

Modelos Matemáticos e Aplicações

Módulo 2: Modelos lineares mistos

Linear Mixed Models

Some particular cases and respective application - II

2017-2018

Elsa Gonçalves
ISA/UL

Case 3

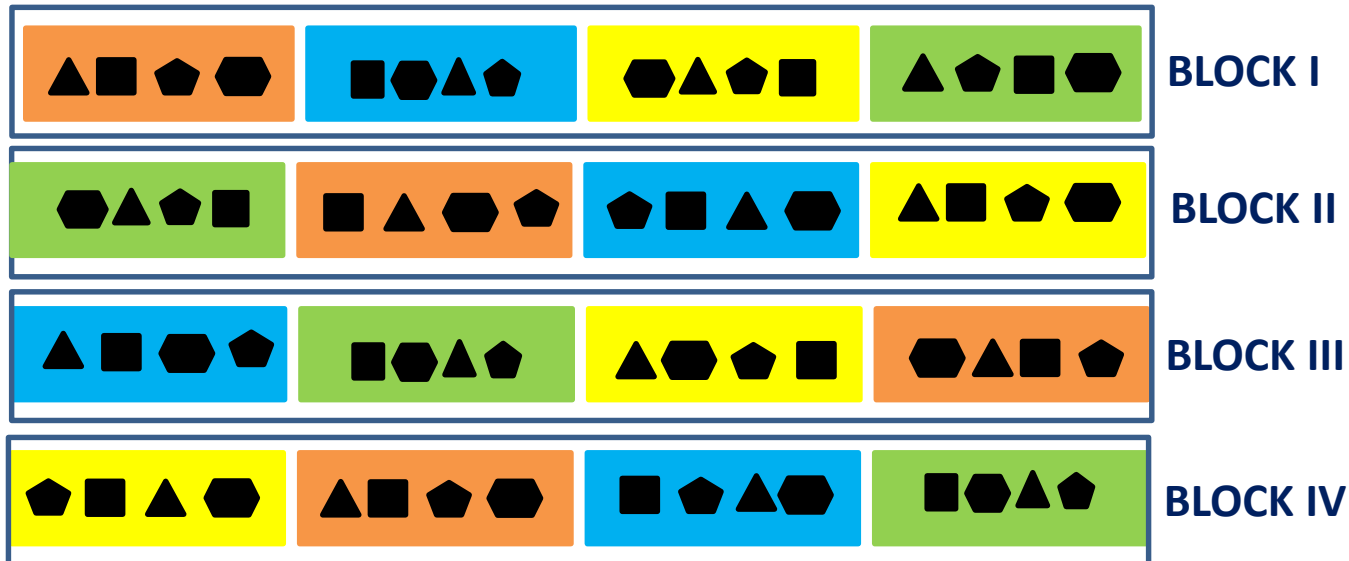
Linear mixed models for analysis of split plot experiments

The split-plot design on a RCB

- Main treatments (levels of factor A) are assigned at random within blocks, each treatment once per block; they are divided further into additional independent units (subplots) to which another set of treatments (levels of factor B) are randomly assigned.
- The number of blocks is the number of replications.
- Any main treatment can be adjacent to any other treatment, but not to the same treatment within the block.

Example:

Different colors represent different main treatments (levels of factor A) ; each row represents a block. There are 4 blocks (I-IV) each of 4 main treatments (colors) divided into 4 additional independent units (subplots) to which another set of treatments (levels of factor B, symbols) are randomly assigned.



