THE ESSENTIALS ON LINEAR REGRESSION, ANOVA, GENERAL LINEAR AND LINEAR MIXED MODELS FOR THE CHEMIST

Govaerts, R., Francq, B., Marion, R.,
Martin, M. and M. Thiel

DISCUSSION PAPER | 2020 / 12

The essentials on linear regression, ANOVA, general linear and linear mixed models for the chemist

Bernadette Govaerts^{a,*}, Bernard Francq^{a,b}, Rebecca Marion^a, Manon Martin^{a,c}, Michel Thiel^{a,d}

^a*Institute of Statistics, Biostatistics and Actuarial Sciences (ISBA), Louvain Institute for Data Analysis and Modeling (LIDAM), Université catholique de Louvain (UCLouvain), Voie du Roman Pays 20, bte L1.04.01, B-1348 Louvain-la-Neuve, Belgium*

^b*CMC Statistical Sciences, GSK Vaccines, Rue de l'Institut 89, 1330 Rixensart, Belgium*

^c*Fond National de la Recherche Scientifique, Rue d'Egmont 5, 1000 Brussels, Belgium*

^d*Statistics and Decision Sciences, Janssen Pharmaceutica, Turnhoutseweg 30, 2340 Beerse, Belgium*

Abstract

This text provides a brief and accessible guide for implementing general, ANOVA and linear mixed models for the analysis of real world data from chemistry, industrial, bio or life-sciences experiments. The reader is introduced to the main concepts and vocabulary of the subject and to the main statistical methods available for formalizing, estimating and performing statistical inference on these models. Mathematical notations are not avoided but kept accessible, methods are illustrated on several detailed applications and the use of appropriate computer software is discussed.

Keywords:

Analysis of Variance (ANOVA), Experimental design analysis, General linear model (GLM), Least squares estimators, Linear Mixed Model (LMM), Linear Regression, Maximum Likelihood Estimation, Random effect, R Software, SAS, Statistical Modeling

Contents

1	Introduction	3
2	Methodology	4
2.1	Empirical (linear) statistical modeling	4
2.2	Factors, responses and design matrix	6
2.2.1	Response and factors	6
2.2.2	Design matrix	6
2.2.3	Factor attributes	7
2.3	The General Linear Model: definition, mathematical formulation and model matrix	7
2.3.1	The General Linear Model family	7
2.3.2	Structure and assumptions of the GLM	7
2.3.3	Construction of the model matrix $\mathbf{X}$	8
2.4	Estimation, statistical inference and prediction in the fixed GLM	9
2.4.1	Estimation	9
2.4.2	Inference and sampling distribution of parameter estimators	10

*Corresponding author

Email addresses: bernadette.govaerts@uclouvain.be (Bernadette Govaerts), bernard.x.francq@gsk.com (Bernard Francq), rebecca.marion@uclouvain.be (Rebecca Marion), mm.manon.martin@gmail.com (Manon Martin), mthiel2@its.jnj.com (Michel Thiel)

2.4.3	<i>F</i> -tests for global significance of the model and for nested models	10
2.4.4	The <i>t</i> test on single model parameters and confidence intervals	11
2.4.5	Variable selection and model hierarchy	11
2.4.6	Prediction: confidence, prediction and tolerance intervals	12
2.4.7	Model validation	13
2.5	ANOVA decomposition and related F testing	13
2.5.1	Goal of ANOVA	13
2.5.2	Anova decomposition in the crossed ANOVA 2 model	14
2.5.3	F-test principle in ANOVA modeling	15
2.5.4	ANOVA model including random factors	16
2.5.5	Nested ANOVA	17
2.5.6	General remarks about ANOVA in balanced and unbalanced designs	17
2.5.7	Link between ANOVA and GLMs in fixed factor designs	17
2.5.8	Estimating variance components in an ANOVA framework	18
2.6	The Linear Mixed Model (LMM)	19
2.6.1	Goal and history	19
2.6.2	The LMM equation	19
2.6.3	Estimation of LMM parameters	20
2.6.4	Statistical inference for linear mixed models	21
2.6.5	Prediction and Tolerance intervals	22
3	Case studies	23
3.1	Case study 1: Paper bleaching DOE experiment	23
3.1.1	Context, experimental design and data	23
3.1.2	Model definition and estimation	24
3.1.3	Model validation	24
3.1.4	Interpretation of the model effects and related inference	25
3.1.5	Model visualization	26
3.2	Case study 2: Interlaboratory repeatability analysis	27
3.2.1	Context, experimental design and data	27
3.2.2	Model definition in the ANOVA and LMM forms	27
3.2.3	ANOVA results	28
3.2.4	LMM results: global model significance and residual graph	29
3.2.5	LMM results: analysis of fixed effects	29
3.2.6	LMM results: analysis of the random part of the model	30
3.3	Case study 3: Candy Sensory Experiment	31
3.3.1	Context, experimental design and data	31
3.3.2	Model definition in the ANOVA and LMM form	31
3.3.3	ANOVA results	32
3.3.4	LMM results: global model significance and residual graphs	33
3.3.5	LMM results: analysis of fixed effects	34
3.3.6	LMM results: analysis of the random part of the model	34
4	Statistical software	35
5	Final Remarks	36
6	Appendix 1 : Generic equations for the two-way fixed and mixed crossed ANOVA models	37
7	Appendix 2 : Variance-covariance matrix for a nested two-factor random model	38

8 Appendix 3 : Covariance Matrices for the Estimated Parameters in the Linear Mixed Model 39

9 Appendix 4 : Small sample null distribution of LR test statistic 40

1. Introduction

In experimental chemistry, it is common to study the effect of changes in one or more variables of interest on a measured response. This may serve numerous possible purposes: comparing the properties of several products, setting up a calibration line, optimizing the yield of a reaction, study the shape of the relationship between the inputs and outputs of a process, quantifying the repeatability and reproducibility of a measurement method, conducting an inter-laboratory study, analyzing the effect of a treatment on several subjects over time, and more.

An established scientific approach to answering these questions consists of first defining a statistical model. This model aims to describe the system of interest using a simple mathematical equation containing one or more random terms, which account for experimental and measurement errors or natural variation between subjects. Once the model is defined, it is then fitted to a set of process data collected according to a given experimental design. Finally, the fitted model is used to answer the questions of interest.

In this context, the linear model family is certainly the most commonly used model because of its ease of implementation and versatility. This family of models includes many models taught in graduate chemistry programs: simple and multiple regression, analysis of variance, screening and response surface models, variance component models, etc. Unfortunately, few chemists are aware of the fact that all of these models share a common general structure. Understanding how to express the random part of the process is also of the utmost importance in contexts where the chemist is involved in life science projects (e.g. -omics sciences) or must be able to accurately quantify the uncertainty of measurement systems.

This chapter provides a review of the main classes of linear models used by chemists for the analysis of experimental results. It essentially discusses four classes of models – linear regression, analysis of variance, general linear models and mixed models – and explains their specific features and the relationships between them. These models are applied on several case studies to illustrate each concept presented and to show how statistical software can be used for their implementation. This presentation is limited to situations in which the observed response is quantitative and the random part of the models is normally distributed. The keys to addressing more advanced models will also be given. The text is written to be accessible to chemists or other experimenters with basic statistical knowledge but will not overlook the main mathematical formulas in order to address the key concepts rigorously.

The literature available on linear models is extremely extensive. It includes books of all levels specifically dedicated to this subject (e.g. [1], [2], [3]), as well as books where the use of these models is discussed for specific applications, such as experimental designs (e.g. [4], [5]) or analytical chemistry ([6], [7]). This text is therefore not intended to provide a vast review of the literature but rather to give some useful references in which the reader will find more details. Note also that some methodological books cited here develop a vast, chapter-by-chapter list of variations of linear models (e.g. [2], [8]), while others are more concise, presenting the family of linear models in general terms and introducing each single model as a special cases (e.g. [9], [10]). These last books are of course less accessible to non-statisticians.

For the applications in this text, the open source software R will be highlighted as a possible tool for estimating linear models and deriving related solutions (see [11], [12], [13]). Other existing software will, however, be presented in a dedicated section. Throughout this chapter, *italics* will be used to emphasize new terminology (as well as the mathematical formulae), and **boldface** will be used for vectors and matrices.

On the basis of this text, the practitioners should be able to (1) choose and fit an appropriate linear model to real data issued from a research or industrial experimental design, (2) understand the underlying statistical concepts, (3) have a general view of the main mathematical notations and formulae used in the literature to formalize these concepts and (4) read and interpret the associated outputs from statistical software. The reader for whom formulae seem obscure should limit his or her reading to the concepts introduced and take more time on the case studies presented at the end of the chapter.

2. Methodology

2.1. Empirical (linear) statistical modeling

The aim of empirical statistical modeling is to approximate the behavior of a process of interest with a mathematical equation that accounts for the systematic and the random behavior of observed process data (see Figure 1). One possible goal of such a model is to understand the effect of changes in a set of predictors ($\xi_1, \xi_2, \dots, \xi_k$) (the input of the process), on one or more observed responses Y (the process output). Another common objective is to estimate an equation for predicting the value of the response for a given set of factors values without having to actually run the process and quantify the uncertainty of these predictions.

Figure 1: Statistical modeling of an experimental process.

In this context, the family of linear models provides a large variety of potential models for answering scientific questions using experimental data. These models have been popular since their introduction (which can be traced back to Adrien-Marie Legendre or Carl Friedrich Gauss for the simple ordinary least square method [14], [15]) and are still widely used today. Linear models can take different forms depending on the problem of interest. Over the years, different generic classes of models have emerged that have been assigned different names and mathematical equations. The practitioner, however, may not always see the relationships between these model classes and equations. Basic examples of linear model equations are given below and Figure 2 illustrates some of them graphically.

- Simple linear model, used for example in calibration models: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$
- Single variable polynomial model: $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_{(p-1)} x_i^{(p-1)} + \epsilon_i$
- Multiple regression model: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{(p-1)} x_{i(p-1)} + \epsilon_i$
- Response surface model: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \epsilon_i$
- ANOVA 1 model: $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$
- Crossed ANOVA 2 model: $y_{ijh} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijh}$
- Nested ANOVA 2 model: $y_{ijh} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijh}$
- Analysis of covariance model: $y_{ij} = \beta_0 + \beta_1 x_{ij} + \alpha_i + \delta_i x_{ij} + \epsilon_{ij}$

The families of linear models discussed in this text share three common characteristics: (1) the response Y explained by the model is always assumed to be observed on a quantitative continuous scale, (2) the expression of the model equation is a sum of parameters multiplied by the values of some predictors of interest in the study and (3) the last term added to the model is an error term ϵ , which accounts for some or all of the random components of the process of interest. These shared characteristics may not necessarily be apparent in the set of equation above, but this will be further explained in the next sections. Note also that the word *linear* does not mean that the systematic part of the model has a linear shape. For example, the response surface model (also called quadratic polynomial model) has a non linear shape but is a linear statistical model because it is expressed as a linear function: a sum of parameters multiplied by

Figure 2: Examples of linear models.

transformations of the factors of interest (X_1^2 and X_1X_2 in the example). An example of a non-linear model which can not be expressed in this form is the Michaelis-Menten equation: $V_i = \frac{v_{\max}[S]_{0i}}{K_M + [S]_{0i}} + \epsilon_i$.

In chemometrics, linear models are mostly used in *experimental* studies, i.e. studies where data are intentionally generated from a given process (analytical, experimental or industrial) in order to answer a specific scientific question. Ideally, these data are generated on the basis of an experimental design to guarantee that they will contain the information targeted by the study and have a balanced structure. Data used to develop linear models can also come from an *observational* study, where data are not produced for a given purpose, but are rather collected from a process or in a certain context. Such data often have less appealing properties in terms of statistical modeling: they may present a high degree of collinearity between predictors, contain very influential observations and outliers, and be badly spread in the domain of the factors of interest. This text will mostly be focused on data coming from experimentally designed studies, though not exclusively. For this reason, the problem of model and data validation will not be discussed outside of the small section 2.4.7, which is mainly devoted to references.

The potential applications of this chapter are of course not restricted to the simple models listed above. Indeed, the framework of general and mixed linear models is able to describe many more complex experimental situations, such as split plot, latin squares, cross-over or incomplete blocked designs, repeated measures and longitudinal studies or variance component analysis.

Figure 3 is a diagram that summarizes the steps of a statistical modeling, from the study design to the interpretation of results. First, an experimental design is chosen based on sample size constraints and the mathematical equation (model) that must be estimated. The different inputs or predictors of the process of interest are listed as columns of the design matrix and the experiment trials in rows. The experiment is then run and the response variable(s) recorded. At this stage, descriptive analyses should be performed to assess data quality and to obtain a first glimpse of the relationship between the inputs and outputs of the studied process. The model (specified before the experiment) is then fitted to the data with the appropriate software and validated with various possible criteria. If needed, the model is simplified or modified and re-fitted. Finally, the final model results can be interpreted using various possible tables and graphs that

can be easily generated by dedicated statistical software. The experimenter is then able to answer to the research questions and confirm the significance of his or her results with statistical inference tools. The last step is usually the optimization of the process by choosing the more impactful predictors and the predictor values that maximize or minimize the response variable, or attain a target response value.

Figure 3: Model building steps

2.2. Factors, responses and design matrix

This section defines some concepts, notations and vocabulary related to the inputs and outputs of the model and the nature of the data available for fitting it.

2.2.1. Response and factors

First, at the output of the process, one or several variables of interest are observed. Depending on the context, these variables are called the *responses* or the *dependent* variables. In this text, models for a single response, noted Y , will be studied, but many concepts discussed here remain valid for multiple response cases treated by e.g. MANOVA (Multivariate ANalysis Of Variance), ASCA (Anova Simultaneous Component Analysis) or PLS (Partial Least Square) models ([16], [17], [18]). The response of interest is assumed to be quantitative and available for n observations $(y_1, y_2, \dots, y_n)$ gathered in a $(n \times 1)$ vector $\mathbf{y}$. In the model, the response is sometimes expressed in a transformed form in order to reflect the real behavior of the process: $\log(Y), 1/Y, \sqrt{Y} \dots$

Next, the variables of interest at the input of the process (or the model), are given several names in the literature: *predictors*, *explanatory* variables, *independent* variables or *factors*. To avoid confusion, this text uses the word *factor* to refer to the k variables of interest in the experiment and denotes them $(\xi_1, \xi_2, \dots, \xi_k)$. The word “variable” will be reserved to name the variables appearing in the model equation and denoted X_j .

2.2.2. Design matrix

The values taken by the different factors in the n experiments used to estimate the model are gathered together in a $(n \times k)$ table called the *design matrix* ξ . The value of factor j for observation i is ξ_{ij} . The terminology *design matrix* does not necessarily suppose that the data have been generated from a well organized Design of Experiment (DOE). Many types of linear models are, in any case, directly linked to

typical experimental designs that guarantee their identifiability and quality. In chemometrics applications, there are several possible designs commonly associated with linear models, such as screening, full or fractional factorial, response surface, split plot or latin square designs or designs for cross over or longitudinal studies ([2], [5], [19]).

2.2.3. Factor attributes

A factor ξ_j at the input of a process can be described by several attributes. It may be either quantitative or categorical/qualitative. A quantitative factor ξ_j takes values on a continuous or an ordered discrete numerical scale (e.g. a temperature, a dose), and a categorical factor takes values within a (not necessarily ordered) list of l_j possible factor *levels* (e.g. a type of reactive, an operator).

A factor may also be defined as *fixed* or *random* in the model. It is fixed when its values are chosen from the set of possible levels of the factor and the goal of the experiment is to understand and predict how the mean response changes when switching from one level to another (e.g. the yield of a chemical process for different possible values of a temperature or different types of catalysts). A factor is random when its values have been randomly chosen from a large population of possible values and the goal of the study is not to be able to predict the behavior of the response for one given value of the factor but rather to study the effect of the factor on the variability of the response in the population (e.g. the between-day variability of an analytical method).

In the design matrix, a given factor may be either *crossed* with or *nested* within another factor. Two factors are crossed when all or some possible combinations of the levels of the factors are tested in the design. A factor ξ_B is nested in factor ξ_A if levels of factor ξ_B are hierarchically subsumed within the levels of factor ξ_A (e.g. measurement devices are nested in different laboratories). Finally, experiments are sometimes organized in *blocks* in order to take into account the fact that experimental conditions may differ from one group of experiments to another (e.g. day, reactor, reagent kit).

2.3. The General Linear Model: definition, mathematical formulation and model matrix

2.3.1. The General Linear Model family

The General Linear Model (GLM) offers a broad framework for explaining a response Y based on a set of factors $(\xi_1, \xi_2, \dots, \xi_k)$. This framework includes a large set of possible generic models but there is no clear consensus in the literature as to the exact boundaries of this set.

This text is restricted to the types of general linear models that are most commonly used in the field of chemometrics, in particular the models including only one response variable Y that is explained by k fixed quantitative or categorical factors. It also assumes that the observed responses $(y_1, y_2 \dots y_n)$ are independent of each other and have the same variance. These models include simple or multiple linear regressions, polynomial models, fixed ANOVA models and analysis of covariance models. The literature and statistical software sometimes include additional models in the GLM class, such as cases where multiple responses must be explained simultaneously by the model, where the model errors have a general covariance structure and where some input factors are random. This latter problem will be discussed in Section 2.6, and the other two are beyond the scope of this text. Note also that the acronym GLM can lead to some ambiguity because it designates also another very large class of statistical models called *generalized linear models*, a generalization of the general linear model to non-normal responses [20, 10]. This topic will not be developed at all in this text.

