- Test statistic : $F = \frac{QMA}{QMRE} \cap F_{(a-1,n-(a+b-1)\,)}$, sob H_0

- Significance level : α

- Rejection region : upper (right-hand) tail

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

Note: the test for fixed effects is identical to what was described in the context of fixed effects ANOVA

Exercise 3

Consider the data *data.frame* terrenos. The objective of the study is to compare the yield among four wheat varieties. In addition, 13 sites with different soil conditions were identified. Consider that those sites are a random sample of the sites where the four varieties of wheat will be grown. The four varieties were assigned at random within sites, each variety once per site.

a) Fit the adequate model for this study (for example, using *package nlme*, function *lme*).

ei) Carry out the hypothesis test for fixed effects of the model. For the calculation of the test statistic recall the hypothesis tests for linear combinations of fixed effects of the linear mixed model given in the theoretical classes. Consider the estimated covariance matrix of the fixed effects estimators (vcov (terrenolme1)), define the matrix L, create the vector with the fixed effects estimates and, with the help of R, compute the test statistic. For your conclusions, use the significance level of 0.05. At the end, run anova (terrenolme1).

eii) Is there a decrease in yield of variety B compared to variety A (for $\alpha = 0.05$)?

Linear mixed model: one factor of fixed effects (factor A), one factor with random effects (factor B), balanced, with interaction

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

for $i = 1, \ldots, a, j = 1, \ldots, b, k = 1, \ldots, c, n = abc$, with $\beta_1 = 0$.

Y_{ijk} is the kth observation in the ith level of factor A and jth level of factor B;

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

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

u_j is the effect of the level j of the factor B, random;

$(\beta u)_{ij}$ is the interaction effect of the ith level of factor A with the jth level of factor B, random;

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

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

The sums of squares are defined as :

$$SQT = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\left(Y_{ijk} - \bar{Y}_{...}\right)^2$$

$$SQA = \sum_{i=1}^{a} bc\,(\bar{Y}_{i..} - \bar{Y}_{...})^2$$

$$SQB = \sum_{j=1}^{b} ac\left(\bar{Y}_{.j.} - \bar{Y}_{...}\right)^2$$

$$SQAB = \sum_{i=1}^{a}\sum_{j=1}^{b} c\left(Y_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}\right)^2$$

$$SQRE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\left(Y_{ijk} - \bar{Y}_{ij.}\right)^2$$

$$SQT = SQA + SQB + SQAB + SQRE$$

- **Estimators for variance components:** procedure based on expected mean squares from the analysis of variance (ANOVA)

$$\mathrm{E}[SQB] = (b-1)\left(ac{\sigma_u}^2 + c{\sigma_{\beta u}}^2 + {\sigma_e}^2\right)$$

$$\mathrm{E}[QMB] = \frac{E[SQB]}{(b-1)} = ac{\sigma_u}^2 + c{\sigma_{\beta u}}^2 + {\sigma_e}^2$$

$$\mathrm{E}[SQAB] = (a-1)(b-1)\left(c{\sigma_{\beta u}}^2 + {\sigma_e}^2\right)$$

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

$$E[SQRE] = ab(c-1){\sigma_e}^2$$

$$E[QMRE] = \frac{E[SQRE]}{ab(c-1)} = {\sigma_e}^2$$

These yield the estimators

$$\hat{\sigma}_e{}^2 = QMRE$$

$$\hat{\sigma}_{\beta u}{}^2 = \frac{QMAB - QMRE}{c}$$

$$\hat{\sigma}_u{}^2 = \frac{QMB - QMAB}{ac}$$

- **The maximum likelihood estimators for variance components are ($\hat{\sigma}_u{}^2 \geq 0, \hat{\sigma}_{\beta u}{}^2 \geq 0$)**

$$\hat{\sigma}_e{}^2 = QMRE$$

$$\hat{\sigma}_u{}^2 = \frac{(1-\frac{1}{b})(QMB-QMAB)}{ac}$$

$$\hat{\sigma}_{\beta u}{}^2 = \frac{(1-\frac{1}{b})QMAB - QMRE}{c}$$

- **The restricted maximum likelihood estimators for variance components are ($\hat{\sigma}_u{}^2 \geq 0, \hat{\sigma}_{\beta u}{}^2 \geq 0$)**

$$\hat{\sigma}_e{}^2 = QMRE$$

$$\hat{\sigma}_{\beta u}{}^2 = \frac{QMAB - QMRE}{c}$$

$$\hat{\sigma}_u{}^2 = \frac{QMB - QMAB}{ac}$$

Asymptotic variance matrix for REML estimators

$$\operatorname{var}\begin{bmatrix}\hat{\sigma}_e{}^2\\ \hat{\sigma}_u{}^2\\ \hat{\sigma}_{\beta u}{}^2\end{bmatrix} \approx \frac{2}{b}\begin{bmatrix}\dfrac{\sigma_e{}^4}{a(c-1)} & 0 & \dfrac{-\sigma_e{}^4}{ac(c-1)}\\ & \dfrac{\dfrac{(\sigma_e{}^2+c\sigma_{\beta u}{}^2)^2}{a-1}+(\sigma_e{}^2+c\sigma_{\beta u}{}^2+ac\sigma_u{}^2)^2}{a^2c^2} & \dfrac{-(\sigma_e{}^2+c\sigma_{\beta u}{}^2)^2}{ac^2(a-1)}\\ & & \dfrac{1}{c^2}\left[\dfrac{(\sigma_e{}^2+c\sigma_\beta{}^2)^2}{a-1}+\dfrac{\sigma_e{}^4}{a(c-1)}\right]\end{bmatrix}$$