Considering two factors with fixed effects (factors A and B) and random blocks. The model can be described as:

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

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

Where:

Y_{ijk} , is the observation from i^{th} level of factor A (*whole-plot i*), block j , and k^{th} level of factor B (*sub-plot or split-plot k*);

μ_{11} , is the general mean (population) in the level 1 of factor A with level 1 of factor B;

α_i , is the effect of the level i of the factor A (increase), assigned to whole-plot **(fixed)**;

u_j , is the effect of block j **(random)**;

$(\alpha u)_{ij}$, is the interaction effect of the i^{th} level of factor A with block j , named as *whole-plot error* **(random)**;

β_k , is the effect of the level k of the factor B (increase), assigned to *sub-plot* **(fixed)**;

$(\alpha\beta)_{ik}$, is the interaction effect of the i^{th} level of factor A with the k^{th} level of factor B (increase) **(fixed)**;

$(\beta u)_{kj}$, is the interaction effect of the k^{th} level of factor B with block j **(random)**;

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

In the common approach, the effect $(\beta u)_{jk}$ is set to zero (thus, $(\beta u)_{jk}$ is incorporated in e_{ijk}). The random error includes $(\beta u)_{jk}$ and $(\alpha\beta u)_{ijk}$, and is called as *within plot error*.

Therefore, the common assumptions are:

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

$$e_{ijk}, i. i. d., \mathcal{N}\left(0, \sigma^2_e\right), \forall ijk; Cov(u_j, (\alpha u)_{ij}) = 0; Cov(u_j, e_{ijk}) = 0;$$

$$Cov\left((\alpha u)_{ij}, e_{ijk}\right) = 0.$$

ANOVA TABLE, considering a balance design:

| | G.L. | S.Q. | QM | E[QM] | F |
|--|-------------------|------------|------------|--|-----------------------|
| Factor A | $a - 1$ | SQA | QMA | $c\sigma_{\alpha u}^2 + \sigma_e^2 + bc \frac{\sum_{i=1}^a (\alpha_i - \bar{\alpha})^2}{a - 1}$ | $\frac{QMA}{QMError}$ |
| Block | $b - 1$ | $SQBL$ | $QMBL$ | $ac\sigma_u^2 c\sigma_{\alpha u}^2 + \sigma_e^2$ | |
| Interaction FactorA×Block (Whole-plot error) | $(a - 1)(b - 1)$ | $SQWError$ | $QMWError$ | $c\sigma_{\alpha u}^2 + \sigma_e^2$ | |
| Factor B | $c - 1$ | SQB | QMB | $\sigma_e^2 + ab \frac{\sum_{k=1}^c (\beta_k - \bar{\beta})^2}{c - 1}$ | $\frac{QMB}{QMRE}$ |
| Interaction FactorA×FactorB | $(a - 1)(c - 1)$ | $SQAB$ | $QMAB$ | $\sigma_e^2 + b \frac{\sum_{i=1}^a \sum_{k=1}^c (\alpha\beta_{ik} - \bar{\alpha}\bar{\beta})^2}{(a - 1)(c - 1)}$ | $\frac{QMAB}{QMRE}$ |
| Residuals (Within plot error) | $a(b - 1)(c - 1)$ | $SQRE$ | $QMRE$ | σ_e^2 | |

Exercise 6

In *package* “nlme” do R, there is a data set named as “Alfalfa”.

```
> head(Alfalfa)
```

Grouped Data: Yield ~ Date | Block/Variety

| | Variety | Date | Block | Yield |
|---|---------|------|-------|-------|
| 1 | Ladak | None | 1 | 2.17 |
| 2 | Ladak | S1 | 1 | 1.58 |
| 3 | Ladak | S20 | 1 | 2.29 |
| 4 | Ladak | O7 | 1 | 2.23 |
| 5 | Ladak | None | 2 | 1.88 |
| 6 | Ladak | S1 | 2 | 1.26 |
| | | | | ... |