2.3.2. Structure and assumptions of the GLM

A general linear model is built from the addition of a series of F model *effects*, each defined as a function of the experimental factors $g_f(\xi_1, \xi_2, \dots, \xi_k)$, $f = 1, \dots, F$. In addition to the intercept 1, common effects include main effects ξ_j , interaction effects $\xi_i \times \xi_j$, quadratic effects ξ_j^2 , etc... If all factors ξ_j included in a given model effect g_f are quantitative, the effect will corresponds to one *term* in the model equation i.e. one unknown parameter β_f and one model variable denoted as X_f . The modeling of categorical factors is less straightforward : each model effect including categorical factors can be associated with several model terms, which are themselves associated with a vector of parameters β_f . This is discussed later in this section.

The univariate form for the general linear model is commonly written as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots \beta_{p-1} x_{i,(p-1)} + \epsilon_i$$

where i is the observation or experiment index ($i = 1, \dots, n$) and the errors ϵ_i are assumed to be independent of each other, have a zero mean and constant variance σ^2 and follow a normal distribution. This last assumption is only necessary for performing inference on model parameters. The number of parameters in the model is p and, if all model effects are quantitative, p is equal to $F + 1$ the number of model effects plus the constant term (the intercept).

In matrix form, the general linear model takes the following well-known form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{y}' = (y_1, y_2, \dots, y_n)$ is the vector of all observed responses, $\mathbf{X}$ is the *model matrix* of size $n \times p$ containing the values x_{ij} of the model variables X_j for each observation i , $\boldsymbol{\beta}' = (\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1})$ is the $p \times 1$ vector of parameters and $\boldsymbol{\epsilon}$ is the $(n \times 1)$ vector of errors, assumed to follow a multivariate normal distribution $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. In order to estimate parameters, we will see in Section 2.4 that it is also necessary that the matrix $\mathbf{X}$ is of full rank (i.e. non-singular), meaning that no column of $\mathbf{X}$ can be expressed as a linear combination of the other columns.

2.3.3. Construction of the model matrix $\mathbf{X}$

Each single or combination of quantitative factors included in a GLM corresponds to one column of $\mathbf{X}$. In the analysis of designed experiments, these factors are often *coded* before being introduced in $\mathbf{X}$ in order to improve the interpretation of the results and/or the numerical stability of the model. The two most common transformations applied are *centering* by the factor mean or domain center and *scaling* by the factor standard deviation or half the size of the factor domain.

When the model includes categorical factors, each factor is first coded appropriately before being incorporated into the model. Let's take, for example, the simple ANOVA model (discussed in detail in Section 2.5): $y_{ij} = \mu + \alpha_j + \epsilon_{ij}$, $i = 1, 2, \dots, n_j$ and $j = 1, \dots, l$. This model aims to study by which amount α_j the overall mean μ of the response Y changes according to the $j = 1, \dots, l$ possible levels of a given categorical factor ξ . The number of replicated observations available for each level of ξ is given by n_j . If $l = 3$ and $n_j = 2$, the ANOVA 1 model can be expressed as a general linear model as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \begin{pmatrix} y_{11} \\ y_{21} \\ y_{12} \\ y_{22} \\ y_{13} \\ y_{13} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{21} \\ \epsilon_{12} \\ \epsilon_{22} \\ \epsilon_{13} \\ \epsilon_{23} \end{pmatrix}$$

A categorical factor with three levels is then *coded* in the model matrix with three *dummy variables*. But this is not the end of the story : as this model matrix $\mathbf{X}$ is not of full rank (its 4th column c_4 can be calculated as $c_4 = c_1 - c_2 - c_3$), the model is then *non-indentifiable* and will accept an infinite set of equivalent possible values for the parameters. In practice, this problem is solved by imposing constraints on the model parameters. The most commonly used and natural constraint in chemical experiments supposes that the effects of the factor levels on the overall mean sum to zero: $\sum_{j=1}^l \alpha_j = 0$. This implies that any parameter (e.g. the last) can be calculated from the others as follows: $\alpha_l = -\sum_{j=1}^{l-1} \alpha_j$. Another approach sets one level of the factor as the reference level (for example the last level) and estimates the effect of changing the factor level from this reference to another level. This is especially useful in medical studies when different drugs are compared to a placebo.

These two approaches lead to two equivalent reparameterizations of the initial model, which have the advantage of yielding full rank $\mathbf{X}$ matrices. In each matrix the main effect of the categorical factor is coded by $(l - 1)$ columns linked to $(l - 1)$ parameters being estimated.

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mu + \alpha_3 \\ \alpha_1 - \alpha_3 \\ \alpha_2 - \alpha_3 \end{pmatrix}$$

Depending on the literature or the software used, the first coding approach is called sum-to-zero or deviation coding and the second is called set-to-zero or treatment coding ([21], [8], [9]). Some software also propose other less common types of coding (e.g. Helmert contrasts or orthogonal polynomials) which are of less interest in chemometrics. Statistical and experimental design software usually generate these model matrices automatically from a list of effects provided by the user. This technical problem of coding is then quite transparent for the user. When there is an interaction effect between two categorical factors in the model (e.g. ξ_A and ξ_B), the effect will be included in a given line of the model matrix by computing simply the element-wise product or the *Kronecker* product between the $(l_A - 1)$ coded values linked to ξ_A with every $(l_B - 1)$ codes values linked to factor ξ_B . This results in $(l_A - 1)(l_B - 1)$ interaction columns in the model matrix or $(l_A - 1)(l_B - 1)$ parameters in the model. The interaction between ξ_A and ξ_B is often denoted $\xi_A \times \xi_B$. When a categorical factor ξ_B is nested within another categorical factor ξ_A , statistical software commonly use the notation $\xi_B(\xi_A)$ where ξ_B is the *child* effect and ξ_A is the *parent* effect.

For a given GLM model with F effects, the construction of the model matrix $\mathbf{X}$, results then in a matrix organized into $(F + 1)$ blocks $\mathbf{X} = (\mathbf{1} \mid \mathbf{X}_1 \mid \mathbf{X}_2 \mid \dots \mid \mathbf{X}_F)$, each block corresponding to an effect f in the model. $\mathbf{X}_f$ is of size $n \times p_f$. $p_f = 1$ if the effect f contains only quantitative variables (e.g. X^2), but it may be larger than one if categorical factors are included in the effect. The vector of parameters may be decomposed accordingly $\beta' = (\beta_0 \mid \beta'_1 \mid \beta'_2 \mid \dots \mid \beta'_F)$.

2.4. Estimation, statistical inference and prediction in the fixed GLM

When the equation of the linear model of interest is defined, the next step consists of fitting the model to the data, validating the model quality and using the fitted model to interpret the effect of the factors on the response and to make predictions. The statistical concepts related to these modeling steps are developed below for the general linear model. Most of the formulae correspond to the outputs automatically calculated by established statistical modeling software. These outputs can be arranged in very standard and convenient tables, such as the model parameter table or analysis of variance table. A case study in section 3.1 provides an application of most of these formulae and illustrations of the related tables.

2.4.1. Estimation

The general linear model involves two sets of unknown parameters: the vector β and the variance of the errors σ^2 . In GLMs, the best way to estimate the regression parameters β is the *Ordinary Least Squares (OLS)* approach. The least square estimate of β minimizes the sum of the squares of the residuals of the model $SSE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} \dots - \hat{\beta}_{p-1} x_{i(p-1)})^2$ and is calculated, in matrix form, as:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

The Gauss-Markov theorem establishes that $\hat{\beta}$ is the best linear unbiased estimate (BLUE) when the errors ϵ_i are uncorrelated and have equal variances σ^2 . *Linear* means that $\hat{\beta}$ is a linear function of the response $\mathbf{y}$, *unbiased* means that the expected value of the estimator is the unknown true value of the parameters $E(\hat{\beta}) = \beta$ and, finally, *best* means that the uncertainty of $\hat{\beta}$ (i.e. its variance) is minimal. The OLS estimator is equivalent to the *maximum likelihood estimator* (ML) when the errors are independent and normally distributed. A ML estimator can be interpreted as the value of β that maximizes the probability of observing the available data [9].

From these estimators, the *predicted* or *fitted* values of the responses $\hat{\mathbf{y}}$ are calculated as follows

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

and may be interpreted as the orthogonal projection of $\mathbf{y}$ onto the space spanned by the columns of $\mathbf{X}$.

An unbiased estimate of the error term variance σ^2 is then obtained with the following formula:

$$S^2 = \hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{1}{n-p} SSE = MSE.$$

The residual variance, S^2 , quantifies the closeness between the observed and fitted responses. The degrees of freedom, $n-p$, represent the amount of independent information left for estimating the variance of the residuals after estimating the p parameters of the model. SSE and MSE denote respectively *sum of squares* and *mean square* of the errors.

The formula for S^2 may also be obtained with the maximum likelihood (ML) estimation approach, but the ML denominator of S^2 is n in this case and must further be corrected to $n-p$ to obtain an unbiased estimator. This is a common problem in the derivation of ML estimators of variance components.

As already discussed above, the OLS estimator assumes that the model matrix is not singular, which means that the matrix $\mathbf{X}'\mathbf{X}$ can be inverted. If this is not the case, some solutions based on generalized inverses exist but are beyond the scope of this text [22].

2.4.2. Inference and sampling distribution of parameter estimators

Statistical theory for linear models is not restricted to the estimation of model parameters. It also provides powerful *inference* tools for assessing model pertinence, comparing different possible models, testing the significance of individual model parameters or quantifying the accuracy of response prediction. In addition to the assumptions of independence and constant variance of the model errors ϵ_i , which are necessary for OLS estimation, these tools also assume the normality of the errors. This hypothesis makes it possible to derive the *sampling distribution* of the regression parameters:

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

which must be read as follows: “the estimators $\hat{\boldsymbol{\beta}}$ follow a multivariate normal distribution with mean $\boldsymbol{\beta}$ (which means that the estimators are unbiased) and with variance-covariance matrix: $V(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ of size $(p \times p)$ ”. The diagonal elements of $V(\hat{\boldsymbol{\beta}})$ are the variances of the parameters, which are used to evaluate the uncertainty of the parameters estimated in the model. The off-diagonal terms provide the covariances between parameters, which give an indication of the quality of the experimental design. It can also be shown that the sampling distribution of the residual variance estimator is proportional to a Chi-squared distribution with $n-p$ degrees of freedom: $(n-p)S^2/\sigma^2 \sim \chi_{n-p}^2$ and that $\hat{\boldsymbol{\beta}}$ and S^2 are independent of each other.

2.4.3. F-tests for global significance of the model and for nested models

The first inferential tool usually applied after fitting a linear model is the *global F-Test*. This test simply aims to answer the question: “does the model explain any part of the response ? ” or “would a model containing only a constant term be as good as the chosen model ?”. This may be expressed in terms of a statistical hypothesis as follows:

$$H_0 : \beta_j = 0 \text{ for } j = 1, \dots, (p-1), \quad H_1 : \exists j \text{ such that } \beta_j \neq 0$$

The test statistics available for evaluating this hypothesis are based on a simple principle. SSE_0 , the *sum of squares of the residuals* of the model under H_0 (i.e. containing only a constant term $y_i = \beta_0 + \epsilon_i$) is compared to the SSE_1 of the full model (i.e. the model containing all of the model terms). A test statistic is then built to measure if SSE_0 is *significantly* greater than SSE_1 , which is an indication that the full model explains at least some variability in the response. The test statistic takes the following form:

$$F_{obs} = \frac{(SSE_0 - SSE_1)/(p-1)}{SSE_1/(n-p)} \sim F_{(p-1);(n-p)} \text{ under } H_0 \text{ and p-value} = P(F_{obs} > F_{(p-1);(n-p)})$$

Thanks to the normality assumption, one can show that this statistic follows an F distribution with $(p - 1)$ and $(n - p)$ degrees of freedom under H_0 , making it possible to derive a related p -value for the test. A small p -value (e.g. ≤ 0.05) indicates that the null hypothesis H_0 may be rejected.

One can also show that $SSE_0 = \sum_{i=1}^n (y_i - \bar{y})^2$, which is also called the *total* sum of squares SST. $SSE_0 - SSE_1$ is also called *SSM* the *model* sum of squares. This will be further detailed in Section 2.5.

The F -test of global significance of the model may be generalized to the comparison of any pair of models M_0 and M_1 as long as model M_0 is a linearly restricted form of M_1 . In mathematical form, the related test hypotheses can be written as follows:

$$H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}, H_1 : \mathbf{L}\boldsymbol{\beta} \neq \mathbf{0} \quad F_{obs} = \frac{(SSE_0 - SSE_1)/(p_1 - p_0)}{MSE_1} \sim F_{(p_1 - p_0);(n - p_1)} \text{ under } H_0$$

where $\mathbf{L}$, the *contrast* matrix, expresses which linear combination of the M_1 model parameters are null in M_0 . The number of parameters in model M_1 is p_1 and the number of parameters in M_0 is p_0 . This test is very widely used in regression and analysis of variance because of its generality. The $\mathbf{L}$ matrix may, for example, be defined to simply test if one parameter of the model is *null* or not: $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$, or if the $q - 1$ parameters associated with a categorical variable are non-significant, or if two parameters of the model are equal or differ from each other: $H_0 : \beta_j = \beta_{j'}$ vs. $H_1 : \beta_j \neq \beta_{j'}$. It is also important to note that, in the GLM context, this F -test is equivalent to another test commonly used in the comparison of embedded models: the *likelihood ratio* test [9]. This test will reveal its importance when random factors are introduced in the model, and it will be discussed in Section 2.6.

2.4.4. The t test on single model parameters and confidence intervals

The t test is another test widely used in regression analysis and related software outputs. It makes it possible to answer the question: “Can a predictor X_j be removed from the model ?” or “Does the variable X_j have a significant effect on the response ?” Or, equivalently, it can test the hypothesis $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$. This test is equivalent to the F -test presented above but has the advantage of allowing also to test unilateral hypotheses like $H_0 : \beta_j \leq 0$ vs. $H_1 : \beta_j > 0$ or $H_0 : \beta_j \geq 0$ vs. $H_1 : \beta_j < 0$. It can also be used to test if a parameter is equal to a given non null constant β_{j0} :

$$H_0 : \beta_j = \beta_{j0}, H_1 : \beta_j \neq \beta_{j0}$$

The general form of the t -statistic is

$$t_{obs} = \frac{\hat{\beta}_j - \beta_{j0}}{s(\hat{\beta}_j)} \text{ and with } s(\hat{\beta}_j) = S\sqrt{(\mathbf{X}'\mathbf{X})_{(j+1)(j+1)}^{-1}},$$

and it can be shown that, under H_0 and the normality assumption for the model error, it is distributed as a t random variable with $(n - p)$ degrees of freedom.

When the normality assumption is not verified for model errors, the tests presented above are no longer valid but solutions exist, like permutation or bootstrap tests, for answering similar questions [23, 24].

The confidence interval is another useful inferential tool in statistical modeling. Confidence intervals make it possible to express the uncertainty of the model parameters by providing an interval around the estimator $\hat{\beta}_j$ that has a given probability $1 - \alpha$ (e.g. 0.95) to include the real value of the parameter β_j . This interval can be calculated as follows:

$$\hat{\beta}_j \pm t_{n-p;1-\alpha/2} s(\hat{\beta}_j)$$

where $t_{n-p;1-\alpha/2}$ is the $(1 - \alpha/2)\%$ quantile of a t distribution with $(n - p)$ degrees of freedom. Note that the confidence interval provides more information than a single hypothesis test about possible values for the model parameters.

2.4.5. Variable selection and model hierarchy

These inferential procedures provide measures of the importance of effects introduced into a model and can be used to decide whether to remove the effects that are not significant. Based on the principle of

parsimony, a model with few terms will indeed always be preferred to a larger model because it will be simpler to interpret and will give predictions of better quality.

Nevertheless, caution should be taken for this type of approach. Firstly, in modeling experimental results, the hierarchy between effects should always be respected within the model terms in order to avoid interpretation difficulties. For example, a non-significant first order effect X_i should not be removed from the model if the quadratic effect X_i^2 or interaction effect $X_i \times X_j$ is present. Second, p-values are not considered to be the best way to assess the relevance of a term in a model. Many other methods and criteria for variable selection exist and are discussed in the statistical literature [25, 11].