ANOVA TABLE: linear mixed model, one factor of fixed effects (factor A) and one factor of random effects (factor B), balanced, with interaction

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

for $i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, c, n = abc$, with $\beta_1 = 0$.

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

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

Hypothesis test for variance component associated to interaction

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

- Test statistic: $F = \frac{QMAB}{QMRE} \cap \mathcal{F}_{\left((a-1)(b-1),ab(c-1)\right)}$, under H_0

- Significance level: α

- Rejection region: upper (right-hand) tail

$$\text{Reject } H_0 \text{ if } F_{calc} > f_{\alpha\left((a-1)(b-1),ab(c-1)\right)}$$

Or, a **Likelihood ratio test.**

Hypothesis test for variance component associated to factor B

- Hypotheses: $H_0: {\sigma_u}^2 = 0 \quad vs \quad H_1: {\sigma_u}^2 > 0$
- Test statistic: $F = \frac{QMB}{QMAB} \cap \mathcal{F}_{(b-1,(a-1)(b-1))}$, under H_0
- Significance level: α
- Rejection region: upper (right-hand) tail

Reject H_0 if $F_{calc} > f_{\alpha(b-1,(a-1)(b-1))}$

Or, a **Likelihood ratio test.**

Tests of hypotheses , fixed effects of factor A

- Hypotheses: $H_0: \beta_i = 0, \ \forall_{i=2,\ldots a} \ \ vs \ \ H_1: \exists_{i=2,\ldots,a}$ 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: α
- Rejection region: upper (right-hand) tail: unilateral direita

Rejeitar H_0 se $F_{calc} > f_{\alpha(a-1,(a-1)(b-1))}$

Some considerations

- For random or complex mixed models there are no exact statistical tests for certain model effects (the numerator and denominator of the F statistics are linear combinations of mean squares). In these cases, approximate F tests are performed. One of the classic methods most used for this approach is the method of Satterthwaite (1941). However, other methods are implemented in more complex mixed models frequently reported in the literature and commonly used in several packages, for example, the methods of Giesbrecht and Burns (1985) and Kenward and Roger (1997). (next slide, additional information)

Example: Satterthwaite Degrees of freedom Approximation

Satterthwaite showed that given the ratio

$$\frac{X^2_{num}/\nu_1}{X^*_2/\upsilon^*_2}$$

where $X^2_{num} \cap \chi^2_{\nu_1}$ and X^*_2 is a linear combination of chi-square random variable all independent of X^2_{num}, the $X^*_2 \cap \chi^2_{\upsilon^*_2}$, where

$$\upsilon^*_2 \cong \frac{\left(\sum_m l_m X^2_m\right)^2}{\sum_m (l_m X^2_m)/df_m},$$

X^2_m denotes the $\chi^2_{df_m}$ random variables, l_m denote the constants in the linear combination, df_m the degrees of freedom for the respective X^2_m.

Note

- There are no exact confidence intervals for the variance components associated with the random effects of the model (the distribution of the estimator of variance components is a linear combination of chi-square random variables, remember these estimators for the classic cases in slides 5, 21, 30).

Exercise 4

In *library nlme* and *lme4* of R is available the data set *Machines* (Pinheiro e Bates, 2000). The objective of the experiment is to compare three brands of machines used in an industrial process. Six workers were chosen randomly among the employees of a factory to operate each machine three times. The response variable is an overall productuvity score taking into account the number and quality of components produced.

a) Describe the appropriate model for this study. Fit the model using R, function *lmer* of package *lme4*. Use the commands plot.design (Machines) and interaction.plot (Machine,Worker,score) and comment.

b) What are the restrict maximum likelihood estimates for variance components of the model?

c) Would the values of the variance components estimates obtained by the maximum likelihood method be higher or lower than the estimates given in the previous item ?

Exercise 4 (cont.)

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

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

f) What are the values of the fixed effects estimates of the model? Explain the meaning of those estimates.

g) Carry out an appropriate hypothesis test to assess if there are any major effects associated with machine brands. Use a significance level of 0.01.

Exercise 4 (cont.)

Note: use the commands and comment the results.

```
plot(machines1r)
residuos<-resid(machines1r)
qqnorm(residuos)
eblupsworker<-ranef(machines1r)$Worker
qqnorm(eblupsworker[,1])
eblupsinteraccao<-ranef(machines1r)$`Worker:Machine`
qqnorm(eblupsinteraccao[,1])
```