

Example Dataset

Pablo Hernández Alonso

05-06-2019

Contents

0. Report at a glance	2
1. Selection of data	3
1.1 Descriptives of data	3
1.1.1 Uploaded data	3
1.1.2 Data analysed	3
1.2 Transformation of data	3
1.3 Exploratory Data Analysis (EDA)	4
1.3.1 Variables	4
Data Frame Summary	4
1.3.2 Missing values	5
1.3.3 Histogram of Continuous Variables	6
1.3.4 Correlation plots	7
1.3.5 Barplots of categorical data	8
1.3.6 Boxplots	9
2. Linear Mixed Model	10
2.1 Step 0. Base models	10
2.2 Step 1. Definition of the ‘Beyond Optimal Model’	11
2.3 Step 2. Structure of RANDOM effects	12
2.4 Step 3. Structure of FIXED effects	13
2.5 Step 4. Final model	14
2.5.1 SUMMARY OF THE DATA	14
2.5.1.1 RANDOM EFFECTS	14
2.5.1.2 FIXED EFFECTS - ANOVA BASED ON T-VALUE	14
2.5.2 OVERALL OUTPUT FROM THE MODEL	14
2.5.2.1 FULL SUMMARY	14
2.5.2.2 ANOVA BASED ON F-TEST	15
2.5.2.3 R-SQUARED (based on Nakagawa and Schielzeth (2013))	15
2.5.2.4 95% CONFIDENCE INTERVALS FOR THE COEFFICIENTS	15
3. Assumptions of LMM	16
3.1 Normality, linearity and Homoscedasticity	16
4. Getting information from the model	17
4.1 A) Plot of random effects	17
4.2 B) Predicted values for the data	17

0. Report at a glance

This report consists of the analysis (tests and graphs) of repeated measures data using linear-mixed model (LMM) implemented in SISREM. It is divided into the different sections that should be taken into account in the context of performing a repeated measures analysis.

We will be really glad about receiving any suggestions for improving our Report. Please, contact sisrem@gmail.com or pablo1280@gmail.com

1. Selection of data

This section describes the variables included in the data for analysis.

1.1 Descriptives of data

1.1.1 Uploaded data

This is the initial dataset that uploaded or loaded into the program:

```
## 'data.frame':   428 obs. of  4 variables:
## $ TAG   : int  86 86 101 101 155 155 155 160 160 192 ...
## $ DBH   : int  454 454 369 369 450 450 450 413 413 486 ...
## $ FRUIT: int  45 662 6 7 700 225 96 67 43 61 ...
## $ YEAR  : int  1 2 1 2 1 2 3 1 3 1 ...
```

1.1.2 Data analysed

This is the dataset that you have used for conducting the analysis

```
## 'data.frame':   428 obs. of  4 variables:
## $ TAG   : Factor w/ 179 levels "86","101","155",...: 1 1 2 2 3 3 3 4 4 5 ...
## $ DBH   : int  454 454 369 369 450 450 450 413 413 486 ...
## $ FRUIT: num  3.83 6.5 1.95 2.08 6.55 ...
## $ YEAR  : Factor w/ 3 levels "1","2","3": 1 2 1 2 1 2 3 1 3 1 ...
```

1.2 Transformation of data

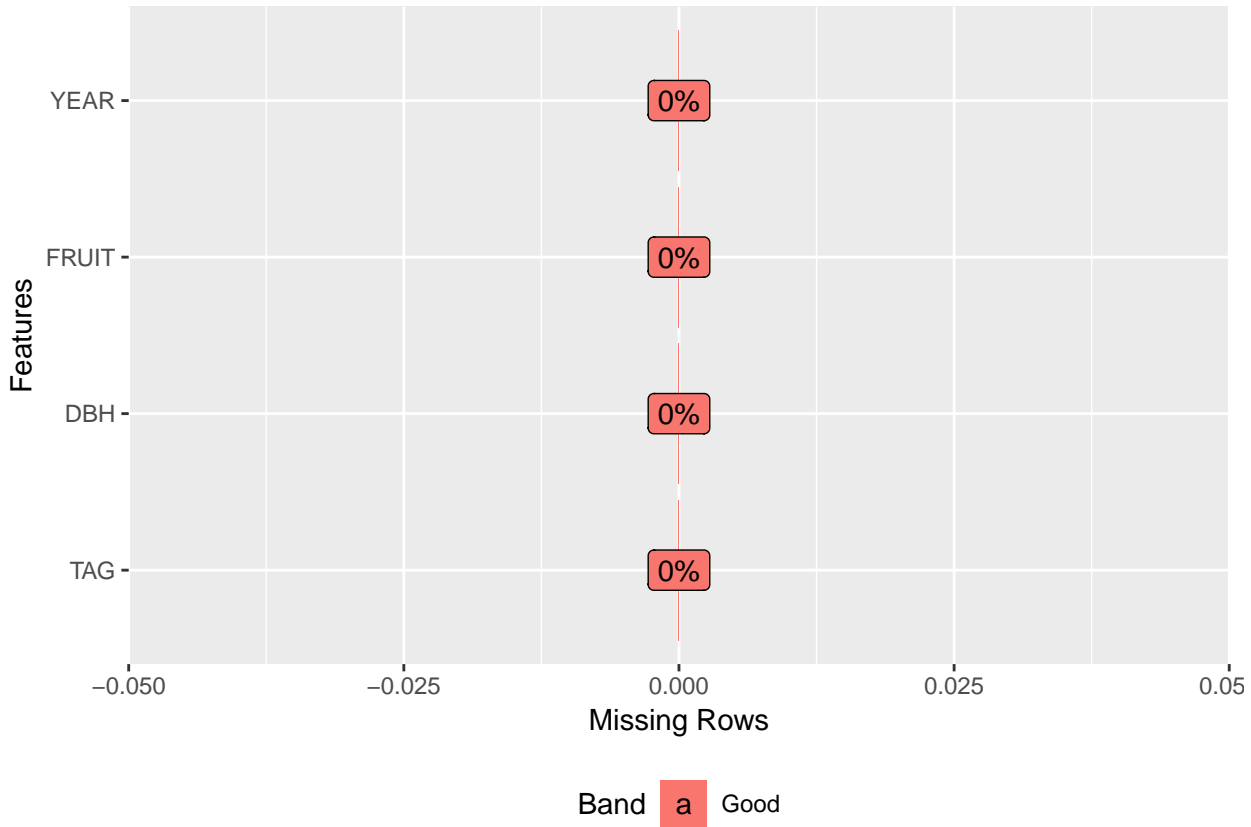
Data usually needs to be transformed before applying a statistical procedure. In this application, we allow the user to transform the “Y” variable after conducting a Box-Cox to determine the optimum transformation. Moreover, SISSREM also allows the user to transform the “X” predictors.