2.4.6. Prediction: confidence, prediction and tolerance intervals

The concept of confidence intervals is also useful for the quantification of model prediction uncertainty. For a given set of (new) values of model predictors $\mathbf{x}'_* = (1, x_{*1}, x_{*2}, \dots, x_{*(p-1)})$, the predicted response is simply calculated as follows:

$$\hat{y}_* = \mathbf{x}'_* \hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \hat{\beta}_1 x_{*1} + \hat{\beta}_2 x_{*2} + \dots + \hat{\beta}_{p-1} x_{*(p-1)}.$$

The uncertainty of $\hat{y}_*$ is quantified by a standard error based on the property that $\hat{y}_*$ is a simple linear combination of $\hat{\boldsymbol{\beta}}$:

$$S(\hat{y}_*) = S \sqrt{\mathbf{x}'_* (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_*}.$$

Two types of intervals can be derived from this standard error. A *confidence interval for the mean response* i.e. the mean of the responses one would obtain after conducting an infinite number of trials in $\mathbf{x}_*$:

$$\hat{y}_* \pm t_{n-p; 1-\alpha/2} S \sqrt{\mathbf{x}'_* (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_*}$$

and the *prediction interval for a single observation* of the response:

$$\hat{y}_* \pm t_{n-p; 1-\alpha/2} S \sqrt{1 + \mathbf{x}'_* (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_*}.$$

In this formula, the 1 expresses that the prediction of a single observation is affected by the uncertainty of the model parameters + the uncertainty on the error term ϵ_* .

Both concepts can actually be generalized into the same mathematical formula as follows:

$$\hat{y}_* \pm t_{n-p; 1-\alpha/2} S \sqrt{1/q + \mathbf{x}'_* (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_*},$$

where $q = \infty$ for the confidence interval and $q = 1$ for the prediction interval. This last interval is very useful in chemistry, i.e. for deriving an interval for the mean of q future samples.

The prediction interval is also analog to the *beta-expectation tolerance interval* [26, 27, 28, 29]. Both have the same mathematical formula but their definition and interpretation are different. A beta-expectation tolerance interval is an interval that should contain a given proportion of the population, on average. For example, a 95% beta-expectation interval will contain 95% of future values, on average. This means that in practice, the interval can sometimes be too narrow (it contains less than 95% of the population) or too large (it contains more than 95% of the population) but will encompass the desired prediction level on average. Another (and more general) approach is the *beta-gamma content tolerance interval*, as it contains a given proportion of the population (the prediction level) with a given confidence level. For example, a 95%-80% beta-gamma content tolerance interval will contain at least 95% of the population in 80% of cases. This means that it will be too large in 80% of cases (and too narrow in 20% of cases). This confidence interval can then be interpreted as a safety margin added to the beta-expectation tolerance interval. It can be shown that beta-gamma content tolerance intervals are usually larger than beta-expectation tolerance intervals when gamma is greater than 50%.

Tolerance intervals are very useful for small sample sizes as the uncertainty of the prediction is accounted for and may be controlled with a confidence level. They are widely applied in product quality control, analytical method validation or risk analysis in industry (i.e. the pharmaceutical or chemical industries). When the sample size increases, the safety margin collapses and the beta-gamma content tolerance intervals

move closer to the beta-expectation tolerance interval or equivalently the prediction interval. Note finally that it may be very useful in practice to present the three statistical intervals “side-by-side” – the confidence, the prediction, and the (beta-gamma content) tolerance intervals – as this will provide a “full picture” to the practitioner.

2.4.7. Model validation

When a statistical model has been set up for a certain application, it is essential to examine its adequacy with respect to the data before drawing any inferential conclusions. Whole chapters of linear model books are devoted to the subject of model validation, which is beyond the scope of this text ([30], [31], [2], [32]). The section below enumerates the most common tools applied when analyzing the results of designs of experiments. When data are issued from observational studies, other diagnostic tools are recommended in order to detect influent observations in the data set, as well as to evaluate colinearities between variables, potential ill conditioning and identifiability problems in the model matrix. Nowadays, in regression-type models, cross validation tools have also become popular for assessing the predictive power of models, but they are not of central interest in the analysis of experimental results.

Goodness of fit tools should first be used to quantify both model significance and the closeness of the data to the model. For this purpose, the most well-known index is the coefficient of determination R^2 , which measures the percentage of variance in the responses that is explained by the model. It is basically calculated as $R^2 = 1 - \frac{SSE}{SST}$ but often replaced by its variation $R^2_{\text{Adjusted}} = 1 - \frac{SSE/(n-p)}{SST/(n-1)}$, which accounts for the degrees of freedom of the model and decreases when the model is less parsimonious. What is considered to be a good value for the R^2 depends on the context. For a calibration line, a R^2 larger than 99% is common, while for biological data, an R^2 close to 70% is already considered high. The R^2 is usually complemented with the global F -test presented in Section 2.4.3, which assess the model’s significance, and by a plot of the responses y_i with respect to the fitted responses $\hat{y}_i$ (Y - Y fit plot), which shows the quality of the fit.

Next, in general linear models, the errors are assumed to be independent, normally distributed and of constant variance σ^2 . Independence is usually assumed to be a direct consequence of the experimental set up, where each trial is conducted independently of the others, or where existing dependencies in the model structure are taken directly into account in the design (e.g. split-plot designs or longitudinal studies). The normality hypothesis may be assessed by graphically studying the residuals $e_i = y_i - \hat{y}_i$ or their standardized values $e_i/s(e_i)$. A quantile-quantile plot of the residuals provides a visual representation of the normality of the errors. The graph of the residuals with respect to the predicted responses makes it possible to detect several other potential problems: outlier data points, variance heterogeneity (e.g. when the points have a funnel shape) or inadequacy of the model equation when, for example, the points form a “banana” shape.

When the experiment contains replicated trials, the *Lack-of-Fit* test also makes it possible to check the linearity assumption of the general linear model. Several approaches exist also for testing the homogeneity of variances (e.g. Levene, Bartlett, etc.).

When these tools detect that model assumptions are not fulfilled, solutions exist for improving, or attempting to improve, the situation. Normality and heterogeneity of variance problems may often be resolved by applying a *Box-Cox* family transformation to the response (e.g. *log* or *square root*). Weighted least squares may also be applied in order to account for heterogeneity of variances or other covariance structures.

All of the tools discussed in this section are also useful for the models discussed in the next sections, along with potential adaptations.

2.5. ANOVA decomposition and related F testing

2.5.1. Goal of ANOVA

Analysis of variance (ANOVA) is a family of statistical methods that is mainly used when analyzing the statistical influence of a set of categorical factors on a quantitative response, ideally in studies with a balanced experimental design. The total variation observed in the response of interest is decomposed into the variability explained by the different effects of a model chosen according to the structure of the

underlying experimental design. The main goal of the method is to measure the importance and significance of each effect but several other results, presented in this section, can be derived.

ANOVA for a design with only one fixed factor with q levels is called ANOVA 1 and is a generalization of the t -test for comparing two independent means. In its more general form, ANOVA may be applied to any multifactor experimental or observational studies including fixed or random factors. These factors may be combined with each other using a crossed and/or hierarchical structure.

ANOVA was originally developed by the statistician and evolutionary biologist Ronald Fisher at the beginning of the 20th century [33, 34]. Nowadays, ANOVA is sometimes considered obsolete in the statistics community, due to the development of general linear and mixed linear models, which include ANOVA models as special cases. Despite this, the ease of use and certain characteristics of ANOVA models remain of interest and will be highlighted in this text, as well as the links between ANOVA and GL/LM models.

In chemometrics, ANOVA may not be well-known among practitioners, despite the fact that some early publications show an interest in it (see e.g. [35]). ANOVA is, moreover, the basis of several methods used to estimate variance components in the development and validation of analytical methods. In addition, ANOVA has recently received increased interest with the development of ASCA-like methods used to analyze experimental designs with multidimensionnal responses ([17], [36], [37], [38]).

2.5.2. Anova decomposition in the crossed ANOVA 2 model

A typical ANOVA analysis is based on the following ingredients: the ANOVA model, an equation for the decomposition of the total response variation, an ANOVA table and F -tests to test the significance of each model effect. ANOVA with two factors makes it possible to illustrate the different facets of these concepts in detail.

For example, let's suppose that the effect of two fixed categorical factors ξ_A and ξ_B on a response Y are studied with a full factorial design including r replicates for each combination of the l_A and l_B levels of the two factors. The related two-way crossed ANOVA model is usually written as follows:

$$y_{ijh} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijh} = \mu_{ij} + \epsilon_{ijh}$$

where y_{ijk} is the value of the response variable for the i -th level of factor ξ_A ($i=1$ to l_A), the j -th level of factor ξ_B ($j=1$ to l_B) and the h -th observation ($h=1$ to r). The overall mean value of the response variable is μ and the response mean for a given combination of factor levels, or *treatments*, ij is μ_{ij} . The α_i 's and β_j 's are parameters related to the main effects of ξ_A and ξ_B and the $\alpha\beta_{ij}$'s are the interaction effect parameters. The ϵ_{ijh} 's are the errors, which are assumed to be $iN(0, \sigma^2)$, like with the general linear model.

The *analysis of variance equation* is built as follows (see also Figure 4). First, the *Total Sum of Squares SST* corresponding to the total variation of the data is decomposed into two terms: the systematic variation due to the variation from one treatment to the other, called *SSM*, the sum of squares of the model, and *SSE*, the sum of squares of the errors, i.e. the random variation between the replicates within each treatment ij .

$$SST = \sum_{ijh} (y_{ijh} - \bar{y}_{..})^2 = SSM + SSE = \sum_{ijh} (\bar{y}_{ij} - \bar{y}_{..})^2 + \sum_{ijh} (y_{ijh} - \bar{y}_{ij})^2 \quad (1)$$

where $\bar{y}_{..}$ is the overall arithmetic mean of the data, $\bar{y}_{ij}$ is the mean of the data for treatment ij .

When the design is balanced (i.e. the number r of replicates for each treatment is identical), *SSM* may then be decomposed into parts of variation linked to each effect in the model:

$$SSM = SSA + SSB + SSAB = \sum_{ijh} (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{ijh} (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{ijh} (\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \quad (2)$$

where $\bar{y}_{i.}$ (resp. $\bar{y}_{.j}$) is the mean of the responses for level i (resp. j) of factor ξ_A (resp. ξ_B). This decomposition is illustrated in Figure 4.

This decomposition in sum of squares allow to directly quantify the part of the total variation of the response SST explained by each term of the ANOVA decomposition. Two other concepts are then related to the sum of squares included in this decomposition. The *degrees of freedom* of each *SS* indicates the number

Figure 4: ANOVA 2 decomposition

of independent items used to calculate it. For example, the degrees of freedom of SST is $(n - 1)$ because it is based on n observations minus 1 degree of freedom lost for the estimation of the overall mean. $(n - 1)$ is then broken down to the independent amount of information linked to each term. The *Mean Squares (MS)* are then simply calculated as the sum of squares divided by their respective degrees of freedom and take the same form as sample or residual variances. The sections below explain the usefulness of these concepts.

All of these results are traditionally summarized in an *analysis of variance table*, which takes the form illustrated in Table 1 for the two-way model when the two factors are fixed.

Table 1: Crossed two ways ANOVA table

Effect	Degrees of freedom	Sum of Squares	Mean Square	F-ratio
A	$l_A - 1$	SSA	MSA	MSA/MSE
B	$l_B - 1$	SSB	MSB	MSB/MSE
AB	$(l_A - 1)(l_B - 1)$	SSAB	MSAB	MSAB/MSE
Error	$n - l_A l_B$	SSE	MSE	
Total	$n - 1$	SST		

2.5.3. F-test principle in ANOVA modeling

The last column of the ANOVA table (Table 1) necessitates a little more explanation. It contains the test statistics used for testing the significance of each effect included in the model: the two main effects and the interaction effect. Traditionally, the choice of these tests statistics is justified on the basis of a concept called the *Expected Mean Square (EMS)*.

The expected value of the mean square associated with a model effect f ($E(MS_f)$) represents the combination of model parameters quantified by the mean square when it is calculated of the data. One can show that, for the balanced fixed effect crossed ANOVA 2 model, these are given by the formulae provided in the first column of Table 2.

On the basis of these formulae, the MSE can then be considered as an unbiased estimator of σ^2 and the other MS's as unbiased estimators of the corresponding combinations of model parameters. These *EMS*'s are usefull to define the model effect test statistics as follows. For example, to test if the main effect of factor ξ_A is significant, the hypotheses are written as follows:

$$H_0 : \alpha_1 = \alpha_2 = \dots \alpha_{l_A} = 0 \text{ against } H_1 : \text{at least one } \alpha_j \neq 0$$

Table 2: Expected mean squares for the fixed, random and mixed balanced ANOVA 2 models

Effect	Fixed Model	Random Model	Mixed Model (ξ_A fixed, ξ_B random)
$E(MSA)$	$\sigma^2 + rl_B \frac{\sum \alpha_i^2}{(l_A-1)}$	$\sigma^2 + rl_B \sigma_A^2 + r \sigma_{AB}^2$	$\sigma^2 + rl_B \frac{\sum \alpha_i^2}{(l_A-1)} + r \sigma_{AB}^2$
$E(MSB)$	$\sigma^2 + rl_A \frac{\sum \beta_j^2}{(l_B-1)}$	$\sigma^2 + rl_A \sigma_B^2 + r \sigma_{AB}^2$	$\sigma^2 + rl_A \sigma_B^2$
$E(MSAB)$	$\sigma^2 + r \frac{\sum \sum \alpha \beta_{ij}^2}{(l_A-1)(l_B-1)}$	$\sigma^2 + r \sigma_{AB}^2$	$\sigma^2 + r \sigma_{AB}^2$
$E(MSE)$	σ^2	σ^2	σ^2

From the *EMS* given in Table 2, one can see that, if H_0 is true, $E(MSA) = E(MSE) = \sigma^2$ and both mean squares are then estimators of the residual variance. They should then take numerical values of the same order of magnitude and the ratio MSA/MSE should be close to one. If, on the contrary, H_1 is true, $E(MSA)$ is larger than $E(MSE)$ and the ratio is then expected to be larger than one and to quantify the force of the impact of factor ξ_A on the response variable. In addition, when the design is balanced, one can also show that the SS_f 's of the analysis of variance decomposition are independent of each other and that each of them follows a chi-squared distribution with df_f degrees of freedom.

These ingredients justify the construction of the F statistic to test the significance of each effect. The null hypothesis will be rejected if the p-value $P(F_{df_f, df_E} > MS_f/MSE)$ is low enough, where F_{df_f, df_E} is an F random variable with df_f and df_E degrees of freedom.

When the effect of a factor in an ANOVA model is significant, the ANOVA analysis will usually be complemented by finer tests to define which levels of the factor are different from the others. *Multiple test* and *contrast* procedures exist for this purpose. They are mostly adaptations of the t and F tests discussed above but need to be adapted to the fact that, when many tests are performed, multiplicity corrections have to be applied to the α significance level of the tests in order to control the risk of false positives (see e.g. [2]).

2.5.4. ANOVA model including random factors

When the ANOVA 2 model includes random factors, the approach is somewhat different. If the two factors are random, ANOVA 2 is called a *random or variance components* model and, if only one factor is random and the other fixed, it becomes a *mixed anova* model.

For these models, the model effects are interpreted differently.

For example, if ξ_A is random, the values α_i are interpreted as realizations of a random variable and express the variability between levels of the factors (e.g. days, operators, ... in an analytical method validation study). They are no longer fixed parameters representing the change in the value of the overall mean of the response when the level of A changes. The α_i 's are assumed to be independent and drawn from a normal distribution $N(0, \sigma_A^2)$. The same is assumed for ξ_B if both factors are random. The interaction is automatically a random effect when one of the two factors is random.

For these random and mixed models, the ANOVA decomposition, the degrees of freedom and the mean squares in the ANOVA table remain the same as for the fixed model but the *EMS*'s change, as summarized in Table 2.

The statistical hypotheses tested and the related F-test statistics change also. For example, the following hypotheses are used when testing the significance of the main effect of the random factor ξ_A :

$$H_0 : \sigma_A^2 = 0 \text{ against } H_1 : \sigma_A^2 > 0$$

The EMS table illustrates, thus, an important point with respect to ANOVA: when the model contains random effects, some F statistics for testing the significance of model effects do not have MSE in their denominator any more, but rather other mean squares terms. The chosen ratio should be equal to 1 under H_0 , which means that, in the random model, for example, the test statistic for the effect of ξ_A is defined as $F_{obs} = MSA/MSAB$.

2.5.5. Nested ANOVA

When dealing with two factor experiments, the nested ANOVA 2 model is also worth mentioning. This model is used when the levels tested for factor ξ_B are different when the level of factor ξ_A changes. In chemistry, such designs are of special interest in method validation studies that study the embedded effect of several factors on measurement uncertainty e.g. different machines and operators nested within different laboratories.

The nested ANOVA 2 model takes the following form:

$$y_{ijh} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijh}$$

where $\beta_{j(i)}$ is the effect of level j of factor ξ_B on the response when $\xi_A = i$. Details about the construction of the ANOVA table and related tests for this model may be found in [2], for example. A detailed example of a nested ANOVA 2 model with two random factors will be presented in Section 3.1.

