

Applied Statistics with R Mixed models

Final Exam - solutions

Date: **9 March 2018** Student: _____ PhD Course: _____

Email address: _____

1. Considering the dataset “needle_litter_fall_rbd.txt” (Exercise 1), just 2 contrasts are allowed for comparing the levels of “treatment” effect, because the treatments are 3. Looking at the Lsmeans of the treatment effect and at the contrasts you have already run in the Exercise (**lmer** function), what 2 contrasts do you think to choose to test what is/are the significantly different level(s) of the “treatment” effect (because different from the other levels? N.B.: more than one answer can be ok)

- ☒ treatment 1 vs. treatment 2 (firstly we check the difference between treatment 1 and 2; then, since they are not different, we can compare each one of them with treatment 3. Don't matter if you don't have correctly answered, it was not so easy!)
- ☒ treatment 1 vs. treatment 3 => this answer and the following one are both ok
- ☒ treatment 2 vs. treatment 3 => this answer and the previous one are both ok
- ☐ treatment 1+2 vs. treatment 3

2. Try to compare the results of the mixed model with a linear model (**lm** function) including the block as a fixed effect: looking at the ANOVA, do you find a similar significance for the treatment effect in the linear and mixed model or not? ☒ YES / ☐ NO Do the Lsmeans differ? ☐ YES / ☒ NO
 In both linear and mixed models, treatment is highly significant (F: 96.353; Pr(>F): 2.753e-05 ***)
 Lsmeans are: C: 7.5; I: 7.4; IF: 9.1 (only SE differ, but that's expected)

```
> model.2<-lm(needle_litter_fall ~ treatment + block ,data = rbd_litter)
> Anova(model.2, type=3,test.statistic = "F") # ANOVA table SS type III
Anova Table (Type III tests)
```

```
Response: needle_litter_fall
              Sum Sq Df F value    Pr(>F)
(Intercept) 153.709  1 4068.765 9.982e-10 ***
treatment    7.280  2   96.353 2.753e-05 ***
block        1.513  3   13.353 0.004593 **
Residuals    0.227  6
```

```
> #comparison with model 1:
> Anova(model.1, type=3,test.statistic = "F") # ANOVA table SS type III
Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger df)
```

```
Response: needle_litter_fall
              F Df Df.res    Pr(>F)
(Intercept) 1713.310  1  3.922 2.503e-06 ***
treatment    96.353  2  6.000 2.753e-05 ***
```

3. The dataset “yield_rbd” has been already seen at the lesson, last 22 February. The experimental design is similar to that one of Exercise 1, obviously, data differ. Focusing on the Lsmeans of “Fertilizer” effect, could you compare the different levels (Fertilizer 1, Fertilizer 2 and Fertilizer 3) using the contrasts to determine if the three fertilizers differ each other (i.e., Fertilizer 1 from Fertilizer 2 etc.)? What levels are different? Lsmeans for fertilizer effect are: 9.5 (Fertilizer 1), 10.6 (Fertilizer 2), 12.3 (Fertilizer 3). Looking at the contrasts, all the levels of Fertilizer are significantly different among each other:

```
> library(multcomp)
> K <- matrix(c(0, 1, -1), 1) # Contrast Fertilizer1 vs Fertilizer2
> t <- glht(model.3, linfct = K) # fit a general linear hypothesis test
> summary(t) # See the results
```

Simultaneous Tests for General Linear Hypotheses

```
Fit: lme4::lmer(formula = yield ~ fertilizer + (1 | block), data = rbd, REML = TRUE)
```

Linear Hypotheses:

	Estimate	Std. Error	z value	Pr(> z)
1 == 0	-1.1667	0.3801	-3.07	0.00214 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```
> K <- matrix(c(0, 1, 0), 1) # Contrast Fertilizer1 vs Fertilizer3
> t <- glht(model.3, linfct = K) # fit a general linear hypothesis test
> summary(t) # See the results
```

Simultaneous Tests for General Linear Hypotheses

```
Fit: lme4::lmer(formula = yield ~ fertilizer + (1 | block), data = rbd, REML = TRUE)
```

Linear Hypotheses:

	Estimate	Std. Error	z value	Pr(> z)
1 == 0	-2.8333	0.3801	-7.455	8.99e-14 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```
> K <- matrix(c(0, 0, 1), 1) # Contrast Fertilizer2 vs Fertilizer3
> t <- glht(model.3, linfct = K) # fit a general linear hypothesis test
> summary(t) # See the results
```

Simultaneous Tests for General Linear Hypotheses

```
Fit: lme4::lmer(formula = yield ~ fertilizer + (1 | block), data = rbd, REML = TRUE)
```

Linear Hypotheses:

	Estimate	Std. Error	z value	Pr(> z)
1 == 0	-1.6667	0.3801	-4.385	1.16e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