You have transformed your Y variable: FRUIT by applying Ln . You have not further transformed your variables.

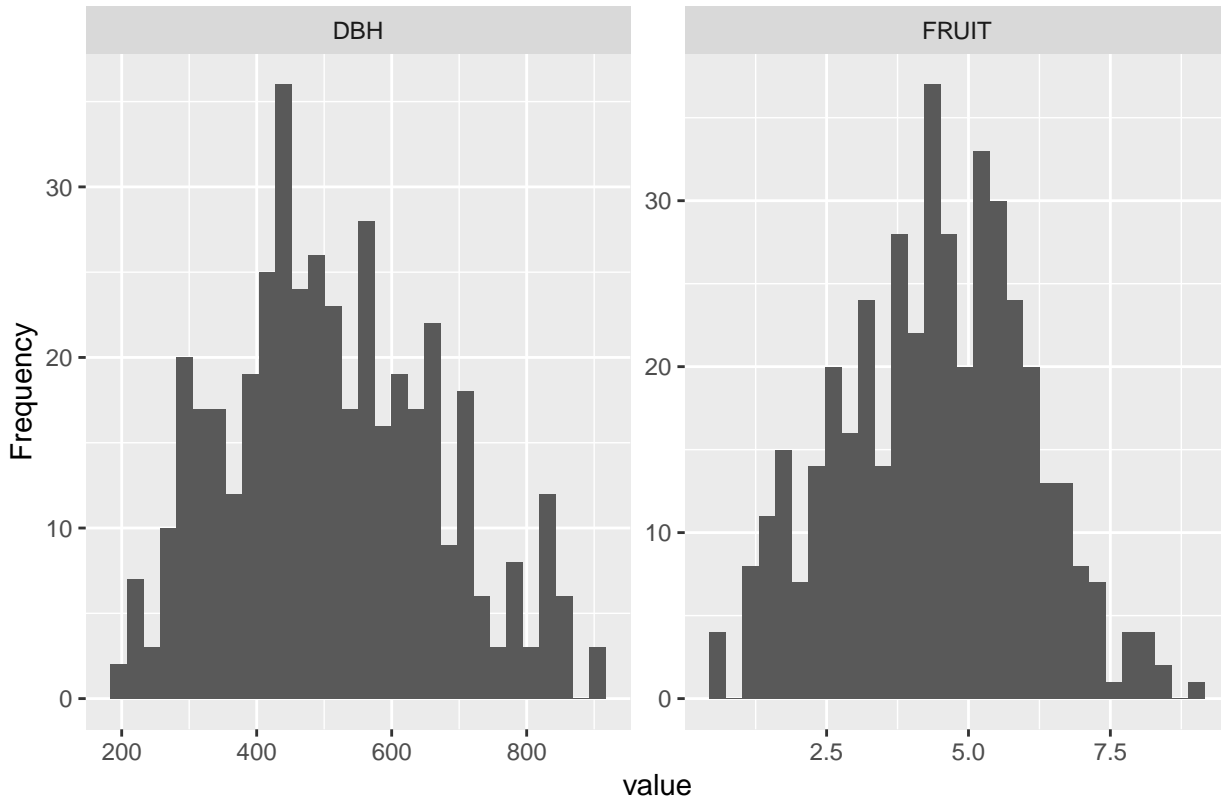
1.3.2 Missing values

Missing values are of important concern in data. LMM can deal with missing values, which is a huge advantage compared to other alternatives such as repeated measures ANOVA.