2.5.6. General remarks about ANOVA in balanced and unbalanced designs

The ANOVA principles illustrated above in the case of the balanced crossed ANOVA 2 model can be generalized to many other possible balanced designs. However, detailed ANOVA variance decomposition equations and related expected mean squares become more complex when the number of factors increases. General algorithms, derived over the years from the original paper by Henderson [39], are available to numerically derive the ANOVA table and the expected mean squares equations. They are based on the fact that the ANOVA sum of squares can be expressed as quadratic forms, $\mathbf{y}: SS_f = \mathbf{y}'\mathbf{Q}_f\mathbf{y}$, and one can apply general matrix formulae to derive their expectation (see [40], [41], [42]). Some old references also propose manual procedures for deriving the ANOVA decomposition and EMSs in the balanced case (see e.g. [2]).

When the design is unbalanced, things become more tricky and the users of ANOVA methods should really be aware that, for most models, the ANOVA decomposition becomes questionable. The main problem with the ANOVA approach is that the partitioning of total sum of squares into effects sum of squares is no longer straightforward. In most cases, there is no unique solution and the calculated sum of squares are neither independent nor (exactly) chi-squared distributed. If one takes for example the crossed ANOVA 2 model, Equation (1) always remains valid, but Equation (2) no longer holds for unbalanced designs. The derivation of the expected mean squares nevertheless remains possible, as explained in [43] or [44]. In addition, one must realize that the various means appearing in the ANOVA equations (e.g. $\bar{y}_i$ and $\bar{y}_{.j}$ in Equation (2)) are no longer unbiased estimators of response means for the corresponding levels of the factors of interest ($\mu + \alpha_i$ and $\mu + \beta_j$). This is illustrated in [37] and in Section 3.3. The problem of unbalanced designs is the subject of a very abundant literature, especially for random and mixed models (see e.g. [3] or [43]).

Thus, nowadays, except in certain special cases, it is recommended to avoid using the traditional ANOVA-based inferential and variance components estimation methods for unbalanced designs. Instead, one should rely directly on the more elegant general or mixed linear model solutions detailed below.

2.5.7. Link between ANOVA and GLMs in fixed factor designs

The link between traditional ANOVA and the general linear model is quite straightforward when all factors in the ANOVA model are fixed and the design is balanced. The GLM makes it possible to recover most of the results provided by classical balanced ANOVA and offers a framework for its generalization to unbalanced cases. When, on the other hand, the design is unbalanced and contains random factors, the use of mixed linear models, presented in Section 2.6, is recommended.

The first step to transforming a fixed factor ANOVA model into a GLM consists of writing it as a regression-type equation and building a model matrix through an adequate coding of the design factors, as explained in Section 2.3. The GLM parameter vector β can then be estimated simply with the general OLS formula. For example, for the balanced fixed ANOVA 2 model, the structure of the model matrix $\mathbf{X}$ is illustrated in Appendix 1 and one can show that the estimators of the model parameters take the following

form when a sum-to-zero coding is used in the construction of $\mathbf{X} = (\mathbf{1} \mid \mathbf{X}_A \mid \mathbf{X}_B \mid \mathbf{X}_{AB})$:

$$\begin{aligned}\hat{\mu} &= \bar{y}_{..} & \hat{\alpha}_i &= \bar{y}_{i.} - \bar{y}_{..} \\ \hat{\beta}_j &= \bar{y}_{.j} - \bar{y}_{..} & \hat{\alpha}\hat{\beta}_{ij} &= \bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}\end{aligned}$$

and the following constraints are verified for these estimators:

$$\sum_{i=1}^{l_A} \hat{\alpha}_i = 0, \sum_{j=1}^{l_B} \hat{\beta}_j = 0, \sum_{i=1}^{l_A} \hat{\alpha}\hat{\beta}_{ij} = 0 \quad \forall j \quad \text{and} \quad \sum_{j=1}^{l_B} \hat{\alpha}\hat{\beta}_{ij} = 0 \quad \forall i.$$

In addition, the blocks $\mathbf{1}$, $\mathbf{X}_A$, $\mathbf{X}_B$ and $\mathbf{X}_{AB}$ are orthogonal to each other which is a key property in recent ASCA type methods [37].

Traditional ANOVA tables are then built in two steps in a GLM framework. First, SST is decomposed into SSM and SSE and the general ANOVA table is calculated (see Table 3). This decomposition is valid for all general linear models, whether the design is balanced or not.

Table 3: General ANOVA table

Effect	Degrees of freedom	Sum of Squares	Mean Squares	F-ratio
<i>Model</i>	$p - 1$	SSM	MSM	MSM/MSE
<i>Error</i>	$n - p$	SSE	MSE	
<i>Total</i>	$n - 1$	SST		

The *F-Test* provided in this table is the test of global model significance, which should be used before interpreting any particular parameter of the model. This test, which was presented in Section 2.4.3, makes it possible to check if the model explains some variation of the response.

Then, in a second table, SSM is decomposed into SSA , SSB and $SSAB$ to obtain the three first lines of the ANOVA table shown in Table 1. In the GLM framework, this is straightforward and does not require the elaboration of an analysis of variance equation. Rather, the principle of *F*-tests for nested models, which was also discussed in 2.4.3, should simply be applied to the model effect of interest f . Each effect f is in turn removed from the model and SSE_0 of this restricted model calculated. SS_f is defined as $SSE_0 - SSE_1$, where SSE_1 is the error sum of squares of the full GLM. In the balanced case, it can be shown that this will give the same result as the ANOVA decomposition given in Equation (2). The rest of each line of the ANOVA table is a natural consequence of the definition of the *F-test* for embedded models: the difference in the number of parameters between the full and restricted model is df_f , $MS_f = SS_f/df_f$ and the *F-Statistic* = MS_f/MSE .

This approach has the advantage of being applicable to non-balanced designs, with the only consequence being that the analysis of variance equation is no longer valid. This means that, in most cases, the effect sums of squares will no longer sum to SSM or be independent of each other. The sums of squares calculated with this method are called *Type 3 sums of squares* instead of simple *SS*. The statistical literature and software also propose the less commonly used type 1, 2 and 4 *SS*, which are not discussed in this text but are all equivalent when the design is balanced. It can be shown that a general formula can be used to derive SS_f , rather than estimating all restricted models. If $\mathbf{L}$ is the $(df_f \times p)$ contrast matrix to be applied to $\boldsymbol{\beta}$ to extract $\boldsymbol{\beta}_f$, SS_f can be calculated as $\hat{\boldsymbol{\beta}}' \mathbf{L}' (\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{L}')^{-1} \mathbf{L} \hat{\boldsymbol{\beta}}$.

2.5.8. Estimating variance components in an ANOVA framework

When the ANOVA model contains random effects, the Expected Mean Squares also provide a way to derive estimators of the random effects. They are obtained in three steps: (1) for each effect of the ANOVA decomposition, the EMS is calculated as a function of the model variance components, (2) the obtained system of equation is then solved in order to express the variance components as functions of EMSs and (3) EMSs are replaced by corresponding mean squares numerical values to obtain estimated variance components. In the balanced case, these estimators can be derived analytically in an explicit

algebraical form. For example, for the random ANOVA 2 model, the estimators for the variance components can be derived based on the second column of Table 3:

$$\begin{aligned}\hat{\sigma}_A^2 &= (MSA - MSAB)/rl_B & \hat{\sigma}_B^2 &= (MSB - MSAB)/rl_A \\ \hat{\sigma}_{AB}^2 &= (MSAB - MSE)/r & \hat{\sigma}^2 &= MSE\end{aligned}$$

In chemometrics, such estimators are very useful in repeatability/reproducibility studies and many other applications where variance components are estimated for measurement methods. These variance components estimators obtained from the ANOVA table have three advantages. First, they are unbiased, second, they are equivalent to the restricted maximum likelihood estimators (presented in the next section) and third, they can be expressed with simple analytical formulae. On the other hand, their use becomes complicated in most unbalanced designs, which is a clear disadvantage. Furthermore, they may sometimes take negative values when the variance being estimated is low or null (e.g. in such a case, it is possible that $MSA < MSAB$).

2.6. The Linear Mixed Model (LMM)

2.6.1. Goal and history

The class of linear mixed models (LLMs) provides a framework for defining very general linear models with both fixed and random factors, estimating their parameters and interpreting their results with inference and prediction tools. In the statistical world, LMMs have now taken over traditional Anova models because of their generality and their ability to deal with unbalanced designs and more advanced model equations and covariance structures. Although early versions of this model were introduced very early by Fisher and then Henderson [39], the main references on the subject are attributed to Searle ([3], [45], [46], [22], [10]).

In chemometrics, LMMs are not well-known, despite their usefulness in many situations, such as variance components analysis in analytical chemistry, experiments organized in blocks, repeated measure or longitudinal studies and the increasingly popular analysis of experimental data from the life sciences (e.g. -omics studies).

Compared to general linear and ANOVA models, linear mixed models come hand in hand with many technical estimation and inference problems, which is beyond the scope of this paper. This section will thus provide only general principles and results about this family of models, restricting its presentation to essential concepts and formulae. These general results will be illustrated in detail for a variance components model and a mixed model in Section 3.2 and 3.3. The reader is invited to consult related literature to develop his knowledge in this vast field.

2.6.2. The LMM equation

In matrix notation, the linear mixed model is usually written as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \quad (3)$$

This equation contains three blocks. The fixed effects are included in $\mathbf{X}\boldsymbol{\beta}$, where the model matrix $\mathbf{X}$ is built exactly as in the case of general linear models on the basis of the F effects included in the model equation (main effects, interactions effects etc..) : $\mathbf{X} = (\mathbf{1} \mid \mathbf{X}_1 \mid \mathbf{X}_2 \mid \dots \mid \mathbf{X}_F)$.

The random effects are included in $\mathbf{Z}\boldsymbol{\gamma}$ and the error term, $\boldsymbol{\epsilon}$, represents all the random variation in the response that is not included in the other parts of the model. The $(n \times q)$ model matrix $\mathbf{Z}$ contains the values of the variables related to the R random effects of the model, using dummy coding for the (random) effects containing categorical factors $\mathbf{Z} = (\mathbf{Z}_1 \mid \mathbf{Z}_2 \mid \dots \mid \mathbf{Z}_R)$. The vector $\boldsymbol{\gamma}$ is not a classical vector of parameters but rather a vector of random variables associated with each random effect.

The general assumptions for the random parts of the model are $\boldsymbol{\gamma} \sim N(0, \mathbf{G})$ and $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma})$, where $\mathbf{G}$ and $\boldsymbol{\Sigma}$ are, respectively, the $(q \times q)$ and $(n \times n)$ variance-covariance matrices for $\boldsymbol{\gamma}$ and $\boldsymbol{\epsilon}$. In chemometrics applications, simpler hypotheses are usually assumed for the random effects, so we will restrict the following explanations to these cases. First, the pure error terms ϵ_i are usually assumed to be independent with a common variance, just like in the general linear model : $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. Then, for each of the R random

effects included in the model (which corresponds to a block of q_j columns $\mathbf{Z}_j$ in $\mathbf{Z}$), the elements in the corresponding vector of random terms $\boldsymbol{\gamma}_j$ are also considered as independent to each other with a common variance: $\boldsymbol{\gamma}_j \sim N(0, \sigma_j^2 \mathbf{I}_{q_j})$.

Equation 3 is sometimes called the *conditional* form of the linear mixed model in the sense that $\mathbf{y}$ is expressed conditionally with respect to the values of the random terms $\boldsymbol{\gamma}$. The *marginal* form of this model integrates all of the random parts of the model into a common variance-covariance matrix $\mathbf{V}$ and is written as

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}),$$

where $\mathbf{V}$ is defined in its generic form and simplified form as:

$$\mathbf{V} = \text{var}(\mathbf{y}) = \text{var}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}) = \mathbf{Z}\text{var}(\boldsymbol{\alpha})\mathbf{Z}' + \text{var}(\boldsymbol{\epsilon}) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \boldsymbol{\Sigma} = \sum_{j=1}^R \sigma_j^2 \mathbf{Z}_j \mathbf{Z}_j' + \sigma^2 \mathbf{I}_n \quad (4)$$

This marginal distribution makes it possible to see mixed models in a different way than ANOVA models. The term $\mathbf{X}\boldsymbol{\beta}$ represents the mean behavior of $\mathbf{y}$ for the values taken by the fixed effects in the data set, while the sum of the random effects is summarized in the variance-covariance matrix $\mathbf{V}$. This matrix is no longer diagonal and includes some covariances between observations. These covariances come from the random effects, which express the fact that some observations share more in common with each other than with the other observations. For example, two repeated measures issued from the same laboratory are more likely to be correlated (because of the random laboratory effect), whereas two measures taken in different laboratories should be independent.

In Section 3.2, this model will be illustrated with an interlaboratory study that investigates the effect of two nested random effects (the laboratory and the measurement devices within each laboratory) on the variability of a measurement method. In Section 3.3, a sensory study is used to illustrate the two-way mixed model. Appendix sections 6 and 7 also demonstrate the format of the different blocks of the LMM matrices for these two models.

2.6.3. Estimation of LMM parameters

The model fitting process for the LMM aims to estimate all of the model parameters, i.e. the parameters associated with the fixed effects $\boldsymbol{\beta}$, the variance components ($\sigma_1^2, \sigma_2^2, \dots, \sigma_R^2$) and the residual variance σ^2 . Unfortunately, because of the presence of the variance components in the model, OLS estimation can no longer be applied in this context. The well-known maximum likelihood estimation approach is, however, an adequate alternative. The log-likelihood function for the linear mixed model based on the multivariate normal marginal distribution of $\mathbf{y}$ takes the following form:

$$l = \log(L) = C - \frac{1}{2}[\log|\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})], \quad (5)$$

where C is a constant term.

Estimators of the model parameters can be obtained by finding the values of the model parameters that maximize this log likelihood function. However, this approach has three drawbacks: global estimation is a tedious numerical task and estimators of the resulting variance components are biased and may be negative. It is thus recommended to estimate LMM parameters in two steps. First, the variance parameters included in $\mathbf{V}$ are estimated using *Restricted (or Residual) Maximum Likelihood Estimation (REML)*. The restricted maximum likelihood function is a reformulation of the likelihood for a linear combination of the responses such that the fixed parameters are no longer involved in the function. The estimators of the variance components are obtained using an iterative algorithm because, in general, no analytical solution exists for the optimization criteria. It can be shown that these estimators are identical to the estimators obtained using the formulae for expected mean squares in the balanced case. Moreover, to avoid the problem of potential negativity encountered in ANOVA estimation, REML estimation is usually performed under the constraint that the variance components have to be positive. The second step is the estimation of the fixed

parameters β . These parameters are estimated by maximising the classical likelihood, where the variance V is considered to be known and equal to $\hat{V}$. This estimator is given by:

$$\hat{\beta} = (X' \hat{V}^{-1} X)^{-1} X' \hat{V}^{-1} y. \quad (6)$$

Note that $\hat{\beta}$ can be viewed as a Weighted Least Squares (WLS) estimator of the fixed effects with weights $\hat{V}^{-1}$. $\hat{\beta}$ is called the *Empirical Best Linear Unbiased Estimators* (EBLUEs) of the fixed effects, where the first E indicates that V is unknown and needs to be estimated.

Two other quantities are of interest in the model fit. First, obtaining values for the (unknown) random effect variables γ provides information about how the response changes when passing from one realization of the random factors to another. Such values can not properly be called *estimators* because these are random variables but their is rather called *predictions* and can be derived with the following formula:

$$\hat{\gamma} = \hat{G} Z' \hat{V}^{-1} (y - X \hat{\beta}). \quad (7)$$

These predictors are called *Empirical Best Linear Unbiased Predictors* (EBLUPs) and are also often characterized as *shrinkage predictors* because they take values closer to zero than the parameter estimators obtained when these effects are treated as fixed in the model. This will be illustrated on Case Studies 2 and 3.

Finally, the estimators of conditional errors, called the *conditional residuals*, are simply: $e = y - X \hat{\beta} - Z \hat{\alpha}$.

Another way of deriving equivalent estimators for β and γ is very frequently encountered in the literature and deserves to be quoted. Indeed, they may be obtained as a solution to the so called *Henderson's Mixed Model Equations* (MME) ([39, 47]):

$$\begin{bmatrix} X' \Sigma^{-1} X & X' \Sigma^{-1} Z \\ Z' \Sigma^{-1} X & Z' \Sigma^{-1} Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} X' \Sigma^{-1} y \\ Z' \Sigma^{-1} y \end{bmatrix} \quad (8)$$

Several other methods for estimating the variance components of LMMs, as well as for implementing efficient algorithms for numerically calculating model parameters, are discussed and compared in the literature. See for example [43, 10, 47].

2.6.4. Statistical inference for linear mixed models

For linear mixed models, statistical inference and testing are not easy tasks because no exact solutions exist, not even for balanced designs. Moreover, statistical results available in maximum likelihood theory are only valid in large samples. For different software, the available inference methods may vary, and the results generated using the same method may sometimes be different. This section presents the most important results that are of practical interest to the experimenter.

First, to test the significance of fixed effects and parameters in the model, the F -test and t -test approaches presented in the GLM framework have been adapted in a transparent way for the practitioner.