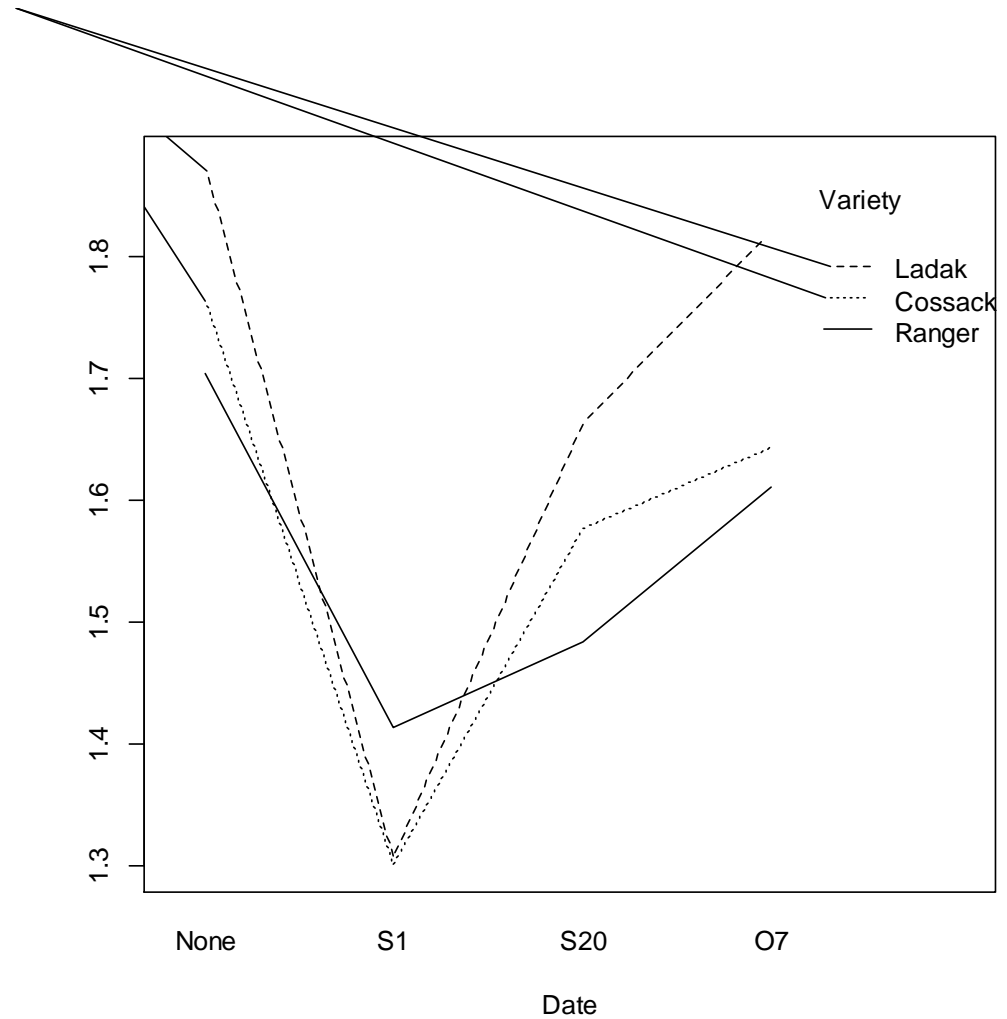
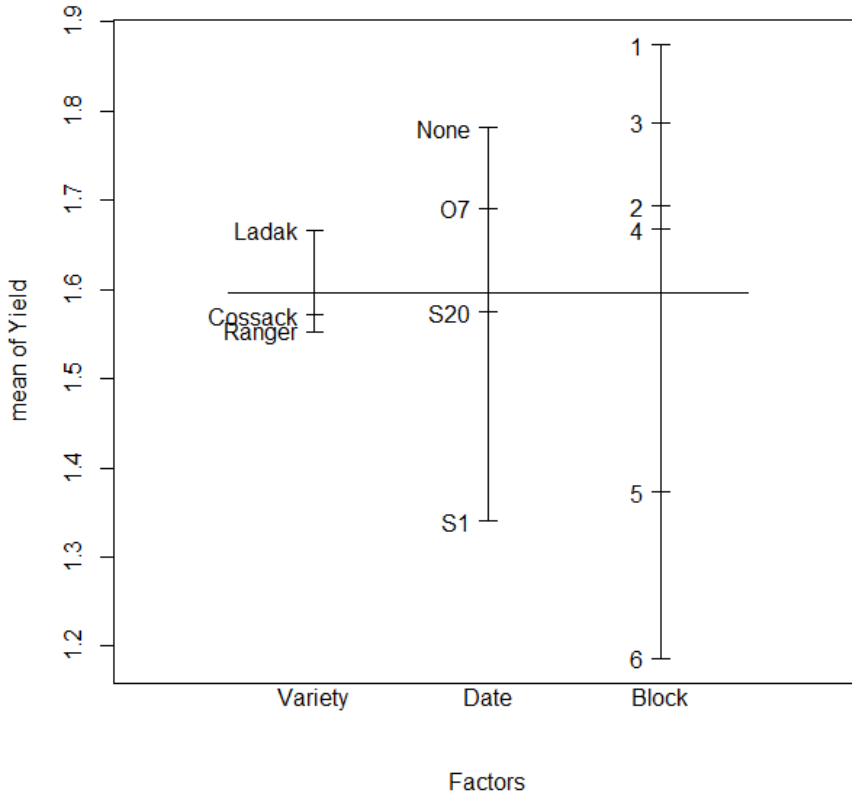
Exercise 6