Here is a summary of the number of missing values in your data

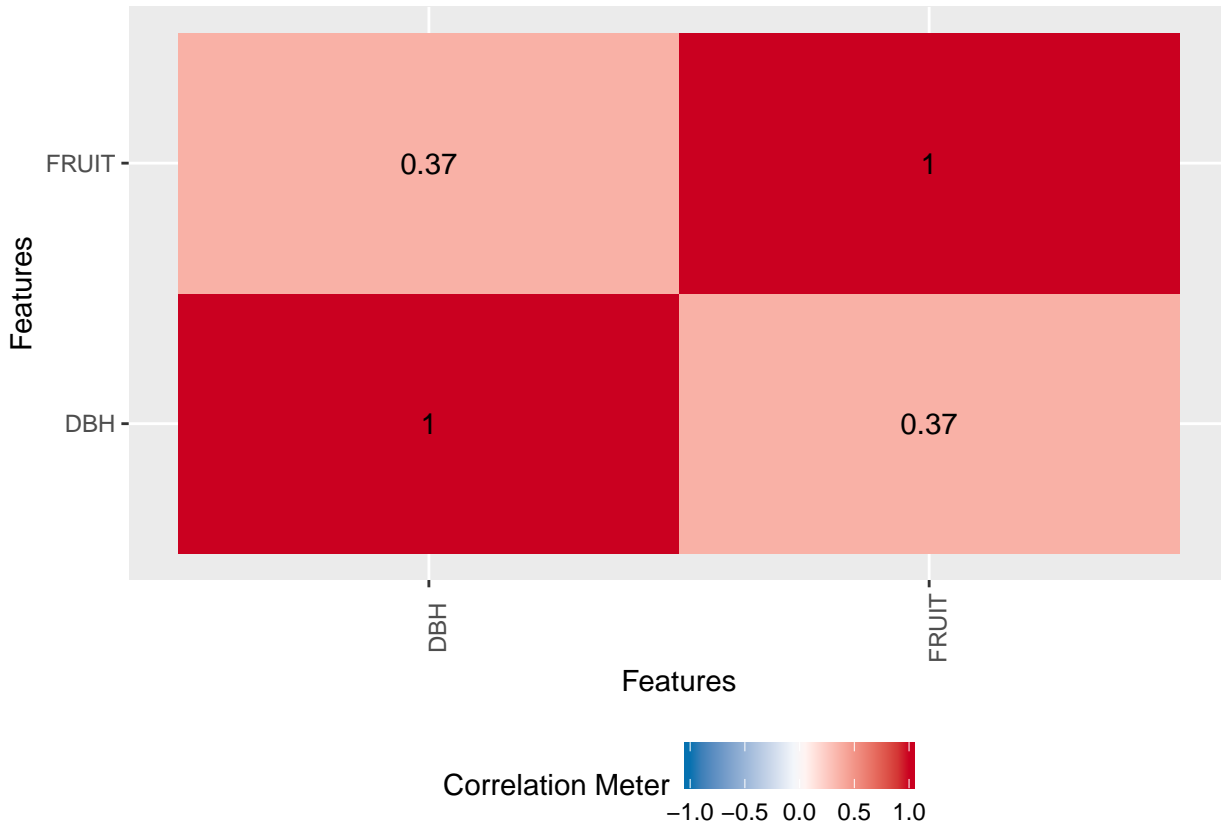


1.3.3 Histogram of Continuous Variables



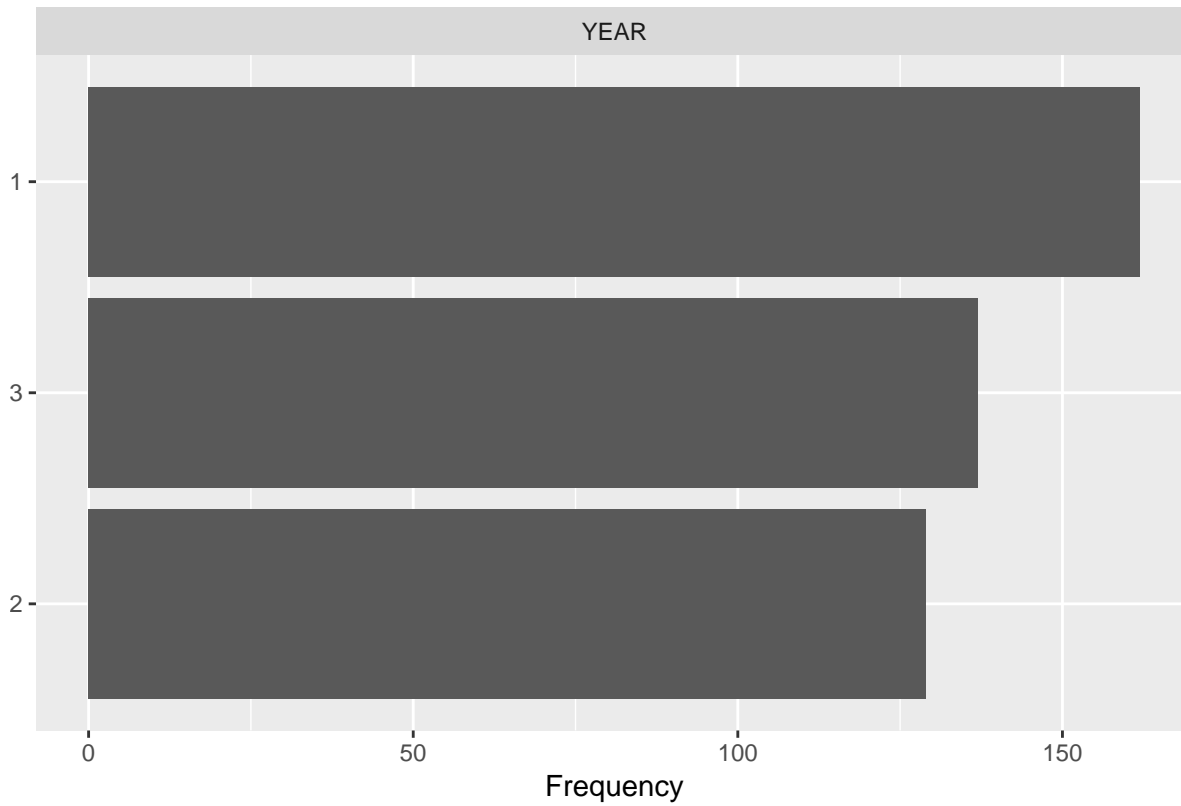
1.3.4 Correlation plots

Correlation between variables may generate several problems in the final LMM. Therefore, it is important to assess the correlation between the continuous variables before including them into the model.