The formula for the F -statistic to test the significance of a fixed effect f in the model is the following :

$$\hat{\beta}' L_f' (L_f (X' \hat{V}^{-1} X)^{-1} L_f')^{-1} L_f \hat{\beta} / p_f$$

where L_f is the contrast matrix selecting the parameters in $\hat{\beta}$ linked to the effect tested and p_f the corresponding number of parameters or the rank of L_f . This statistic asymptotically follows an F distribution with p_f and d degrees of freedom. The Satterthwaite's method [48], and Kenward and Roger's method (KR) [49] are two well-known solutions for calculating the degrees of freedom d [50, 51] for the denominator i.e. linked to matrix $(L_f (X' \hat{V}^{-1} X)^{-1} L_f')$. These two procedures handle the multiple random parts of the model and unbalanced designs. The KR approach is recommended for fixed effects, as it also accounts for uncertainty in V .

The t -statistic takes the same form as it does for the GLM, but the standard errors for the parameters are computed with asymptotic formulae (explained in Appendix 3) and the number of degrees of freedom is

calculated with the Satterthwaite or KR approach. Confidence intervals for a single fixed parameter β_j can also be derived in the same spirit.

The likelihood ratio test is another standard solution for hypothesis testing in mixed models and has the advantage of being applicable to both fixed effects and random components. This test makes it possible to compare two nested models and is based on the log likelihood test ratio (LLR) test statistic: $2\ln(L_{H_1}/L_{H_0}) = 2(\ln(L_{H_1}) - \ln(L_{H_0}))$, where L_{H_1} and L_{H_0} are the values of the likelihood for the two nested models, with the tested effect(s) included (under H_1) and excluded (under H_0). For fixed effects, maximum likelihood estimates must be used in the test statistic and, for variance parameters, the REML likelihood values [19]. When testing for a fixed effect, the LLR test statistic asymptotically follows, under H_0 , a χ_q^2 distribution, where q is the difference in the number of parameters between the two models. This approach is frequently used in this context, but also produces non-conservative p-values for moderate to small sample sizes [52]. On the other hand, when testing for the positiveness of variance parameters, H_0 lies at the boundary of the parameter space, meaning that standard asymptotic theory no longer holds. Thus, when the statistic is near the significance threshold, results should be interpreted with caution. It has been demonstrated [53] that, under standard conditions, the test statistic for testing the nullity of one variance component can be well approximated by a mixture of χ^2 distributions under H_0 : $0.5\chi_0^2 + 0.5\chi_1^2$.

Some software also propose a Wald test for testing the significance of a single variance component σ_j^2 . This test is a Z test based on the asymptotic normality of (RE)ML estimators, and the test statistic takes the form $z_{obs} = \hat{\sigma}_j^2 / s(\hat{\sigma}_j^2)$, where the computation of the asymptotic standard error $s(\hat{\sigma}_j^2)$ is based on the Fisher information matrix described in Appendix 3. This approach is not recommended for small samples, as it assumes that the sampling distribution of $\hat{\sigma}_j^2$ is symmetric, which is not the case.

Calculating confidence intervals for a variance component σ_j^2 in a linear mixed model is also a subject of interest. A first approach adopted by statistical software relies on the asymptotic normality property of the ML estimators, as does the Wald test. The confidence interval takes the following form: $\hat{\sigma}_j^2 \pm z_{1-\alpha/2} s(\hat{\sigma}_j^2)$. One drawback of this confidence interval is that it may contain negative values and may be too narrow on the right-hand side. An alternative solution relies on the fact that variance component estimators approximately follow a chi-squared distribution. A confidence interval is then proposed with the following classical formula:

$$\left[\frac{d\hat{\sigma}_j^2}{\chi_{d,1-\alpha/2}^2}, \frac{d\hat{\sigma}_j^2}{\chi_{d,\alpha/2}^2} \right]$$

where d is the number of degrees of freedom obtained using the generalized Satterthwaite method [29, 54, 48] : $d = 2\hat{\sigma}_j^4 / \text{Var}(\hat{\sigma}_j^2)$. Note that a one-sided confidence interval is usually preferred in practice, as upper bounds are more meaningful for variance components.

All of the methods presented above are in most cases only valid in large samples. An alternative solution for significance tests, which can be applied for both fixed effects and variance parameters, consists of determining the full and null distributions of the (R)LLR test statistic using a parametric bootstrap approach (see [55]). Confidence intervals can also be derived with related techniques. Bayesian approaches are also often chosen as an alternative solution [56] .

2.6.5. Prediction and Tolerance intervals

Like for general linear models, prediction and tolerance intervals for the response can be derived for linear mixed models. For a given set of values of the model predictors $\mathbf{x}'_* = (1, x_{*1}, x_{*2}, \dots, x_{*(p-1)})$, the (marginal) predicted response is simply calculated as $\hat{y}_* = \mathbf{x}'_* \hat{\beta}$, using only the fixed part of the model. The uncertainty of prediction is then derived as the square root of the sum of two values: first, the variance of estimation of the fixed effects $\text{var}(\mathbf{x}'_* \hat{\beta})$ and second, the marginal variance of the response $\hat{\sigma}_y^2 = \sum_{j=1}^R \hat{\sigma}_j^2 + \hat{\sigma}^2$ based on all variance components. According to the confidence levels of the interval, the student distribution it still used but the degrees of freedom are calculated with the generalized Satterthwaite method (see [29]).

3. Case studies

This section presents three case studies aimed at illustrating most of the concepts discussed in the previous sections. The data, the R and SAS code and the resulting statistical analyses are available from the authors and in a Github folder referred to in Section 4.

The first case study comes from an industrial experiment and deals with the estimation and interpretation of a response surface model including two (fixed) quantitative factors and one categorical one. The second case study illustrates how a two-factor nested random model can be applied to model and analyze data from an interlaboratory repeatability/reproducibility experiment. The last case study is issued from a sensory experiment and illustrates a mixed model based on a balanced crossed two-way factorial design.

The three designs discussed here are balanced designs so that it is possible to easily apply and compare all methods presented in the theory sections. For unbalanced designs, ANOVA (through its GLM generalization) remains applicable for fixed models, as advocated above, but for mixed or random models, ANOVA is not recommended (LMM should in most cases be preferred). The consequence of unbalanced designs will be briefly illustrated in the supplementary material.

3.1. Case study 1: Paper bleaching DOE experiment

3.1.1. Context, experimental design and data

Paper bleaching is the chemical processing of wood pulp or recycled paper for the purpose of lightening its color and whitening the pulp, producing paper with a whiteness adapted to the final usage of the paper. Nowadays, this long-standing process is still evolving, thanks to efforts to minimize its environmental impact and to reduce its dependence on chlorine bleaching formulae [57].

The data in this first case study come from a real industrial experiment aimed at analyzing the effect of three formulation factors on the brightness of the produced paper (expressed in ISO degrees). The ultimate goal of the experiment is to optimize the bleaching formula. The three factors of interest are the amount of hydrogen peroxide (the most commonly used bleaching agent), the type of alkali and the quantity of alkali used in combination with H_2O_2 .

The effect of the three factors was studied with a 3^3 full factorial design with 3 levels for each factor, plus 2 replicates at the center point for each alkali. The total number of runs is thus $n = 33$, and the design matrix ξ is a (33×3) matrix.

Figure 5: Case study 1: Experimental data with simple polynomial interpolation between experimental points

The data issued from this experiment are presented in Figure 5, which can already generate some interpretations. The replicates show good experimental repeatability. The highest brightness is obtained

for Alkali A3 for fixed values of the other two factors. The effects of the two quantitative factors are very similar from one alkali to another. A strong interaction effect is observed between the quantity of the alkali and the amount of hydrogen peroxide, as well as a light curvature effect. All of these results confirm that a second order response surface model would probably be appropriate for modeling the phenomenon.

3.1.2. Model definition and estimation

The full 3^3 factorial design allows the inclusion of a large set of possible effects in the model, but we will restrict the statistical analysis here to a full second-order polynomial model, i.e. a model containing a constant term, all first order terms (the main effects) and all second-order interaction and quadratic effects. If the experiment only involves quantitative factors, this model is called a *response surface model* in the DOE literature. However, because a categorical factor is also involved here, the model to be fitted is an extension of both the surface model and the analysis of covariance model.

There are several potential ways to write the model equation without ambiguity, but there is no universal formulation. Thus, we will write it as follows:

$$y_{ij} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \alpha_j + \alpha \beta_{1j} x_{i1} + \alpha \beta_{2j} x_{i2} + \epsilon_{ij}$$

where i is an index that denotes a combination of values for the two quantitative factors x_1 and x_2 , among the $9+2$ possible combinations, and j is the index for the level of the categorical factor. y_{ij} is the observed response for alkali j and the i^{th} combination of QtH_2O_2 and $QtAlka$ levels, and ϵ_{ij} is the corresponding experimental error. In order to ensure the identifiability of the model parameters, the following constraints are imposed on the parameters associated with the categorical factor with $l = 3$ levels:

$$\sum_{j=1}^l \alpha_j = 0, \sum_{j=1}^l \alpha \beta_{1j} = 0 \text{ and } \sum_{j=1}^l \alpha \beta_{2j} = 0. \quad (9)$$

Excluding the variance of the error term, this model contains $6 + 3 \times (l - 1)$ parameters to be estimated, i.e. $p = 12$ because $l = 3$. The model matrix $\mathbf{X}$ is thus of size 33×12 , and there are enough degrees of freedom available for comfortably estimating all of the parameters and validating the model.

When building the model matrix $\mathbf{X}$, the quantitative factors were centered and scaled in order to simplify their interpretation and to avoid collinearity. This transformation is written as follows for observation i of a given factor x .

$$x_{iS} = \frac{x_i - (\max_i(x_i) + \min_i(x_i))/2}{(\max_i(x_i) - \min_i(x_i))/2} \quad (10)$$

The qualitative factor was coded with sum-to-zero coding in order to satisfy the constraints written in Equation (9).

The full model matrix $\mathbf{X}$ is provided in the *R* code results. Because all model effect are fixed, the model was simply estimated with ordinary least squares (OLS).

3.1.3. Model validation

Table 4 and Figure 6 provide the main useful results for evaluating the quality of the model.

Table 4: Case study 1: Global ANOVA table

Source	Degrees of freedom	Sum of Squares	Mean Square	F-ratio	p-value
Model	11	245.10	22.282	268.2	<.0001
Error	21	1.74	0.083		
C. Total	32	246.85			

First, the global analysis of variance table and the related F -test indicate that the model is highly significant ($p < 0.0001$), with $R^2 = \frac{SSM}{SST} = 0.993$ and $RMSE = \sqrt{0.0831} = 0.289$. Second, the Y - $Yfit$ plot (Figure 6A) confirms that the model fits the data very well, and the graph of residuals $e_{ij} = (y_{ij} - \hat{y}_{ij})$

versus predicted values $\hat{y}_{ij}$ in Figure 6B does not show any particular pattern. The normal quantile plot (Figure 6C) suggests a slight but acceptable non-normality of the residuals. The model is thus considered to be valid.

Figure 6: Case study 1: (A) Y-Yfit plot, (B) residual plot and (C) residuals normal quantile plot

3.1.4. Interpretation of the model effects and related inference

Table 5 is the ANOVA table for the model effects and Table 6 presents the model coefficients.

The ANOVA table should be interpreted first. It provides a decomposition of the part of the total variation explained by the model (SSM) into the sum of squares related to each effect in the model. Since the model matrix is not fully orthogonal, type 3 sums of squares were calculated, and one can observe that their sum does not exactly equal SSM . Each effect linked to the categorical factor Alkali has two degrees of freedom because these effects correspond to two parameters in the model. This table provides, for each effect, an F -statistic to check if the effect is significant. All effects are significant except the interactions between the type and the quantity of alkali. This effect could be removed from the model to achieve greater parsimony.

Table 5: Case study 1: ANOVA table of model effects

Effect	Degrees of freedom	Sum of Squares	F-ratio	p-value
QtH2O2S	1	39.28	472.85	<.0001
QtAlkaS	1	2.78	33.43	<.0001
QtH2O2S*QtH2O2S	1	4.61	55.47	<.0001
QtH2O2S*QtAlkaS	1	86.78	1044.66	<.0001
QtAlkaS*QtAlkaS	1	52.38	630.59	<.0001
TypAlka	2	44.92	270.39	<.0001
TypAlka*QtH2O2S	2	0.68	4.07	0.0321
TypAlka*QtAlkaS	2	0.39	2.37	0.1184

Table 6 provides the estimators of the 12 model parameters, along with their standard errors, t -statistics and associated p-values. These p-values confirm the results shown in the ANOVA table: all of the parameters, except for the ones associated with the TypAlka*QtAlkaS interaction, are highly significant. Note also that, for all quantitative terms, the ANOVA F -statistics are the squares of the t -statistics, and their p-values are identical. Indeed, these tests are equivalent. For each effect containing the categorical factor, $l - 1 = 2$ parameters are provided. These parameters correspond to the first two levels of the alkali type factor, and the t -test (along with the associated p-values) evaluates if a given level behaves differently than an average fictive alkali level. For each effect, a parameter can be calculated for the third alkali level A3 based on the sum-to-zero coding. This parameter is equal to the negative sum of the parameters for the other two levels (see the constraints provided in Equation 9).

In a data analysis report, Table 6 should ideally be placed in a technical appendix and the model equation

Table 6: Case study 1: Model coefficients

Term	Estimate	Standard Error	t-ratio	p-value
Intercept	77.033	0.085	902.41	<.0001
QtH2O2S	1.477	0.068	21.75	<.0001
QtAlkaS	0.393	0.068	5.78	<.0001
QtH2O2S*QtH2O2S	-0.779	0.105	-7.45	<.0001
QtH2O2S*QtAlkaS	-2.689	0.083	-32.32	<.0001
QtAlkaS*QtAlkaS	-2.625	0.105	-25.11	<.0001
TypAlka[A1]	-0.886	0.071	-12.49	<.0001
TypAlka[A2]	-0.762	0.071	-10.75	<.0001
TypAlka[A1]*QtH2O2S	0.273	0.096	2.84	<.0001
TypAlka[A2]*QtH2O2S	-0.159	0.096	-1.65	0.1130
TypAlka[A1]*QtAlkaS	-0.01	0.096	-1.04	0.3124
TypAlka[A2]*QtAlkaS	0.209	0.096	2.17	0.0413

written in the main text. A readable way of presenting it is the following:

$$\hat{y} = 77.03 + 1.48QtH_2O_2S + 0.39QtAlkaS - 0.78QtH_2O_2S^2 - 2.63QtAlkaS^2 - 2.69QtH_2O_2S \times QtAlkaS$$

$$+ \begin{pmatrix} A1 : & -0.89 \\ A2 : & -0.76 \\ A3 : & 1.65 \end{pmatrix} + \begin{pmatrix} A1 : & 0.27 \\ A2 : & -0.16 \\ A3 : & -0.11 \end{pmatrix} QtH_2O_2S + \begin{pmatrix} A1 : & -0.1 \\ A2 : & 0.21 \\ A3 : & -0.11 \end{pmatrix} QtAlkaS \quad (11)$$

The first row of Equation 11 represents the mean behavior of the response for a (fictive) average level of the alkali factor. The two quantitative factors are expressed in their standardized form (according to Equation 10). In the second row, each model effect involving the categorical factor is giving three numbers, one for each factor level. These numbers indicate how the intercept and the parameters of the main effects (slopes) of the quantities of H_2O_2 and alkali are affected when a given alkali is chosen.

3.1.5. Model visualization

Statistical and DOE software provide several ways to visualize the results of a response surface fit. The more useful ones for this study are presented in Figure 7. The 3D plot in Figure 7A shows the behavior of the response as a function of the two quantitative factors and for alkali A3, the alkali yielding the highest brightness. The contour plot in Figure 7B provides the same information but with more precise iso-response curves. The interaction plot in Figure 7C shows the estimated mean response with respect to one factor, for different levels of a second factor and for the mean level of the third factor. In this last graph, a significant interaction is characterized by non-parallel lines.

Figure 7: Case study 1: Graphical representations of the estimated model. (A) Response surface plot, (B) Contour plot and (C) Interaction plot.

3.2. Case study 2: Interlaboratory repeatability analysis

3.2.1. Context, experimental design and data

Platelet number counting is a common biological blood analysis in hematology, and in particular in blood donation, to verify the quality of donor blood and to monitor the blood processing process. The data for the second case study are inspired by an inter-laboratory repeatability-reproducibility experiment conducted by the Belgian Red Cross to evaluate the precision of the cell count analytical devices used in their laboratories.

Twelve blood samples were extracted from a homogenized blood pool and four of them were sent to each of three laboratories inside the organization. In each laboratory, each sample was analyzed in quadruplicates in repeatability conditions, resulting in a dataset of $n = 48 = 3 \times 4 \times 4$ platelet numbers, represented in Figure 8. This design of experiment is thus a two-factor nested design with four replicates per condition. The laboratory (factor A) and the sample within the laboratory (factor B) are random factors with $l_A = 3$ and $l_B = 4$ levels respectively, and the design matrix ξ is a (48×2) matrix.

Figure 8: Case study 2 : Platelet counting data

These data show clear variability between laboratories, followed by some variability between samples within each laboratory and some non-negligible variability in terms of pure repeatability. All of the data from this fully balanced experimental design will be used to fit the model, making it possible to compare results for both ANOVA and linear mixed models.