These data are described in Snedecor & Cochran (1980) as an example of a *split-plot design* (Pinheiro and Bates, 2000). The objective is to study if the yield (T/acre) of alfalfa (*Medicago sativa*) is affected by variety and date of third cutting. Therefore, there are two factors: variety of alfalfa, with 3 levels (*Cossac*, *Ladak* e *Ranger*) and date of third cutting, with 4 levels (*none*–sem corte, *S1*–*Sep1*; *S20* – *Sep20*; and *O7* – *Oct7*). The treatment structure used in the experiment was a 3×4 full factorial. The experimental units were arranged into 6 blocks, each block was divided into 3 plots (*whole plots*; *whole plot*, *largest experimental unit*), where the varieties of alfalfa were randomly assigned; and each whole plot was divided into four subplots (split plots), where the dates of third cutting were randomly assigned.

a) Describe the appropriate model for this study.

Exercise 6 (cont.)

b) Plot the data using *plot.design* (*Alfalfa*) and *interaction.plot* (*Date*, *Variety*, *Yield*). Comment.



Exercise 6 (cont.)

- c) Fit the model described in item a) in R using *lmer* of package “*lme4*”.
- d) Carry out the hypothesis tests that answer the objectives of the study.
- e) Compare the previous results with those obtained with the command
“*aov(Yield~Date*Variety+Error(Block*Variety), data=Alfalfa)*”.

Some considerations

- (1) Factors A and B with random effects;
- (2) Factor A with fixed effects and factor B with random effects.

(1) ANOVA table: factors A and B with random effects, balanced:

| | G.L. | QM | E[QM] | F |
|--|-------------------|------------|---|------------------------------------|
| Factor A | $a - 1$ | QMA | $bc\sigma_\alpha^2 + c\sigma_{\alpha u}^2 + b\sigma_{\alpha\beta}^2 + \sigma_e^2$ | $\frac{QMA+QMRE}{QMWError+QMAB}^*$ |
| Block | $b - 1$ | $QMBL$ | $ac\sigma_u^2 + c\sigma_{\alpha u}^2 + \sigma_e^2$ | |
| Interaction FactorA×Block (Whole-plot error) | $(a - 1)(b - 1)$ | $QMWError$ | $c\sigma_{\alpha u}^2 + \sigma_e^2$ | |
| Factor B | $c - 1$ | QMB | $ab\sigma_\beta^2 + b\sigma_{\alpha\beta}^2 + \sigma_e^2$ | $\frac{QMB}{QMAB}$ |
| Interaction FactorA×FactorB | $(a - 1)(c - 1)$ | $QMAB$ | $b\sigma_{\alpha\beta}^2 + \sigma_e^2$ | $\frac{QMAB}{QMRE}$ |
| Residuals | $a(b - 1)(c - 1)$ | $QMRE$ | σ_e^2 | |