1.3.5 Barplots of categorical data

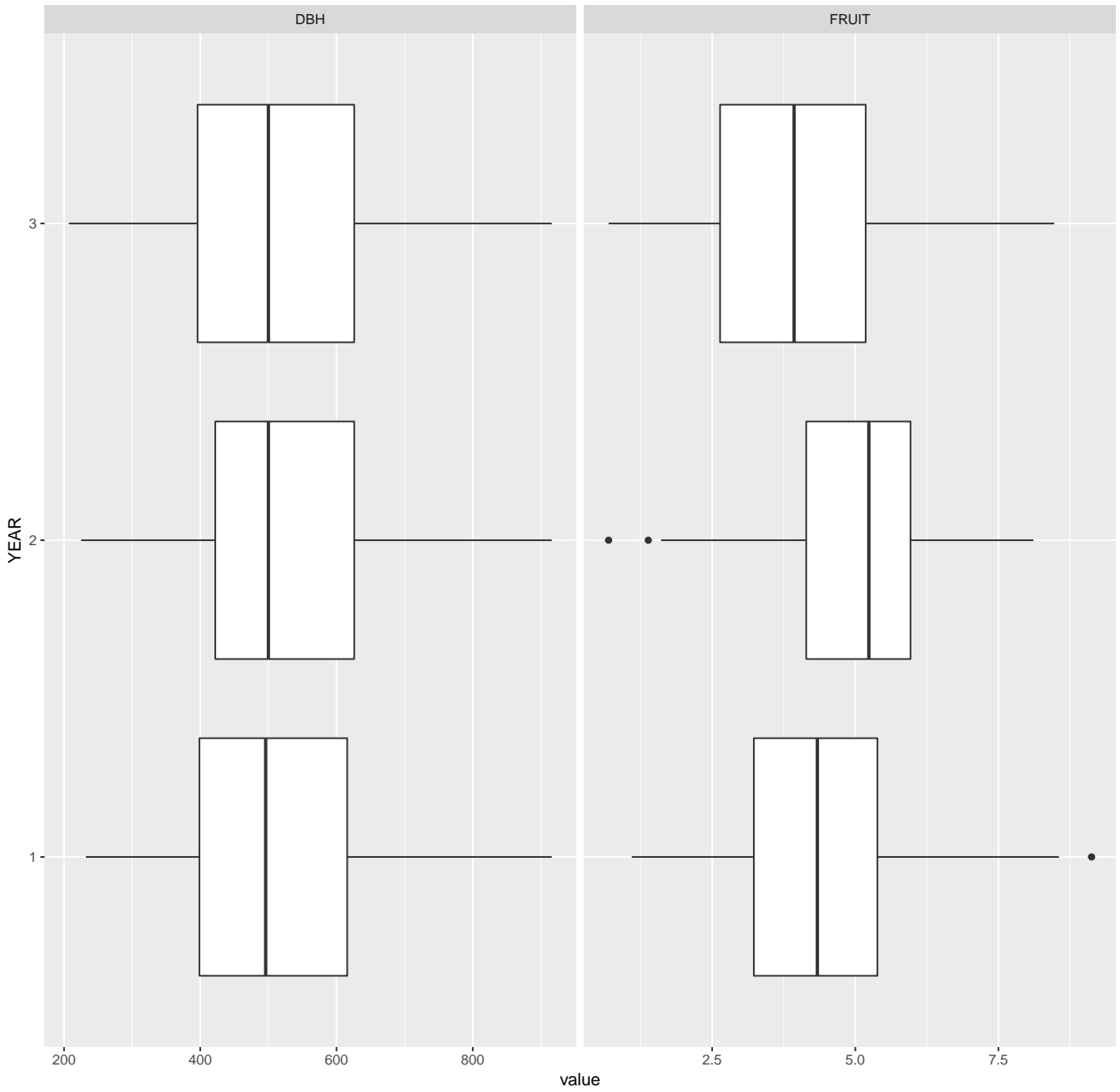
```
## There are no categorical variables in your data
## 1 columns ignored with more than 50 categories.
## TAG: 179 categories
```



1.3.6 Boxplots

Boxplots by the time variable (used as factor)

```
plot_boxplot(data_to_use(), by = input$var_fixed_for_boxplot)
```



2. Linear Mixed Model

In this section, results regarding the LMM that you have performed will be shown.