3.2.2. Model definition in the ANOVA and LMM forms

In statistical terms, the model of interest is a *variance component model*, as the goal here is to estimate the variances corresponding to the three sources of variability in the data: the laboratory, the sample and the pure repeatability error. Such a nested experiment is very common in laboratory method validation, and it may be modeled, if the data are balanced, by two equivalent models: a *nested two-factor random ANOVA* model or a *linear mixed model (LMM)*.

The ANOVA model was discussed in Section 2.5.5 and takes the following form:

$$y_{ijh} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijh}$$

where α_i represents the random change relative to the overall mean μ associated with laboratory i and the α_i 's are assumed to be independent and drawn from a normal distribution $N(0, \sigma_A^2)$. $\beta_{j(i)}$ represents the effect of sample j in laboratory i and all $\beta_{j(i)}$'s are assumed to be independent of each other and issued from a normal distribution $N(0, \sigma_{B(A)}^2)$. Finally, ϵ_{ijh} is the random pure "error" observed by repeating the analysis of sample j in laboratory i in repeatability condition, and all ϵ_{ijh} 's are assumed to be $iN(0, \sigma^2)$.

This ANOVA model may be written in matrix LMM form as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} = \mathbf{1}\mu + \mathbf{Z}_A\boldsymbol{\gamma}_A + \mathbf{Z}_B\boldsymbol{\gamma}_B + \boldsymbol{\epsilon}$$

$\mathbf{X}$ is simply an $n \times 1$ matrix of ones corresponding to the model constant term β_0 . In this case β_0 is the overall mean μ of the response. $\mathbf{Z}$ is the $(n \times 15)$ model matrix for the random part of the model. It may be split into two blocks: $\mathbf{Z}_A$ of size $n \times 3$ for the laboratory random effect and $\mathbf{Z}_B$ of size $n \times 12$ for the sample random effect within the laboratories. The rows of the blocks $\mathbf{Z}_A$ and $\mathbf{Z}_B$ contain dummy variables to indicate to which laboratory i , and to which sample j in laboratory i , the data point corresponds. For example, for Sample 2 in Lab 3, the corresponding row of $\mathbf{Z}_A$ is $(0, 0, 1)$ and for $\mathbf{Z}_B$, it is $(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0)$.

$\boldsymbol{\gamma}_A$ and $\boldsymbol{\gamma}_B$ are (3×1) and (12×1) random effect vectors, respectively. The elements of $\boldsymbol{\gamma}_A$ are the α_i 's of the ANOVA model and the elements of $\boldsymbol{\gamma}_B$ are the $\beta_{j(i)}$'s and have the same definition. $\boldsymbol{\epsilon}$ is the $(n \times 1)$ vector containing all ϵ_{ijh} 's. Appendix 2 provides a detailed description of the $\mathbf{Z}$ matrix and of the marginal variance-covariance matrix $\mathbf{V}$ for $\mathbf{y}$.

3.2.3. ANOVA results

The traditional method, ANOVA, is presented first to analyze these data. It has the advantage, in balanced cases, of providing exact tests of significance of the model effects and explicit formulae for unbiased estimators of the variance components. This method starts with an analysis of variance equation, which takes the following form:

$$SST = \sum_{ijh} (y_{ijh} - \bar{y}_{..})^2 = SSA + SSB(A) + SSE = \sum_{ijh} (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{ijh} (\bar{y}_{j(i)} - \bar{y}_{i.})^2 + \sum_{ijh} (y_{ijh} - \bar{y}_{j(i)})^2$$

where $\bar{y}_{..}$, $\bar{y}_{i.}$ and $\bar{y}_{j(i)}$ are the arithmetic means of the responses for all data, for laboratory i and for sample j within laboratory i , respectively.

The standard ANOVA table derived from this equation is provided in Table 7.

Table 7: Case study 2: Anova table

Source	Degrees of freedom	Sum of Squares	Mean Square	F-ratio	p-value
Model	11	209 014.1	19 001.3	38.02	<.0001
Lab	2	188 172.4	94 086.2	40.63	<.0001
Sample(Lab)	9	20 841.7	2315.7	4.63	0.0004
Error	36	17 989.2	499.7		
Total	47	227 003.3			

In this table, the SS, df and MS columns are identical no matter the type of factor, fixed or random. What differs are the F -ratios provided, which are based on the expected value of the mean squares. For this model (in the balanced case), they are given by :

$$E(MSA) = rl_B\sigma_A^2 + r\sigma_{B(A)}^2 + \sigma^2, \quad E(MSB(A)) = r\sigma_{B(A)}^2 + \sigma^2, \quad E(MSE) = \sigma^2$$

where $r = 4$ and $l_B = 4$.

Following the procedure explained in Section 2.5.4, the two F -ratios are defined as $MSA/MSB(A)$ and $MSB(A)/MSE$ and can be used to test if the variance components σ_A^2 and $\sigma_{B(A)}^2$ are significantly greater than zero. The two very small p-values ($< .0001$ and 0.0004) confirm this affirmation: there is a significant variability between laboratories and between samples within laboratories. The p-value linked to the model, at the first line, tests if at least one of these two variances is significant which is indeed the case.

Next, and as explained in Section 2.5.8, the three EMSs take the form of a system of linear equations, which can be solved to derive estimators of the variance components. These estimators are obtained by dropping the expected values $E(\cdot)$ and replacing the variances σ^2 by their estimators $\hat{\sigma}^2$:

$$\hat{\sigma}_A^2 = \frac{MSA - MSB(A)}{rl_B}, \quad \hat{\sigma}_{B(A)}^2 = \frac{MSB(A) - MSE}{r}, \quad \hat{\sigma}^2 = MSE$$

The variance component estimates are summarized in Table 8, along with the corresponding standard deviations and percentages of total variance explained by each source of variability.

Table 8: Case study 2: Variance Components Estimations

Component	Var Comp Estimate $\hat{\sigma}_j^2$	% of Total variance	Standard deviation $\hat{\sigma}_j$
Lab	5735.7	85.7%	75.7
Sample(Lab)	454.0	6.8%	21.3
Residual	499.7	7.5%	22.4
Total	6689.4	100.0%	81.8

These results confirm what was observed in the raw data. The largest source of variability in the response is the difference between laboratories. The other two sources of variance, which have comparable variability, generate much less variability in the response. Note, however, that estimating the variance between laboratories with only three replicates is not very reliable.

3.2.4. LMM results: global model significance and residual graph

After estimating the model through REML, a first step, advocated in the GLM framework, is to test the global significance of the model and to evaluate the distribution of the residuals. In theory, the maximum likelihood ratio statistic can be used as a replacement for the global F -test in the GLM setting. This statistic is used to compare the null model (a model with only the constant term and the residuals) with the full model. The gain in the likelihood ratio statistic can be calculated as the difference between $2\ln L_1 = -454.6$ and $2\ln L_0 = -535.9$. The LR test statistic is thus equal to $LLR = 81.3$, which is highly significant according to the quantiles of a χ_2^2 distribution. Remind, however, that testing the nullity of variance components, the chi-squared distribution is no longer a good approximation of the test statistic sampling distribution but, in this case, the LLR is so high that one can be confident in this significance.

In order to evaluate the distribution of the residuals, the conditional residuals are usually used in linear mixed models. These residuals are defined as $\mathbf{e} = \mathbf{y} - (\mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\gamma})$ and represented in Figure 9.

Figure 9: Case study 2 : Conditional residuals plots.

Graph **A** shows no unexpected patterns in the residuals and the qq-plot **B** indicates that the residuals have an almost normal distribution, with a slight amount of kurtosis.

Note that, if the residuals had been studied in the ANOVA analysis, the same results would have been obtained.

3.2.5. LMM results: analysis of fixed effects

As this model is a variance components model, the only fixed effect in the model is the constant or intercept term μ . Software usually provide a coefficient table in the form of Table 9.

In this case, since the data are balanced, the intercept is simply the overall arithmetic mean of the data $\hat{\mu} = \bar{y}_{..}$. However, it should be noted that if the data were not balanced, the overall mean $\bar{y}_{..}$ would bias the estimation because of over- or under-representation of some laboratories or samples. The (R)MLE would, however, provide an unbiased estimation of μ .

Table 9: Case study 3: Fixed model coefficients

Term	Estimate	Standard Error	Degrees of freedom	t-ratio	p-value
Intercept	1269.19	44.27	2	28.67	0.001

A comment should also be made concerning the degrees of freedom used to conduct the significance test for this constant term. Indeed, in a general linear model, the number of degrees of freedom linked to the t statistic would be $n - p = 36$. Here the number of degrees of freedom was calculated using the Kenward-Roger procedure discussed in Section 2.6.4, and the very low value of 2 reflects the fact that most of the random portion of the model comes from the 3 laboratories and not from the data within the laboratories.

3.2.6. LMM results: analysis of the random part of the model

Two sets of parameters are of interest when analyzing the random part of the model: the random components σ_A^2 and $\sigma_{B(A)}^2$ and the random parameters α_i and $\beta_{j(i)}$.

The REML estimates of the variance components are the same as the ANOVA estimates for all of the variance components. They are provided in Table 10, along with several inferential results. These results are extracted from SAS and R software outputs and are more difficult to interpret than ANOVA results. This is because none of the procedures used are exact, and the adopted methods and results may differ from one software to another.

Table 10: Case study 2: Inference on random coefficients (Bootstrap p-values are based on 2000 simulations)

Component	$\hat{\sigma}_j^2$	$s(\hat{\sigma}_j^2)$	LLR Statistic	χ^2 p-value	Boot. p-value	df	χ^2 CI-Inf	χ^2 CI-Up
Laboratory	5735.65	5880.8	15.74	<.0001	0.0005	1.90	1521.7	266 406.0
Sample(Lab)	454.01	274.5	10.78	0.001	0.0014	5.47	182.62	2444.61
Residual	499.7	117.8				36.00	330.46	843.15

This table provides the estimations of the variance components along with their standard error calculated on the basis of the formulae provided in Appendix 3. The LR test statistic was then calculated for each variance component and two p-values provided. The first p-value is based on a χ^2 distribution with 1 degree of freedom and the second on a bootstrap procedure. These p-values confirm the high significance of the Lab and Sample variance components. In addition to these results, Appendix 4 illustrates, using bootstrap simulations, why the mixture of 2 chi-squares is better suited to approximating the distribution of this test statistic than a χ_1^2 . Nonetheless, the p-values in this case study are not very affected by this problem. Then 95% confidence intervals are also provided for the variance components which were calculated by SAS. One can observe that these intervals can be very wide when the number of degrees of freedom for the component is small.

LMM also provides EBLUP predictions for the random parameters according to Equations (7) or (8). These predictions are summarized in the graphs in Figure 10 and compared with the estimations obtained by OLS for the fixed general linear model. If the two factors were considered to be fixed in the model, their GLM estimators would simply be, in this balanced case, the sample mean for the laboratories and samples, $\hat{\alpha}_i = \bar{y}_i$ and $\hat{\beta}_{j(i)} = \bar{y}_{j(i)}$.

This graph very clearly illustrates why special estimators are necessary when random factors are introduced to a model. Indeed, one can see that all EBLUP predictions are closer to 0 (i.e. *shrunk* towards 0) compared to classical fixed estimators. This degree of shrinkage is larger for the sample effect, which is due to the fact that the variance of the sample effect is moderate compared to the variance of the residuals. For the laboratory effect, the phenomenon is less spectacular due to the high variability between laboratories. Chemometricians should thus be aware that these EBLUP estimators are better estimators than the sample means when one considers that (random) factor levels represent a sample from a larger population (e.g. three laboratories chosen from a *population* of candidate laboratories).

Figure 10: Case study 2 : Comparison of GLM OLS fixed effect estimations to LMM EBLUP random effect predictions

3.3. Case study 3: Candy Sensory Experiment

3.3.1. Context, experimental design and data

The third dataset is issued from a sensory experiment where a panel of 11 judges rated 5 different candies with respect to a series of 9 attributes: transparency, acidity, sweet taste, raspberry flavor, sugar coated texture tested with a spoon, biting strength in the mouth, hardness, elasticity in the mouth and stick to teeth in the mouth. Each candy was evaluated in triplicates by each judge, and the resulting design was a two-factor full factorial design repeated three times. No data are missing, the design is balanced and the study includes $165 = 11 \times 5 \times 3$ runs. Several names, such as *cross-over study* or *repeated measure study*, are used in the literature to refer to such a design. The particularity of this type of design is that all panel judges test all of the candies, so the data for a given judge will not be independent of each other. Such a design has the significant advantage that each judge acts as his or her own control, and candy effects can be studied within each judge. The drawback is that a more complex statistical analysis is necessary for such data: this analysis must take the particular correlation structure in the data into account.

This dataset was first published in [58], in which the authors compared the PC-ANOVA and ASCA approaches to modeling this multi-response dataset. Both approaches involved a fixed and crossed two-factor ANOVA model. This dataset is also used in [59], where the authors propose a new analytical solution for estimating the multivariate uncertainty in ASCA models using confidence ellipsoids. They also extend their results to multiple testing of effect level differences. Another work in which this dataset appears is [60], where it is used to illustrate the method LMM-PCA, an extension of ASCA to mixed models.

In this text, only one response from the experimental design will be investigated: the attribute “sweet.” The raw data are presented in Figure 11, along with the means of the three replicates for each judge. Graph A shows a great degree of variability in the data and, there is no clear difference in judge perception between candies, except for candy 1. The high degree of dispersion in the data is due to the variability between and within judges. Graph B is more interesting and is a better tool for representing repeated measure data. The profiles of the judges (shown as colored lines) are relatively parallel, suggesting that the 11 judges perceive the 5 candies similarly, even though some judges systematically rate candies higher than the others. Overall, the differences between the 5 candies are more apparent in this graph. The statistical modeling below will objectively confirm or refute these observations.

3.3.2. Model definition in the ANOVA and LMM form

The statistical model chosen for analyzing the results of this repeated measure experiment is a two cross-factor mixed model. The candy is a fixed factor and the judge is included as a random factor. This random factor makes it possible to model the relationship between data for a same judge and generate results that can be generalized to the *population* of potential consumers from which the 11 judges are issued. The modeling

Figure 11: Case study 3 : (A) Raw data and (B) averaged data for the response "sweet" of the candies dataset.

of sensory profiling is discussed in detail in [61]. In particular, the authors address the consequences of defining the judge factor as fixed or random in such studies.

The formulation of the ANOVA model and the analysis of this model were discussed in Section 2.5. The ANOVA model equation takes the following form:

$$y_{ijh} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijh}$$

where y_{ijk} represents the value of the attribute of interest (the sweetness) for candy i , judge j and replicate h . $i = 1, \dots, l_A$, where $l_A = 5$, $j = 1, \dots, l_B$, where $l_B = 11$, and $h = 1, \dots, r$, where $r = 3$. μ represents the overall mean, α_i is the (fixed) effect of candy i and α_i 's are subject to the constraint $\sum_i \alpha_i = 0$. β_j is the (random) effect of judge j , which reflects the fact that each judge has his own response "baseline" when comparing one candy to another. The β_j 's are assumed to be independent and normally distributed *i.e.* $\sim iN(0, \sigma_B^2)$. $(\alpha\beta)_{ij}$ is the (random) interaction between candy i and judge j and the $(\alpha\beta)_{ij}$'s are assumed to be $iN(0, \sigma_{AB}^2)$. The interaction in this model is used to reflect potential disagreements between certain judges with respect to some of the candies. Parallel judge profiles in Figure 11B would indicate the absence of an interaction effect in the data. Note that when an interaction between a fixed and a random term is included in an ANOVA or LM model, it is considered to be a random effect, reflecting an effect in the population of potential judges. Finally, the errors ϵ_{ijk} are assumed to be $iN(0, \sigma_\epsilon^2)$ and represent the repeatability of each judge in a repeated test of the same candy.

Similarly to the model in Section 3.2, the mixed form of the model may be written as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} = \mathbf{1}\mu + \mathbf{X}_A\boldsymbol{\beta}_A + \mathbf{Z}_B\boldsymbol{\gamma}_B + \mathbf{Z}_{AB}\boldsymbol{\gamma}_{AB} + \boldsymbol{\epsilon}$$

$\mathbf{X}_A$ is an $n \times 4$ matrix with coded candy values. $\mathbf{Z}_A$ and $\mathbf{Z}_{AB}$ are, respectively, $n \times 11$ and $n \times 55$ matrices of dummy variables. A full description of the form of these matrices and of the marginal variance-covariance matrix $\mathbf{V}$ for the response is provided in Appendix 1. This last matrix clearly illustrates the structure of dependence between observations which results from the repeated measure characteristic of this sensory study.

3.3.3. ANOVA results

Performing Analysis of Variance on these data begins with the construction of an ANOVA table for two crossed factors. The first three columns of this table (see Table 11) are calculated with the formulae given in Figure 4, regardless of the fixed or random type of the two factors present in the model. The last two columns are calculated according to the procedure explained in Section 2.5.3 and using the expected mean squares provided in Table 2. The four F Ratios, which test the significance of the model effects, are

derived as follows: $F\text{-ratio}(\text{Model}) = MSM/MSE$, $F\text{-ratio}(A) = MSA/MSAB$, $F\text{-ratio}(B) = MSB/MSE$, $F\text{-ratio}(AB) = MSAB/MSE$.