*Approximate degrees of freedom. For example, *Satterthwaite* method:

$$v_1 = \frac{(QMA+QMRE)^2}{\frac{(QMA)^2}{a-1} + \frac{(QMRE)^2}{a(b-1)(c-1)}}, v_2 = \frac{(QMWError+QMAB)^2}{\frac{(QMWError)^2}{(a-1)(b-1)} + \frac{(QMAB)^2}{(a-1)(c-1)}}$$

(2) ANOVA table: factor A with fixed effects and factor B with random effects, balanced:

| | G.L. | QM | E[QM] | F |
|--|-------------------|------------|--|------------------------------------|
| Factor A | $a - 1$ | QMA | $c\sigma_{\alpha u}^2 + b \frac{a}{a-1} \sigma_{\alpha\beta}^2 + \sigma_e^2 + bc \frac{\sum_{i=1}^a (\alpha_i - \bar{\alpha})^2}{a-1}$ | $\frac{QMA+QMRE}{QMWError+QMAB}^*$ |
| Block | $b - 1$ | $QMBL$ | $ac\sigma_u^2 c\sigma_{\alpha u}^2 + \sigma_e^2$ | |
| Interaction FactorA×Block (Whole-plot error) | $(a - 1)(b - 1)$ | $QMWError$ | $c\sigma_{\alpha u}^2 + \sigma_e^2$ | |
| Factor B | $c - 1$ | QMB | $ab\sigma_{\beta}^2 + \sigma_e^2$ | $\frac{QMB}{QMRE}$ |
| Interaction FactorA×FactorB | $(a - 1)(c - 1)$ | $QMAB$ | $b \frac{a}{a-1} \sigma_{\alpha\beta}^2 + \sigma_e^2$ | $\frac{QMAB}{QMRE}$ |
| Residuals | $a(b - 1)(c - 1)$ | $QMRE$ | σ_e^2 | |

*Approximate degrees of freedom. For example, *Satterthwaite* method:

$$v_1 = \frac{(QMA+QMRE)^2}{\frac{(QMA)^2}{a-1} + \frac{(QMRE)^2}{a(b-1)(c-1)}}, v_2 = \frac{(QMWError+QMAB)^2}{\frac{(QMWError)^2}{(a-1)(b-1)} + \frac{(QMAB)^2}{(a-1)(c-1)}}$$

Case 4

The following is an example of the application of linear mixed models with categorical and numerical predictor variables (covariance analysis) and in which the observations are made in the same individual over time*

* The correlation matrices used for this type of analysis are used in time series and spatial statistics. For its understanding would be necessary theoretical bases on time series and spatial statistics, which is not part of this UC. Therefore, we will only exemplify its application, so that it is recorded that these instruments are currently widely used in mixed models context.

Exercise 8

Data set *BodyWeight* (Pinheiro e Bates, 2000) is available in R, and is related to the body weights of rats measured over 64 days. The body weights of the rats (in grams) are measured on day 1 and every seven days thereafter until day 24, with an extra measurement on day 44. There are 3 groups of rats, each on a different diet.

```
> head(BodyWeight)
```

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

```
  weight Time Rat Diet
```