2.1 Step 0. Base models

Let's check the results for the Step 0.

In this initial step, every predictor (fixed variables (FV)) will be included in an independent model together with the variable of TIME and the random variable (RV) assuming a random intercept model. The user should select those FVs with a P-value < 0.2 to be included in the next step.

These are the models evaluated:

```
FRUIT ~ DBH + YEAR , random= list(~1|TAG) .
```

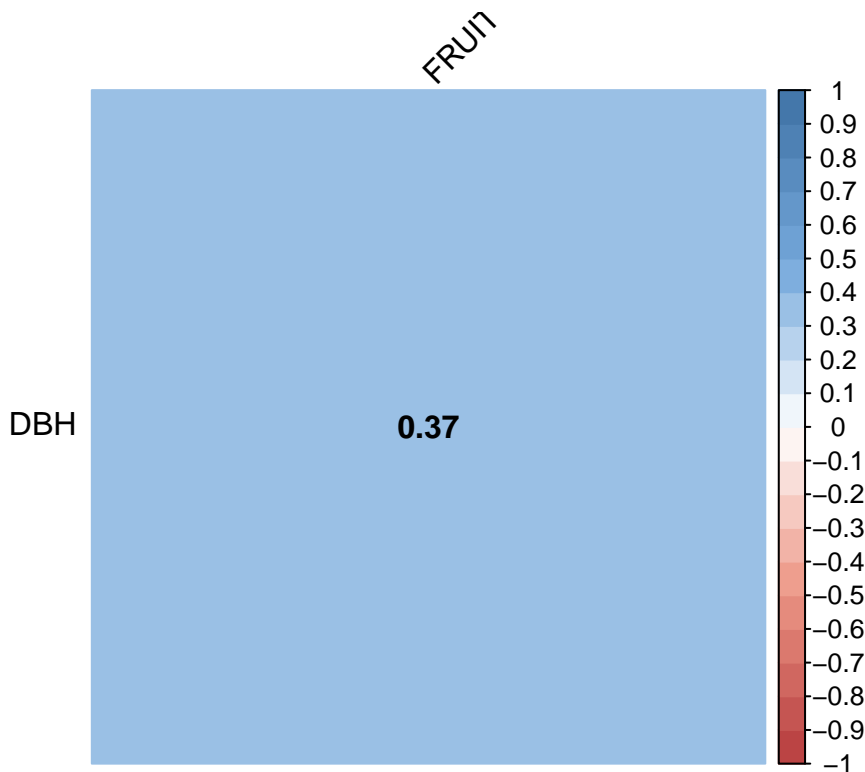
```
## These are the results for each FV and TIME:
```

```
##
```

```
## [[1]]
```

```
##          numDF denDF  F-value p-value
## (Intercept)      1   247 2225.7516 <.0001
## DBH              1   177  51.4726 <.0001
## YEAR            2   247  24.8278 <.0001
```

Importantly, remember not to add variables that are highly correlated ($r > 0.70$ based on Dormann et al., 2013) because when you include correlated variables into a parametric model it will have a problem of coefficients' stability as it will lose accuracy in determining the coefficients. In case that you have correlation between variables, remember to choose one of the correlated variables using biological knowledge/reasoning to select the most meaningful variable or conduct a dimension-reduction analysis (e.g. Principal Components Analysis) leaving a single variable that accounts for most of the shared variance among the correlated variables



No paired correlations with $r > |0.7|$

2.2 Step 1. Definition of the ‘Beyond Optimal Model’

Let’s check the results for the Step 1.

In this step you have to choose all the fixed terms that you want to consider and interactions, together with the main random term (different subjects where the measures are repeated: 1|ID). Remember not to add variables that are highly correlated ($r > 0.70$)!