Table 11: Case study 3: Anova table

Effect	Degrees of freedom	Sum of squares	Mean Square	F-ratio	p-value
Model	54	3170.53	58.71	5.25	<.0001
A: Candies	4	1986.79	496.70	27.49	<.0001
B : Judges	10	460.89	46.09	2.55	0.0174
AxB : Candies*Judges	40	722.85	18.07	1.62	0.0264
Error	110	1229.48	11.18		
Total	164	4400.00			

As all the p-values in this table are low (lower than the conventional level 0.05), one can draw the following conclusions for the 4 tests: (1) At least one fixed parameter or random component in the model (α_i , σ_B^2 or σ_{AB}^2) is significantly different from 0. Thus, the model explains a significant portion of the overall data variability. (2) The candies effect is highly significant, which means that the judges perceive a difference in sweetness between some of the candies (some α_i 's are different from 0). (3) and (4) The two variance components σ_B^2 are σ_{AB}^2 are slightly significant. There is thus some variability in the perception of the judges and their perceptions may differ for some candies.

Note that the candies effect would have been even more significant if the judge effect had been considered fixed in the model. However, in such a case, the results would not have been generalizable to the population of judges who could have tested these candies.

The procedure explained in Section 2.5.8 makes it possible to derive estimators for the variance components. The expected mean squares in Table 2 lead to the following system of equations:

$$E(MSB) = \hat{\sigma}^2 + rl_A \hat{\sigma}_B^2, E(MSB) = \hat{\sigma}^2 + r \hat{\sigma}_{AB}^2, E(MSE) = \hat{\sigma}^2$$

whose solution is :

$$\hat{\sigma}_B^2 = \frac{MSB - MSE}{rl_A}, \hat{\sigma}_{AB}^2 = \frac{MSAB - MSE}{r}, \hat{\sigma}^2 = MSE$$

The resulting estimations are presented in Table 12.

Table 12: Case study 3: Variance Components Estimations

Component	Var Comp Estimate $\hat{\sigma}_j^2$	% of Total variance	Standard deviation $\hat{\sigma}_j$
Judges	1.87	12.1%	1.37
Judges*Candies	2.30	14.9%	1.52
Residual	11.18	72.8%	3.34
Total	15.34	100 %	3.92

This table shows that the variability between judges and the variability due to the judge-candy interaction are much smaller than the variability within judges, i.e. the simple variability of the three scores given by each judge for the same candy (repeatability).

3.3.4. LMM results: global model significance and residual graphs

The REML estimation of the full linear mixed model results in a likelihood value of $-2\ln L_1 = 894.4$, whereas the likelihood for the null model with only the constant term is given by $-2\ln L_0 = 908.9$. The LR test statistic is equal to $\chi_{obs}^2 = 14.6$, with a p-value of 0.024 based on a χ^2 with 6 degrees of freedom for the 4 fixed parameters and 2 random components. The model is thus significant according to this test.

The conditional residuals represented in Figure 12 have rather good behavior. According to the qq-plot, with the exception of three minor outliers, the normal distribution hypothesis is perfectly verified, and the residual graph does not demonstrate any worrisome patterns. The LMM results can thus be interpreted.

Figure 12: Case study 3 : Conditional residuals.

3.3.5. LMM results: analysis of fixed effects

For the fixed part of LLMs, software usually deliver two tables that are very similar to those produced with GLMs: an “ANOVA-like” table (see Table 13) and a coefficient table (see Table 14). These tables may be interpreted in the same way as for fixed models, but it is important to remember that the computations behind their generation are not completely equivalent. In particular, the standard errors of the parameters are derived from asymptotic formulae and the number of degrees of freedom for the denominator of the test statistics are calculated using formulae that depend on the different random parts of the model. The Kenward-Roger’s approach was used here. The p-values are also only approximate in small samples.

Table 13: Case study 3: Fixed effect global F-Test

Effect	Numerator df	Denominator df	F-ratio	p-value
Candies	4	40	27.49	<.0001

Table 14: Case study 3: Fixed model coefficients

Term	Estimate	Standard Error	Degrees of freedom	t-ratio	p-value
Intercept	6.70	0.53	10	12.67	<.0001
Candies1	5.83	0.66	40	8.82	<.0001
Candies2	-3.27	0.66	40	-4.94	<.0001
Candies3	-0.75	0.66	40	-1.13	0.2656
Candies4	-3.48	0.66	40	-5.26	<.0001

These tables confirm the high significance of the candy effect both for the global F -test and for the individual test on most single model parameters α_i . Remember that Candy 5 does not appear in the table because the related coefficient $\hat{\alpha}_5$ can be derived as minus the sum of the others. A better way to visualize these results is through a graph of candies means estimated from the model, along with their confidence intervals, which is shown in Figure 13. The table on the right provides the p-values for the pairwise comparisons of the five means. The p-values were calculated with a Tukey procedure that corrects the p-values for multiple testing.

It is important to notice that, because the data are balanced, the candies means estimations issued from the model correspond simply in this case to the arithmetic means : $\hat{\mu} + \hat{\alpha}_i = \bar{y}_i \dots$. If the data were unbalanced, this would not be the case : the arithmetic means would be biased estimators of the expected candies means but the (RE)ML estimators would be unbiased.

3.3.6. LMM results: analysis of the random part of the model

Table 15 presents results related to the estimation of the model variance components. As expected in the balanced case, the REML variance estimates are identical to those provided by ANOVA. The rest of the table gives inference results for these parameters, which were explained in Section 3.2.6.

Figure 13: Case study 3 : (A) Means with confidence intervals and (B) p-values of Tukey 2-2 comparisons.

The two p-values confirm the relatively low significance of the two variance components. The chi-squared confidence intervals are quite large for the judge and interaction variance components because of the low degrees of freedom and much narrow for the repeatability variance.

Table 15: Case study 3: Inference on random coefficients

Component	$\hat{\sigma}_j^2$	$s(\hat{\sigma}_j^2)$	LLR Statistic	χ^2 p-value	Boot. p-value	df	χ^2 CI-Inf	χ^2 CI-Up
Judges	1.87	1.40	4.14	0.0418	0.020	3.56	0.64	18.73
Judges*Candies	2.30	1.44	3.62	0.0570	0.026	5.11	0.90	13.45
Residual	11.18	1.51				110.00	8.72	14.84

Figure 14 presents the sorted EBLUP predictions for the random Judge parameters (A), as well as a comparison between these EBLUP predictions and the OLS estimators for a model where the Judge effect is fixed (B). There are clearly three groups of judges: judges with low, medium and high (Judge 5) perception of sweetness. Compared to the OLS values, the shrinkage of EBLUP predictions is quite high here due to the fact that the judge effect is small with respect to the residual variance.

Figure 14: Case study 3 : Random Judge EBLUP Coefficients : (A) sorted and (B) compared to OLS coefficients

4. Statistical software

Numerous commercial and freeware software exist to fit, validate and interpret linear models. The best software choice mainly depends upon the complexity of the model of interest, the software that is available in the experimenter's working environment and his/her willingness to do some coding to analyze his/her data (e.g. R, SAS or MATLAB) or his/her preference for point and click software (e.g. MINITAB, JMP,...). Some

statistical software are very general-use, offering numerous statistical procedures, whereas others may be devoted to particular applications, such as the analysis of the results of designed experiments (e.g. MODDE or Design Expert).

Most available software are well suited for regression analysis and basic balanced ANOVA models, but the choice of software must be made with care when data are unbalanced or when the model includes random factors or fixed categorical factors that must be coded in the general linear model. The preparation of this chapter was based on three software, namely SAS, R and JMP, which are able to handle all techniques discussed. Some details are given below about the procedures available in each of them, and a Github repository is also available with the code and the results for the three case studies in this text (https://github.com/bgovaerts/GLM_LMM_ANOVA). JMP is a powerful statistics software for experimenters, and it has a very attractive user interface. SAS and R are less easy to use because they are based on a programming language. SAS is still considered a high level reference in statistical computing and data science in academics and business. Nowadays, however, R has become very popular because it is open source and, in addition to being a very powerful computing language, it comes with a very wide variety of packages in statistics, bioinformatics, machine learning, etc.

In JMP [62], the *fit Y by X* platform is useful for simple linear regression (if X is continuous) or 1-way ANOVA (if X is categorical). Otherwise, the *fit model* platform is the most useful and general one. Predictors can be continuous or qualitative in this platform. Several kinds of effects can be introduced to the model through the user interface, and the standard least squares method is automatically applied when factors are not declared as random. The outputs include classical statistical tables as well as appealing functionalities for representing the model results using graphs, as is the case for response surface models, for example. Changing the attribute of a predictor from fixed to random will automatically change the model to a linear mixed model with the REML procedure, making it possible to generate most results described in this text.

In R [63], the function *lm* is the richest function to fit general linear models. It is linked to many other generic functions for extracting outputs from the model fit (residual analysis, ANOVA tables, predictions, etc.). Syntax is available to describe several types of models including main, interaction and nested effects. For mixed models, the R package *lme4* [64] is the most well-known with *lmer* as the function to fit the model. For advanced inference options, the *lmerTest* [50] or *pbkrtest* [65] packages offer additionnal functionalities.

In SAS [66], the following procedures (Proc) are of interest for linear modeling. Proc REG is a procedure that is used for linear regression with continuous predictors and Proc RSREG for response surface models (where the syntax is much more compact and some tests, like the lack-of-fit test, are provided by default). For qualitative predictors, Proc ANOVA performs ANOVA for certain specific designs (e.g. balanced designs), where factors can be both fixed and random and model effects include main, interaction or nested effects. Proc GLM should be preferred for (un)-balanced data when predictors are continuous or categorical and when the model also includes several kinds of effects, such as quadratic, interactions or nested effects. Random variables can also be used in Proc GLM, but the user should be aware that some results are still computed as if the random factors were fixed, which may be confusing. Proc MIXED is able to estimate a wide variety of linear mixed models, including all types of factors and effects, with the REML (by default) or ML method. It is able to perform most analyses discussed in Case studies 2 and 3 of this chapter, and many more.

5. Final Remarks

The aim of this chapter was to present a series of useful linear modeling methods in a single text: the general linear model (including multiple linear regression), the analysis of variance model and the mixed linear model. The specificities and contributions of each class of model were discussed with accessible language and illustrated with case studies. Moreover, precise mathematical formalism was used to describe these different classes of models. In particular, the text addressed the notion of the random factor, which is a little-known notion that is nonetheless important for a variety of analyses: the estimation of variance components in the field of measurement (e.g. analytical method validation), as well as the analysis of data coming from a design with correlation structure (e.g. split plot designs) or a design where certain factor levels are drawn from a population to which one would like to generalize results.

This chapter also discussed why it is often inappropriate to use classical ANOVA to model unbalanced data. When the factors are fixed, the general linear model offers a direct solution to the problem by providing unbiased estimates of the ANOVA model parameters and adjusted ANOVA tables using the F-Test principle for nested models. When the design contains random factors, the mixed linear model is a solution for modeling unbalanced data but is less easy to use. In addition, because most results rely on asymptotic theory, the practical implementation of inference solutions may differ from one software to another. For this reason, this chapter argues that, for balanced data, ANOVA remains a useful method, despite the fact that it is sometimes set aside by the statistical community.

Finally, it should be noted that this chapter assumes that the model response variable is continuous and follows a normal distribution whereas, in practice, the experimenter might encounter other types of responses, such as binary (e.g. success/failure experiment) or count (e.g. number of defects in a batch) responses. To model this type of variable, the generalized linear model (also abbreviated by GLM in the statistical literature!) offers an interesting extension to the normal-based models seen here. Many notions discussed in this text remain valid for generalized linear models, and the interested reader can consult e.g. references [9], [10] and [20] to learn more about them.

6. Appendix 1 : Generic equations for the two-way fixed and mixed crossed ANOVA models

This appendix includes explicit forms of some matrices used in GLMs and LMMs for the two-way ANOVA model. These matrices are provided in order to help the user clearly visualize their shapes, sizes and content. Let's suppose that we have a design for studying the effect of two (crossed) factors ξ_A and ξ_B on a response Y , where the two factors have $l_A = 3$ and $l_B = 2$ levels respectively. We are interested in the following ANOVA 2 model:

$$y_{ijh} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijh}$$

Suppose also that the design is a full factorial design with two replicates for each combination of the factors ($r = 2$ and $n = 3 \times 2 \times 2 = 12$).

If sum-to-zero or deviation coding is used for the categorical factors, the design matrix and the (fixed) General Linear Model is written as follows. Each block in the $\mathbf{X}$ matrix (separated here by a vertical line) corresponds to one of the model effects: $\mathbf{1}$, $\mathbf{X}_A$, $\mathbf{X}_B$ or $\mathbf{X}_{AB}$.

$$\xi = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 2 & 1 \\ 2 & 1 \\ 3 & 1 \\ 3 & 1 \\ 1 & 2 \\ 1 & 2 \\ 2 & 2 \\ 2 & 2 \\ 3 & 2 \\ 3 & 2 \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \alpha\beta_{11} \\ \alpha\beta_{21} \end{pmatrix} + \boldsymbol{\epsilon}$$

When sum-to-zero coding is used, the missing parameters are related to the ones included in the model according to the following formulae: $\alpha_3 = -\alpha_1 - \alpha_2$, $\beta_2 = -\beta_1$, $\alpha\beta_{31} = -\alpha\beta_{11} - \alpha\beta_{21}$, $\alpha\beta_{i2} = -\alpha\beta_{i1}$. The variance covariance matrix of $\boldsymbol{\epsilon}$ is $\sigma^2 \mathbf{I}_n$.

Now, if we suppose that the model is a mixed model with fixed factor ξ_A and random factor ξ_B , it is written as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \alpha\beta_{11} \\ \alpha\beta_{21} \\ \alpha\beta_{31} \\ \alpha\beta_{12} \\ \alpha\beta_{22} \\ \alpha\beta_{32} \end{pmatrix} + \boldsymbol{\epsilon}$$

In this model, $\mathbf{X}$ contains two blocks, $\mathbf{1}$ and $\mathbf{X}_A$, corresponding, respectively, to the overall mean effect and the fixed effect of ξ_a . $\mathbf{Z}$ contains two blocks, $\mathbf{Z}_B$ and $\mathbf{Z}_{AB}$, which correspond to the random main and interaction effects of ξ_B and $\xi_A \times \xi_B$. μ and the α_i 's are fixed parameters and the β_j 's and $\alpha\beta_{ij}$'s are random variables with the respective distributions $\beta_j \sim iN(0, \sigma_B^2)$ and $\alpha\beta_{ij} \sim iN(0, \sigma_{AB}^2)$. As for the GLM, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.

If we want to write the marginal model of $\mathbf{y}$, the variance-covariance matrix $\mathbf{V}$ is an $n \times n$ matrix organized into the following blocks :

$$\mathbf{V} = \sigma^2 \mathbf{I}_n + \sigma_A^2 \mathbf{Z}_A \mathbf{Z}_A' + \sigma_{AB}^2 \mathbf{Z}_{AB} \mathbf{Z}_{AB}'$$

$$\mathbf{V} = \sigma^2 \mathbf{I}_n + \begin{pmatrix} \sigma_B^2 \mathbf{J}_6 & \mathbf{0} \\ \mathbf{0} & \sigma_B^2 \mathbf{J}_6 \end{pmatrix} + \begin{pmatrix} \sigma_{AB}^2 \mathbf{J}_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{AB}^2 \mathbf{J}_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{AB}^2 \mathbf{J}_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{AB}^2 \mathbf{J}_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{AB}^2 \mathbf{J}_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{AB}^2 \mathbf{J}_2 \end{pmatrix}$$

where $\mathbf{J}_i$ is an $(i \times i)$ matrix containing ones. An alternative but equivalent formulation is the following:

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{11} \end{pmatrix}$$

where

$$\mathbf{V}_{11} = \begin{pmatrix} \sigma^2 + \sigma_B^2 + \sigma_{AB}^2 & \sigma_B^2 + \sigma_{AB}^2 & \sigma_B^2 & \sigma_B^2 & \sigma_B^2 & \sigma_B^2 \\ \sigma_B^2 + \sigma_{AB}^2 & \sigma^2 + \sigma_B^2 + \sigma_{AB}^2 & \sigma_B^2 & \sigma_B^2 & \sigma_B^2 & \sigma_B^2 \\ \sigma_B^2 & \sigma_B^2 & \sigma^2 + \sigma_B^2 + \sigma_{AB}^2 & \sigma_B^2 + \sigma_{AB}^2 & \sigma_B^2 & \sigma_B^2 \\ \sigma_B^2 & \sigma_B^2 & \sigma_B^2 + \sigma_{AB}^2 & \sigma^2 + \sigma_B^2 + \sigma_{AB}^2 & \sigma_B^2 & \sigma_B^2 \\ \sigma_B^2 & \sigma_B^2 & \sigma_B^2 & \sigma_B^2 & \sigma^2 + \sigma_B^2 + \sigma_{AB}^2 & \sigma_B^2 + \sigma_{AB}^2 \\ \sigma_B^2 & \sigma_B^2 & \sigma_B^2 & \sigma_B^2 & \sigma_B^2 + \sigma_{AB}^2 & \sigma^2 + \sigma_B^2 + \sigma_{AB}^2 \end{pmatrix}$$

7. Appendix 2 : Variance-covariance matrix for a nested two-factor random model

This appendix describes the structure of the marginal variance-covariance matrix $\mathbf{V}$ for the response in the model in case study 2. Without loss of generality, and for the sake of simplicity, we assume here that $l_A = 3$, $l_{B(A)} = 2$ and $r = 2$. n is equal to 12.

$$\mathbf{V} = \sigma^2 \mathbf{I}_n + \sigma_A^2 \mathbf{Z}_A \mathbf{Z}_A' + \sigma_{B(A)}^2 \mathbf{Z}_{B(A)} \mathbf{Z}_{B(A)}'$$

We have

$$\mathbf{Z}_A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \mathbf{Z}_{B(A)} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$\mathbf{V}$ thus takes the following form:

$$\mathbf{V} = \sigma^2 \mathbf{I}_n + \begin{pmatrix} \sigma_A^2 \mathbf{J}_4 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_A^2 \mathbf{J}_4 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_A^2 \mathbf{J}_4 \end{pmatrix} + \begin{pmatrix} \sigma_{B(A)}^2 \mathbf{J}_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{B(A)}^2 \mathbf{J}_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{B(A)}^2 \mathbf{J}_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{B(A)}^2 \mathbf{J}_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{B(A)}^2 \mathbf{J}_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{B(A)}^2 \mathbf{J}_2 \end{pmatrix}$$

where $\mathbf{J}_i$ is an $(i \times i)$ matrix containing ones.