4. Let's consider the dataset "nested_trees.txt" (see Figure 1). The dataset includes the measures of height growth for Norway spruce trees (*Picea alba*) in the course of 5 years. At the beginning of that period, part of the target forest has been divided into 2 zones with different treatments: one zone that has been then regularly irrigated for the following 5 years of experiment (Soil A), and the other zone that has been maintained natural, therefore not artificially irrigated (Soil B). Four plots of trees were defined in each zone, and 6 trees for each block were measured at the beginning of the experiment and after the 5 years. The height growth at the end of the 5 years was measured for each tree: data are reported in "nested_trees.txt". Before importing the data in R, hand write a Mixed Model to analyze the differences in Height_growth due to a different soil:

- What effect is nested? ☐ Soil ☒ Plot
- And within what other effect? ☒ Soil ☐ Plot
- Therefore, what effect do you include as random? Plot
- And what do you include as fixed effect? Soil
- Mixed Model: `model.4<-lmer(Height_growth ~ Soil + (1|Soil:Plot), data = nested_trees, REML = TRUE)`

5. Let's import the dataset "nested_trees.txt"; it includes the following columns: Soil; Block; Tree; Height_growth; remember to set the Plot as factor (it is not necessary for Soil because Soil is not numeric). Change the reference grid to SAS for Soil and Plot (see the previous exercises as an example). Write the mixed model in R and print the results from the model

```
# "Read the data" #
nested_trees<-read.table(file = ".../nested_trees.txt",sep = "\t",header =
TRUE,stringsAsFactors = TRUE)
options(contrasts = c("contr.SAS","contr.poly")) # Change contrast options
nested_trees$Plot<-as.factor(nested_trees$Plot) # Set the variable as factor
nested_trees$Soil<-as.factor(nested_trees$Soil) # Set the variable as factor

#### Random Model
model.4<-lmer(Height_growth ~ Soil + (1|Soil:Plot), data = nested_trees ) # Fit a
model with Random effect
summary(model.4) # See the results
```

6. Print the ANOVA for the model using the function ANOVA: is the fixed effect that you included in the model significant? ☒ YES / ☐ NO

```
Anova(model.4, type=3,test.statistic = "F") # ANOVA

> Anova(model.4, type=3,test.statistic = "F") # ANOVA
Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger df)

Response: Height_growth
          F Df Df.res    Pr(>F)
(Intercept) 478.428  1      6 5.965e-07 ***
Soil        22.417  1      6 0.003209 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

7. Let's check the least square means (Lsmeans) for the levels of the fixed effect and write some considerations: Lsmean of Soil B is higher (2.45) than Lsmean of Soil A (1.70), and since the Soil effect is significant (see the Anova), it is possible to conclude that a regular irrigation favors the growth of trees. In this case we don't need to use the contrasts to check the differences between levels because we have just 2 levels (Soil A and Soil B), therefore Anova is enough to see if levels are different

```
lsmeans(model.4,"Soil") # lsmeans

> lsmeans(model.4,"Soil") # lsmeans
Soil lsmean      SE df lower.CL upper.CL
A      1.70 0.1120103  6  1.425921  1.974079
B      2.45 0.1120103  6  2.175921  2.724079
```

Optional questions:

8. Let's consider Exercise 3; in addition to consider the Variance-Covariance Structure NULL (model.a) and corAR1 (model.b) to account for the repeated time measures, an additional Variance-Covariance Structure, that is the compound symmetry, **corCompSymm** (model.c). Could you compare the model fitting using the 3 Variance-Covariance Structures? What is the best fitted model?

- ☐ model.a / ☒ model.b / ☐ model.c (model.b is the best fitted due to the lowest AIC and BIC)

```
model.c<-lme(height ~ alluminium + time + alluminium*time, data = mydata, # Model
             random = ~ 1|seedling,                                # Random Effect
             weights = varIdent(form = ~ 1|time),                  # Constant variance(s)
             correlation = corCompSymm())                          # compound symmetry
```

```
anova(model.a,model.b,model.c)
```

```
> anova(model.a,model.b,model.c)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
model.a	1	21	2479.600	2553.718	-1218.800			
model.b	2	22	2392.086	2469.733	-1174.043	1 vs 2	89.51447	<.0001
model.c	3	22	2481.117	2558.764	-1218.558			

9. Print the **anova** for the best fitted model (but you could find the same conclusions for all the models) and answer the following statements:

- The maple seedling growth significantly varies over time ☒ TRUE / ☐ FALSE
- The treatment effect is significant when considered: ☐ alone / ☒ in its interaction with time

(the treatment is "alluminium")

```
anova(model.a,model.b,model.c)
```

```
> anova(model.b)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	189	270.38679	<.0001
alluminium	3	63	1.02365	0.3882
time	3	189	13.85646	<.0001
alluminium:time	9	189	2.14919	0.0274

Figure 1: Dataset "nested_trees.txt"