Beyond Optimal Model

Equation is:

```
lme(FRUIT ~ YEAR * DBH, random=list(~1|TAG))
```

Model fit by: ML

Missing values: na.action = na.omit

Results from the model are:

```
## Linear mixed-effects model fit by maximum likelihood
## Data: data_to_use()
##      AIC      BIC    logLik
## 1526.595 1559.068 -755.2975
##
## Random effects:
## Formula: ~1 | TAG
##      (Intercept) Residual
## StdDev:  0.9052937 1.181427
##
## Fixed effects: FRUIT ~ YEAR * DBH
##      Value Std.Error DF   t-value p-value
## (Intercept)  2.1082244 0.4038442 245   5.220390 0.0000
## YEAR2        0.4990298 0.5093989 245   0.979644 0.3282
## YEAR3       -0.0663867 0.4847739 245  -0.136944 0.8912
## DBH          0.0042976 0.0007592 177   5.660764 0.0000
## YEAR2:DBH    0.0000575 0.0009387 245   0.061274 0.9512
## YEAR3:DBH   -0.0008926 0.0008964 245  -0.995760 0.3203
## Correlation:
##      (Intr) YEAR2  YEAR3  DBH   YEAR2:
## YEAR2      -0.535
## YEAR3      -0.589  0.457
## DBH        -0.957  0.509  0.559
## YEAR2:DBH  0.519 -0.959 -0.443 -0.537
## YEAR3:DBH  0.567 -0.441 -0.956 -0.583  0.466
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.77482861 -0.63634754  0.05714195  0.65099543  2.40173195
##
## Number of Observations: 428
## Number of Groups: 179
```

ANOVA of the FIXED effects:

```
##      numDF denDF  F-value p-value
## (Intercept)    1   245 2215.3717 <.0001
## YEAR           2   245  26.1948 <.0001
## DBH            1   177  48.5197 <.0001
## YEAR:DBH       2   245   0.6717 0.5118
```

2.3 Step 2. Structure of RANDOM effects

Let's check the results for the Step 2.

In this step we are going to find a proper structure for the RANDOM component of the LMM. We will use the best 'Beyond optimal model' and we will vary the RANDOM component leaving the same FIXED component. When model fits are ranked according to their AIC values, the model with the lowest AIC value being considered the 'best'. REML method is used to properly compare the models using AIC

Structure of RANDOM components

Equations are:

Model 0 (without random effects): `gls(FRUIT ~ YEAR + DBH)`

Model 1 (random intercept model): `lme(FRUIT ~ YEAR + DBH, random=list(~1|TAG))`

Model 2 (random slope (and intercept) model): `lme(FRUIT ~ YEAR + DBH, random=list(YEAR|TAG))`

Models fit by: REML

Missing values: `na.action = na.omit`

MODEL comparison based on AIC

##	df	AIC
## mod0	5	1582.571
## mod1	6	1544.358
## mod2	11	1548.011

2.4 Step 3. Structure of FIXED effects

Let's check the results for the Step 3.

In this step we are going to find a proper structure for the FIXED component of the LMM). We will use the best model from STEP2 and we will vary the FIXED component leaving the same optimized RANDOM component from STEP2. When model fits are ranked according to their AIC values, the model with the lowest AIC value being considered the 'best'. ML method is used to properly compare the models using AIC.

Structure of FIXED effects

Equations are:

Model 3: basic model `lme(FRUIT ~ 1 , random=list(~1|TAG))`

Model 4: main predictor model `lme(FRUIT ~ DBH, random=list(~1|TAG))`

Model 5: addition model `lme(FRUIT ~ YEAR + DBH, random=list(~1|TAG))` Model 6: interaction model `lme(FRUIT ~ YEAR * DBH, random=list(~1|TAG))`

Models fit by: ML

Missing values: `na.action = na.omit`

`## MODEL comparison based on AIC`

```
##      df      AIC
## mod3  3 1608.961
## mod4  4 1566.621
## mod5  6 1523.951
## mod6  8 1526.595
```

2.5 Step 4. Final model

Let's check the results for the Step 4.

In this step we are going to generate the FINAL model based on previous steps. REML method is used to properly report the FINAL model.

Equation is: `lme(FRUIT ~ YEAR + DBH, random=list(~1|TAG))`

Models fit by: REML

Missing values: `na.action = na.omit`

2.5.1 SUMMARY OF THE DATA

2.5.1.1 RANDOM EFFECTS

```
## TAG = pdLogChol(1)
##           Variance StdDev
## (Intercept) 0.8246943 0.9081268
## Residual    1.4161665 1.1900279
```

2.5.1.2 FIXED EFFECTS - ANOVA BASED ON T-VALUE

```
##           Value Std.Error DF t-value p-value
## (Intercept) 2.267676594 0.3110738472 247 7.289834 4.184186e-12
## YEAR2       0.532928157 0.1437375449 247 3.707648 2.583104e-04
## YEAR3      -0.528384953 0.1424065448 247 -3.710398 2.556675e-04
## DBH         0.003991142 0.0005718358 177 6.979525 5.723058e-11
```

2.5.2 OVERALL OUTPUT FROM THE MODEL

2.5.2.1 FULL SUMMARY

```
## Linear mixed-effects model fit by REML
## Data: data_to_use()
##      AIC      BIC    logLik
## 1544.358 1568.657 -766.1792
##
## Random effects:
## Formula: ~1 | TAG
##      (Intercept) Residual
## StdDev:   0.9081268 1.190028
##
## Fixed effects: FRUIT ~ YEAR + DBH
##           Value Std.Error DF t-value p-value
## (Intercept) 2.2676766 0.31107385 247 7.289834 0e+00
## YEAR2       0.5329282 0.14373754 247 3.707648 3e-04
## YEAR3      -0.5283850 0.14240654 247 -3.710398 3e-04
## DBH         0.0039911 0.00057184 177 6.979525 0e+00
## Correlation:
##      (Intr) YEAR2 YEAR3
## YEAR2 -0.180
## YEAR3 -0.214 0.446
## DBH   -0.927 -0.020 0.009
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
```

```
## -2.78647936 -0.60881246 0.03793967 0.65215390 2.38337605
##
## Number of Observations: 428
## Number of Groups: 179
```