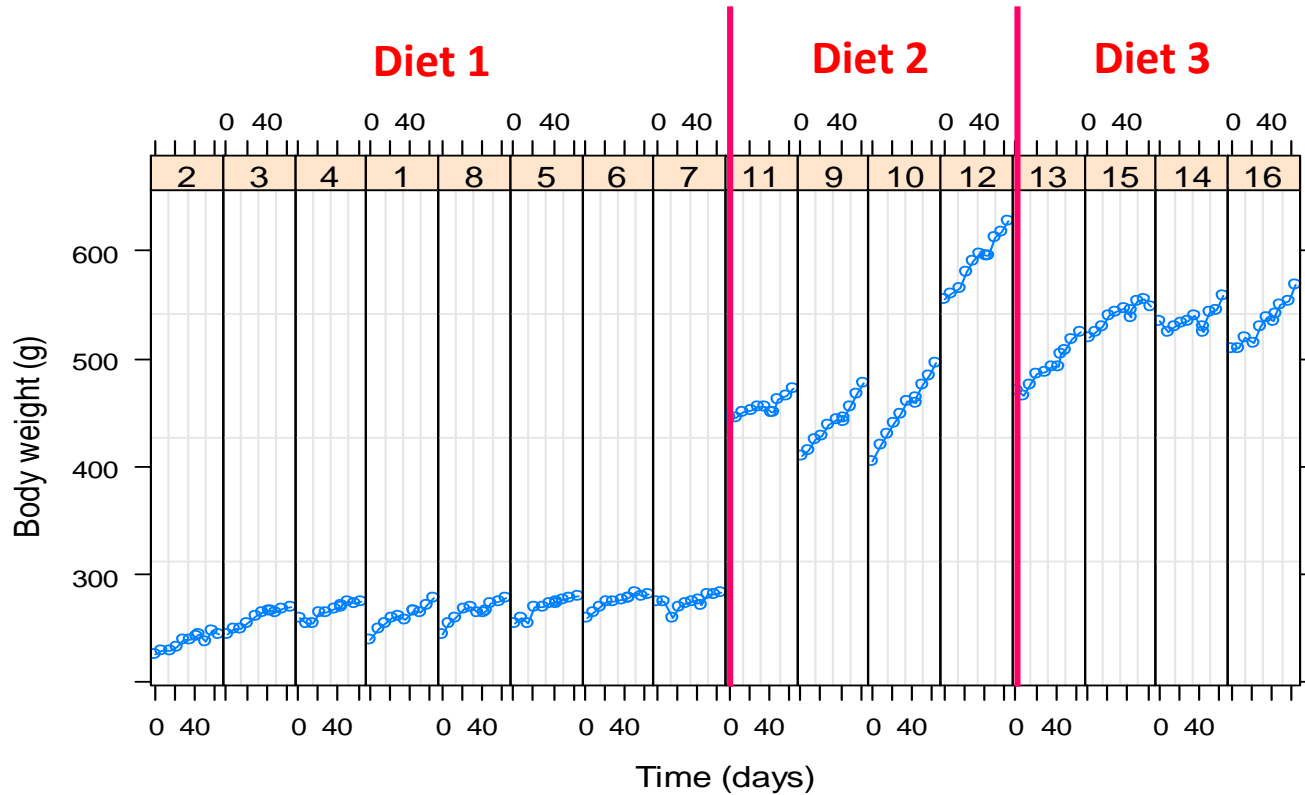
```
1   240   1  1  1
```

```
2   250   8  1  1
```

```
3   255  15  1  1
```

```
4   260  22  1  1
```

a) Plot the data using `plot(BodyWeight)` and comment.



It can be observed:

- differences among the three diet groups;
- 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) In R use *lme* of package “*nlme4*” to fit the appropriate model for this study (consider intercept and slope random effects to account for rat-to-rat variation). Use the commands *summary*, *anova*, *ranef* and *fitted*. Explain how each fitted value is obtained.

c) The observations are made in the same individual over time. In this context it can be model the dependence among the within-group errors. The observations are not equally spaced in time, as an extra observation is taken at 44 days. In this case, we can use a spatial correlation structure for random errors. Several correlation structures are available in package *nlme*, for example, *corExp*, *corGaus*, *corSpher*. Use the commands:

```
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)).
```

According to AIC and BIC criteria, what is the best correlation structure?

d) The model selected in item c) is significantly better than the model fitted in item b)?