$\mathbf{V}$ can also be written as a block diagonal matrix as follows :

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V}_{11} \end{pmatrix}$$

where

$$\mathbf{V}_{11} = \begin{pmatrix} \sigma^2 + \sigma_A^2 + \sigma_{B(A)}^2 & \sigma_A^2 + \sigma_{B(A)}^2 & \sigma_A^2 & \sigma_A^2 \\ \sigma_A^2 + \sigma_{B(A)}^2 & \sigma^2 + \sigma_A^2 + \sigma_{B(A)}^2 & \sigma_A^2 & \sigma_A^2 \\ \sigma_A^2 & \sigma_A^2 & \sigma^2 + \sigma_A^2 + \sigma_{B(A)}^2 & \sigma_A^2 + \sigma_{B(A)}^2 \\ \sigma_A^2 & \sigma_A^2 & \sigma_A^2 + \sigma_{B(A)}^2 & \sigma^2 + \sigma_A^2 + \sigma_{B(A)}^2 \end{pmatrix}$$

8. Appendix 3 : Covariance Matrices for the Estimated Parameters in the Linear Mixed Model

In order to perform inference on the parameters and random effects in the linear mixed model, some analytical results are needed to evaluate the precision of the parameter estimators. This is a delicate problem because, while some results exist, they are only asymptotically valid and may behave poorly in small samples. The two main results are the following.

First, the Henderson equations in Equation 8 make it possible to derive an analytical form for $\mathbf{C}$, the variance-covariance matrix of the vector of estimates $(\hat{\beta}', \hat{\gamma}')'$. Following the notation introduced in Section 2.6.2,

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}'_{21} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix}^{-1},$$

In calculating the inverse of the matrix, the variances of $(\hat{\beta}$ and $\hat{\gamma})$ can be expressed directly as

$$\text{Var}(\hat{\beta}) = \mathbf{C}_{11} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$$

and

$$\text{Var}(\hat{\gamma}) = \mathbf{C}_{22} = \mathbf{GZ}'\mathbf{PZG} \text{ with } \mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$$

In practice, the $\mathbf{R}$ and $\mathbf{G}$ in these equations are replaced with their MLE or REML estimators $\hat{\mathbf{R}}$ and $\hat{\mathbf{G}}$ so that numerical values for $\mathbf{C}$ can be obtained. Several works in the statistical literature discuss the behavior of these estimators in small samples and the ways that they should be used in inference tools (see e.g. [10], [67] or [47]).

Asymptotic expressions for the responses variance-covariance matrix $\hat{\mathbf{V}}$, which contains the maximum likelihood estimates of the variance components, can be obtained using the Fisher information matrix. These expressions take a fairly simple analytical form when the structure of $\mathbf{V}$ is simple (which is the assumption made in this text): $\mathbf{V} = \sum_{j=1}^R \sigma_j^2 \mathbf{Z}_j \mathbf{Z}_j' + \sigma^2 \mathbf{I}_n$. In this case, it may be shown that the asymptotic variance-covariance matrix of $(\hat{\sigma}^2, \hat{\sigma}_1^2, \dots, \hat{\sigma}_R^2)$ is an $(R+1) \times (R+1)$ matrix $2\mathbf{W}^{-1}$, where each element w_{ij} in $\mathbf{W}$ is defined as ([43]):

$$w_{ij} = \text{trace}(\mathbf{Z}_i' \mathbf{V}^{-1} \mathbf{Z}_j (\mathbf{Z}_i' \mathbf{V}^{-1} \mathbf{Z}_j)')$$

9. Appendix 4 : Small sample null distribution of LR test statistic

This appendix illustrates for Case study 2 and 3 the behaviour of the sampling distributions of the LR test statistics on the basis of bootstrap simulations. For the fixed effect Candies of Case study 3, Figure 16 shows that a χ^2 distribution with 4 degrees of freedom fits very well the empirical sampling distribution of the test statistic under the null. For all the other effects of Cases studies 2 and 3, which are random, Figures 15 and 16 show that the mixture a χ_0^2 and a χ_1^2 distribution fits better the sampling distribution of the test statistic than a χ_1^2 .

Figure 15: Case study 2 : Sampling distribution of LR test statistic on variance components

Figure 16: Case study 3 : Sampling distribution of LR test statistic on fixed effect candy and variance components

Acknowledgements

The authors thank the University UCLouvain for its support and Tormod Naes for providing the candies data set. The fourth author gratefully acknowledges funding from the Belgian Fund for Scientific Research (F.R.S.-FNRS) with a FRIA grant.

Conflicts of interests

Bernadette Govaerts and Rebecca Marion are employees of UCLouvain, Bernard Francq is an employee of the GSK group of companies, Manon Martin is an employee of F.R.S.-FNRS, Michel Thiel is an employee of Janssen Pharmaceutica. There is no conflicts of interest between these companies for the subject covered by this paper.

Softwares

The SAS, JMP and R software environment were used to perform the computations in this text. Data, SAS and R code and SAS, JMP RMarkdown outputs of the three case studies treated are available on https://github.com/bgovaerts/GLM_LMM_ANOVA.

References

- [1] N. R. Draper, H. Smith, Applied regression analysis, Vol. 326, John Wiley & Sons, 2014.
- [2] M. H. Kutner, C. J. Nachtsheim, J. Neter, W. Li, Applied linear statistical models, 5th Edition, McGraw-Hill Irwin Boston, 2005.
- [3] S. R. Searle, G. Casella, C. E. McCulloch, Variance components, Wiley series in probability and statistics, Wiley, Hoboken, NJ, 2006.
- [4] G. E. Box, J. S. Hunter, W. G. Hunter, Statistics for experimenters, Wiley Hoboken, NJ, USA, 2005.
- [5] A. I. Khuri, J. A. Cornell, Response surfaces: designs and analyses, Routledge, 2018.
- [6] J. Miller, J. C. Miller, Statistics and chemometrics for analytical chemistry, Pearson education, 2018.
- [7] D. L. Massart, B. G. Vandeginste, L. Buydens, S. De Jong, P. J. Lewi, J. Smeyers-Verbeke, C. K. Mann, Handbook of chemometrics and qualimetrics: Part a, Applied Spectroscopy 52 (1998) 302A.
- [8] G. A. Miliken, D. Johnson, Analysis of Messy Data, Volume I: Designed Experiments, New York: VanNostrand Reinhold, 1984.
- [9] H. Madsen, P. Thyregod, Introduction to general and generalized linear models, CRC Press, 2010.
- [10] C. E. McCulloch, S. R. Searle, Generalized, Linear, and Mixed Models (Wiley Series in Probability and Statistics), Wiley-Interscience, 2001.
- [11] J. J. Faraway, Linear models with R, Chapman and Hall/CRC, 2016.
- [12] D. M. Bates, lme4: Mixed-effects modeling with R, Springer-Preprint, 2010.
- [13] J. Fox, S. Weisberg, An R companion to applied regression, Sage Publications, 2018.
- [14] A.-M. Legendre, Nouvelles méthodes pour la détermination des orbites des comètes. Appendice sur la méthodes des moindres carrés, Firmin-Didot, Paris, 1805.
- [15] C. F. Gauss, Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium, 1809.
- [16] J. P. Stevens, Applied multivariate statistics for the social sciences, Routledge, 2012.
- [17] A. K. Smilde, J. J. Jansen, H. C. Hoefsloot, R.-J. A. Lamers, J. Van Der Greef, M. E. Timmerman, Anova-simultaneous component analysis (asca): a new tool for analyzing designed metabolomics data, Bioinformatics 21 (13) (2005) 3043–3048.
- [18] S. Wold, M. Sjöström, L. Eriksson, Partial least squares projections to latent structures (pls) in chemistry, Encyclopedia of computational chemistry 3 (2002) –.
- [19] G. Verbeke, G. Molenberghs, Linear mixed models for longitudinal data, Springer Science & Business Media, 2009.
- [20] P. McCullagh, J. A. Nelder, Generalized linear models, Chapman and Hall/CRC, 1989.
- [21] J. M. Chambers, T. J. Hastie, et al., Statistical models in S, Vol. 251, Wadsworth & Brooks/Cole Advanced Books & Software Pacific Grove, CA, 1992.
- [22] S. R. Searle, A. I. Khuri, Matrix algebra useful for statistics, John Wiley & Sons, 2017.
- [23] M. Anderson, C. T. Braak, Permutation tests for multi-factorial analysis of variance, Journal of statistical computation and simulation 73 (2) (2003) 85–113.
- [24] B. Efron, R. J. Tibshirani, An Introduction to the Bootstrap, no. 57 in Monographs on Statistics and Applied Probability, Chapman & Hall/CRC, Boca Raton, Florida, USA, 1993.
- [25] E. I. George, The variable selection problem, Journal of the American Statistical Association 95 (452) (2000) 1304–1308.
- [26] R. W. Mee, β -expectation and β -content tolerance limits for balanced one-way anova random model, Technometrics 26 (3) (1984) 251–254.

- [27] S. De Gryze, I. Langhans, M. Vandebroek, Using the correct intervals for prediction: A tutorial on tolerance intervals for ordinary least-squares regression, *Chemometrics and intelligent laboratory systems* 87 (2) (2007) 147–154.
- [28] K. Krishnamoorthy, T. Mathew, *Statistical Tolerance Regions: Theory, Applications, and Computation*, Wiley Series in Probability and Statistics, Wiley, 2009.
- [29] B. Francq, D. Lin, W. Hoyer, Confidence, prediction, and tolerance in linear mixed models, *Statistics in Medicine* 38 (2019) 5603–5622.
- [30] M. Ortiz, M. Sánchez, L. Sarabia, Quality of analytical measurements: univariate regression, *Comprehensive Chemometrics: chemical and biochemical data analysis* 1 (2009) 127–169.
- [31] R. H. Myers, *Classical and modern regression with applications*, Vol. 2, Duxbury press Belmont, CA, 1990.
- [32] J. Fox, *Regression diagnostics: An introduction*, Vol. 79, SAGE Publications, Incorporated, 2019.
- [33] R. Fisher, *Statistical methods for research workers*, Oliver & Boyd, 1925.
- [34] R. Fisher, *The design of experiments*, Oliver & Boyd, 1935.
- [35] L. Stahle, S. Wold, Analysis of variance (anova, *Chemometrics and Intelligent Laboratory Systems* 6 (1989) 259–272.
- [36] G. Zwanenburg, H. C. Hoefsloot, J. A. Westerhuis, J. J. Jansen, A. K. Smilde, Anova–principal component analysis and anova–simultaneous component analysis: a comparison, *Journal of Chemometrics* 25 (10) (2011) 561–567.
- [37] M. Thiel, B. Feraud, B. Govaerts, ASCA+ and APCA+: Extensions of ASCA and APCA in the analysis of unbalanced multifactorial designs, *Journal of Chemometrics* 31 (6) (2017) e2895.
- [38] S. Guisnet, M. Martin, B. Govaerts, Comparison of parafasca, acomdim, and amopls approaches in the multivariate glm modelling of multi-factorial designs, *Chemometrics and Intelligent Laboratory Systems* 184 (2019) 44–63.
- [39] C. R. Henderson, Estimation of variance and covariance components, *Biometrics* 9 (2) (1953) 226–252.
- [40] S. R. Searle, Another look at henderson’s methods of estimating variance components, *Biometrics* (1968) 749–787.
- [41] F. A. Graybill, *Matrices with applications in statistics*, Duxbury classic series., 1983.
- [42] H. Sahai, M. M. Ojeda, *Analysis of Variance for Random Models, Volume 1: Balanced Data: Theory, Methods, Applications, and Data Analysis*, Vol. 1, Springer Science & Business Media, 2004.
- [43] H. Sahai, M. M. Ojeda, *Analysis of Variance for Random Models, Volume 2: Unbalanced Data: Theory, Methods, Applications, and Data Analysis*, Vol. 2, Springer Science & Business Media, 2004.
- [44] J. Goodnight, F. Speed, Computing expected mean squares, *Biometrics* (1980) 123–125.
- [45] S. R. Searle, M. H. Gruber, *Linear models*, John Wiley & Sons, 2016.
- [46] S. R. Searle, *Linear models for unbalanced data*, John Wiley & Sons, 1987.
- [47] F. Gumedze, T. Dunne, Parameter estimation and inference in the linear mixed model, *Linear Algebra and its Applications* 435 (8) (2011) 1920–1944.
- [48] F. E. Satterthwaite, An approximate distribution of estimates of variance components, *Biometrics Bulletin* 2 (6) (1946) 110–114.
- [49] M. Kenward, J. Roger, Small sample inference for fixed effects from restricted maximum likelihood, *Biometrics* 53 (1997) 983–997.
- [50] A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, lmerTest package: Tests in linear mixed effects models, *Journal of Statistical Software* 82 (13) (2017) 1–26.
- [51] S. G. Luke, Evaluating significance in linear mixed-effects models in r, *Behavior Research Methods* 49 (2017) 1494–1502.
- [52] J. C. Pinheiro, D. M. Bates, Linear mixed-effects models: basic concepts and examples, *Mixed-effects models in S and S-Plus* (2000) 3–56.
- [53] D. O. Stram, J. W. Lee, Variance components testing in the longitudinal mixed effects model, *Biometrics* 50 (4) (1994) 1171–1177.
- [54] S. R. Lipsitz, J. G. Ibrahim, M. Parzen, A degrees-of-freedom approximation for a t-statistic with heterogeneous variances, *Journal of the Royal Statistical Society. Series D (The Statistician)* 48 (4) (1999) 495–506.
- [55] A. C. Davison, D. V. Hinkley, et al., *Bootstrap methods and their application*, Vol. 1, Cambridge university press, 1997.
- [56] Y. Fong, H. Rue, J. Wakefield, Bayesian inference for generalized linear mixed models, *Biostatistics* 11 (3) (2010) 397–412.
- [57] Wikipedia, Bleaching of wood pulp, https://en.wikipedia.org/wiki/Bleaching_of_wood_pulp (September 2019).
- [58] G. Luciano, T. Næs, Interpreting sensory data by combining principal component analysis and analysis of variance, *Food Quality and Preference* 20 (3) (2009) 167–175.
- [59] K. H. Liland, A. Smilde, F. Marini, T. Næs, Confidence ellipsoids for asca models based on multivariate regression theory, *Journal of Chemometrics* 32 (5) (2018) e2990.
- [60] M. Martin, B. Govaerts, Limm-pca : combining asca+ and linear mixed models to analyse high dimensional designed data, Discussion Paper DP2019/21, Institute of Statistics, Biostatistics and Actuarial Sciences, UCLouvain, Belgium (2019).
- [61] T. Næs, O. Langsrud, Fixed or random assessors in sensory profiling?, *Food Quality and Preference* 9 (3) (1998) 145–152, sensometric Workshop.
- [62] SAS Institute Inc., Cary, NC, JMP® 14 Fitting Linear Models (2018).
- [63] R. C. Team, R: A Language and Environment for Statistical Computing.
- [64] D. Bates, M. Maechler, B. Bolker, S. Walker, lme4: Linear mixed-effects models using ‘Eigen’ and S4, r package version 1.1-9 (2014).
URL <http://CRAN.R-project.org/package=lme4>
- [65] U. Halekoh, S. Højsgaard, A kenward-roger approximation and parametric bootstrap methods for tests in linear mixed models – the R package pbkrtest, *Journal of Statistical Software* 59 (9) (2014) 1–30.
- [66] SAS Institute Inc., Cary, NC, SAS®9.4 Procedures Guide. (2019).
- [67] R. Wolfinger, R. Tobias, J. Sall, Computing gaussian likelihoods and their derivatives for general linear mixed models, *SIAM Journal on Scientific Computing* 15 (6) (1994) 1294–1310.