2.5.2.2 ANOVA BASED ON F-TEST

```
##          numDF denDF  F-value p-value
## (Intercept)      1   247 2225.7514 <.0001
## YEAR            2   247  26.2072 <.0001
## DBH             1   177  48.7138 <.0001
```

2.5.2.3 R-SQUARED (based on Nakagawa and Schielzeth (2013))

R2m (Marginal R2 (whole model)): it is the marginal R2 for a linear mixed model, meaning that it is concerned with the variance explained by the fixed factors.

R2c (Conditional R2 (whole model)): it is the conditional R2 for a linear mixed model, meaning that it is concerned with the variance explained by the fixed and random factors.

```
## Warning: 'r.squaredGLMM' now calculates a revised statistic. See the help
## page.
```

```
##          R2m      R2c
## [1,] 0.2072428 0.4989978
```

2.5.2.4 95% CONFIDENCE INTERVALS FOR THE COEFFICIENTS

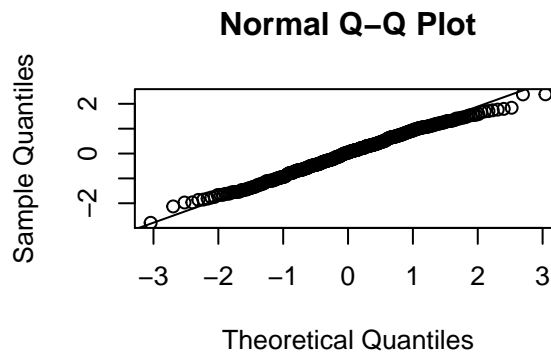
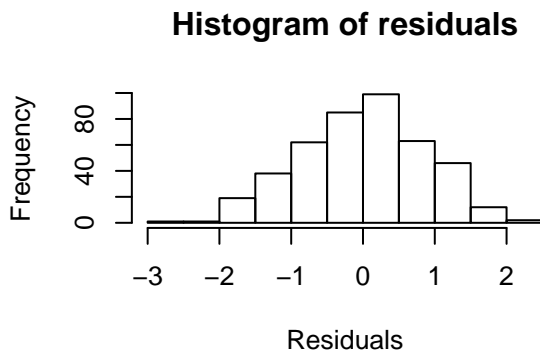
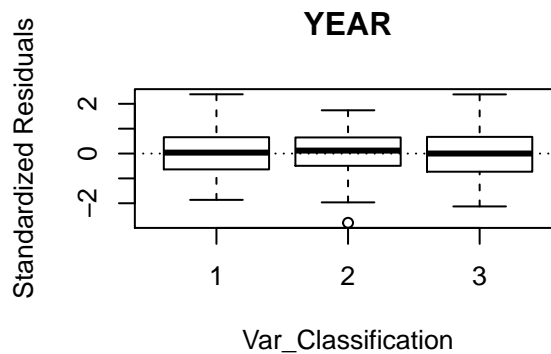
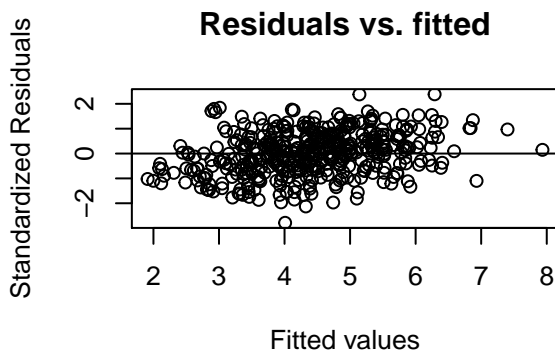
```
## Approximate 95% confidence intervals
##
## Fixed effects:
##          lower      est.      upper
## (Intercept) 1.654980955 2.267676594 2.880372233
## YEAR2       0.249820568 0.532928157 0.816035747
## YEAR3      -0.808870985 -0.528384953 -0.247898921
## DBH         0.002862649 0.003991142 0.005119636
## attr("label")
## [1] "Fixed effects:"
##
## Random Effects:
## Level: TAG
##          lower      est.      upper
## sd((Intercept)) 0.7441771 0.9081268 1.108196
##
## Within-group standard error:
##          lower      est.      upper
## 1.089232 1.190028 1.300152
```

3. Assumptions of LMM

3.1 Normality, linearity and Homoscedasticity

We need to check that the assumptions are met:

- 1) No pattern should be seen in the plots: ‘Fitted values vs Standardized residuals’ (top left plot) and ‘Explanatory variable vs Standardized residuals’ (top right plot).
- 1.a) If we observe increasing or decreasing values in ‘Fitted values vs Standardized residuals’, it inform us about Heteroscedasticity (i.e. variance of data not approx. equal across range of predicted values) that we should correct (try box-cox transformation alternatives).
- 2) Histogram of residuals (bottom left) must follow a Gaussian distribution, otherwise it will inform us about non-normality behaviour.
- 3) QQ-Plot (bottom right) must show dots aligned to the diagonal line, otherwise it will inform us about non-normality.



4) Shapiro-Wilk analysis of residuals **** If P-value < 0.05 -> Non-normality

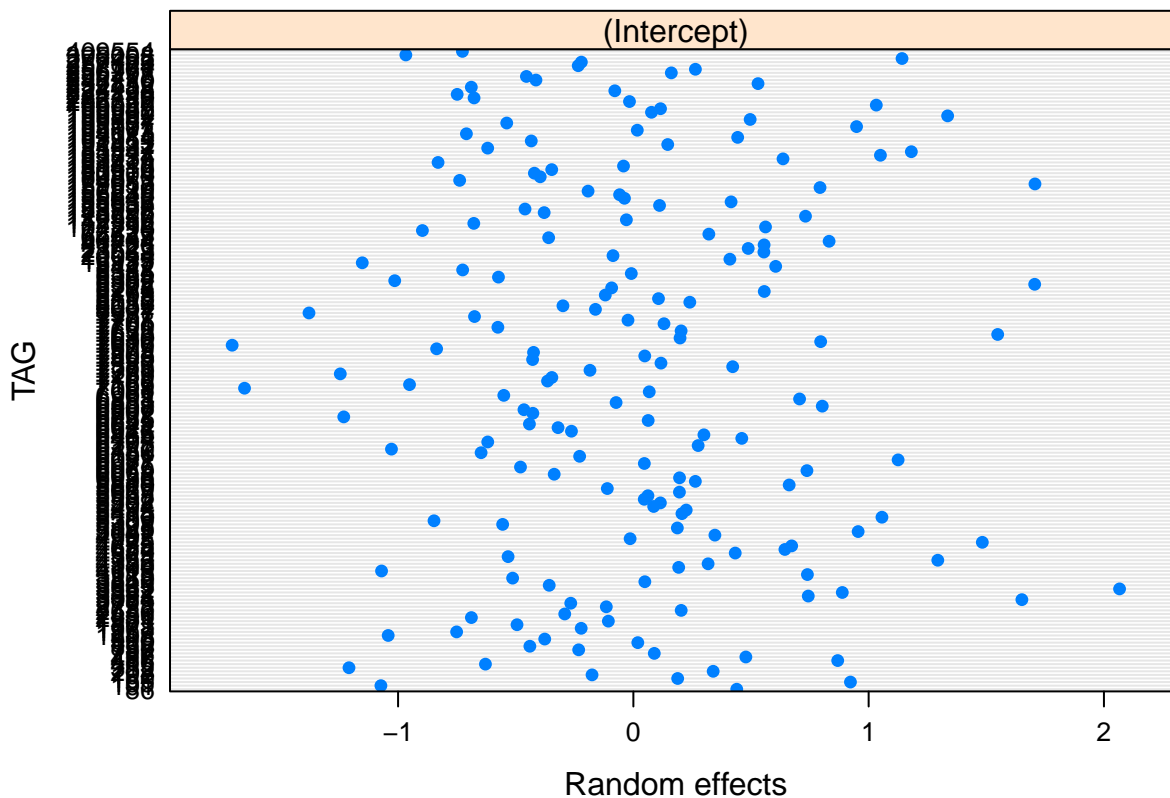
```
##  
## Shapiro-Wilk normality test  
##  
## data: Res  
## W = 0.99555, p-value = 0.2642
```

5) Multi-collinearity

```
## YEAR2 YEAR3 DBH  
## 1.249056 1.248677 1.000783
```

4. Getting information from the model

4.1 A) Plot of random effects



4.2 B) Predicted values for the data

B1) If the structure is 1|RV then you are constructing a RANDOM INTERCEPT MODEL:

Each random variable (subject) is assigned a different intercept cause by-RV (by subject) variability is taken into account. However, the fixed effects are all the same for all RVs.

In this model, we account for baseline-differences in the dependent variable, but we assume that whatever the effect of the fixed variables, it is the same for all the RVs.

B2) If the structure is FX|RV then you are constructing a RANDOM SLOPE MODEL:

In this case RVs are not only allowed to have differing intercepts, but they are also allowed to have different slopes for the effect of FVs.

##	(Intercept)	YEAR2	YEAR3	DBH
## 86	2.707033	0.5329282	-0.528385	0.003991142
## 101	1.194721	0.5329282	-0.528385	0.003991142
## 155	3.190287	0.5329282	-0.528385	0.003991142
## 160	2.456003	0.5329282	-0.528385	0.003991142
## 192	2.092015	0.5329282	-0.528385	0.003991142
## 203	2.606860	0.5329282	-0.528385	0.003991142
## 352	1.059062	0.5329282	-0.528385	0.003991142
## 415	1.639040	0.5329282	-0.528385	0.003991142
## 486	3.136129	0.5329282	-0.528385	0.003991142
## 487	2.745737	0.5329282	-0.528385	0.003991142

SISSREM: Shiny Interactive, Supervised and Systematic report from REpeated Measures data - 2019 - Beta Version 1.0.