‖‖ **eNote 4**

Mixed Model Theory, Part I

Indhold

The previous modules have introduced a number of situations where models including random effects are very useful. This module gives a fairly detailed description of the mixed model framework. Hopefully this will provide the reader with a better understanding of the structure and nature of these models, along with an improved ability to interpret results from these models.

4.1 Design matrix for a systematic linear model

▏▏▏▏ **Example 4.1   Two way ANOVA**

Consider again the systematic (or fixed effects) two way analysis of variance (ANOVA) model described in the first module. To lighten the notation in this example, the case with two treatments and three blocks is used. The usual way to present this model is:

$$y_{ij} = \underbrace{\mu}_{\text{overall mean}} + \overbrace{\alpha_i}^{\text{treatment effect}} + \underbrace{\beta_j}_{\text{block effect}} + \overbrace{\varepsilon_{ij}}^{\text{measurement noise}} , \quad \varepsilon_{ij} \sim N(0, \sigma^2),$$

where $i = 1, 2$, $j = 1, 2, 3$ and ε_{ij} are independent. In this simple case (with only 6 observations) it is possible to write exactly what this model states for every single observation:

$$
\begin{aligned}
y_{11} &= \mu + \alpha_1 & + \beta_1 & & + \varepsilon_{11} \\
y_{21} &= \mu & + \alpha_2 + \beta_1 & & + \varepsilon_{21} \\
y_{12} &= \mu + \alpha_1 & & + \beta_2 & + \varepsilon_{12} \\
y_{22} &= \mu & + \alpha_2 & + \beta_2 & + \varepsilon_{22} \\
y_{13} &= \mu + \alpha_1 & & + \beta_3 & + \varepsilon_{13} \\
y_{23} &= \mu & + \alpha_2 & + \beta_3 & + \varepsilon_{23}
\end{aligned}
\tag{4-1}
$$

This expansion illustrates exactly what parameters are used to predict each observation. For instance the observation from block 3 receiving treatment 1 y_{13} is predicted by adding the common mean parameter μ, the treatment parameter for the first treatment α_1, and the block parameter for the third block β_3. Just as expected. This expanded view inspires the following matrix notation for the model:

$$
\underbrace{\begin{pmatrix} y_{11} \\ y_{21} \\ y_{12} \\ y_{22} \\ y_{13} \\ y_{23} \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{13} \\ \varepsilon_{23} \end{pmatrix}}_{\varepsilon}
$$

> Notice how each column in the **X** matrix corresponds to a model parameter, and each row in the **X** matrix corresponds to a prediction of a single observation. Just like in the expanded model description in (4-1). (Remember from algebra how **X** multiplied by β will result in a 6×1 matrix (a column), where the element in the i'th row is the i'th row of **X** multiplied with β. For instance, the fifth element of $(\mathbf{X}\beta)$ will be equal to $\mu + \alpha_1 + \beta_3$, because each element from β is multiplied with the corresponding element in the fifth row of **X**, and then added together.)

A general linear fixed effects model can be presented in matrix notation by:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \quad \text{where } \varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I}). \tag{4-2}$$

The matrix **X** is known as the *design matrix*. The dimension of **X** is $n \times p$, where n is the number of observations in the data set and p is the number of fixed effect parameters in the model. The vector β consisting of all the fixed effect parameters has dimension $p \times 1$.

Notice how the distribution of the error term is described as:

$$\varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I})$$

Here ε is said to follow a *multivariate normal distribution* with mean vector $\mathbf{0}$ and covariance matrix $\sigma^2\mathbf{I}$. In general a stochastic vector **x** is said to follow a multivariate normal distribution with mean vector μ and covariance matrix Σ, if each coordinate x_i follows a normal distribution with mean μ_i and variance Σ_{ii}, and if the covariance between any two coordinates x_i and x_j is Σ_{ij}. In the case of ε, the mean vector is $\mathbf{0} = (0, \ldots, 0)'$, so each coordinate has mean zero. The covariance matrix is given as σ^2 times the unit matrix **I** with ones in the diagonal and zeroes anywhere else, so each coordinate has variance σ^2, and the covariance between any two coordinates is zero. If the covariance between two coordinates following a multivariate normal distribution is zero, then they are independent. This implication is not always true for stochastic variables following other distributions, but can be shown for the multivariate normal distribution.

The entire vector of observations **y** follows a multivariate normal distribution. The mean vector is $\mathbf{X}\beta$, and the covariance matrix, which is typically denoted **V**, is $\mathbf{V} = \text{var}(\mathbf{y}) = \sigma^2\mathbf{I}$. The covariance matrix V describes the covariance between any two observations, at any position i, j in the matrix it states the covariance between observation number i and observation number j. A few consequences of this is that V will always be symmetric $(V_{i,j} = V_{j,i})$, and the diagonal elements are the variances of the observations $(\text{cov}(y, y) = \text{var}(y))$. For a fixed effect model the covariance matrix is always given as $V = \sigma^2\mathbf{I}$, as all observations are independent, but for mixed models some more interesting structures will appear.

It is a fairly simple task to construct the design matrix corresponding to a given model and a given data set. For each factor and covariate in the model formulation the relevant columns are added to the matrix. For a factor a column is added for each level, with ones in the rows where the corresponding observation in the **y** vector is from that level, and zero otherwise. For a covariate a single column is added with the measurements of that covariate.

The matrix representation of a general linear fixed effects model is extremely useful for many purposes. For instance it reduces calculation of estimates and variances of estimates to simple (for a computer) matrix manipulations. Here a simple way of identifying *over-parameterization* via the matrix representation is given.

A model is said to be over-parameterized (or unidentifiable) if several different values of the parameters give the same model prediction, in which case it is not possible to find a unique set of parameter estimates. In general it is a challenging task to determine if a model is over-parameterized, especially for non-linear models, but for fixed effects linear models this can be done by calculating the rank of the design matrix.

> ⦀ **Example 4.2   Two way ANOVA (continued)**
>
> Consider again the two way (ANOVA) model with two treatments and three blocks:
>
> $$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \text{where } \varepsilon_{ij} \sim N(0, \sigma^2),$$
>
> This model is over-parameterized. Any set of parameter estimates $(\hat{\mu}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$ can for instance be replaced with $(\hat{\mu} + 1, \hat{\alpha}_1 - 1, \hat{\alpha}_2 - 1, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$, and the model would still give the exact same predictions of every single observation (think!).
>
> This is a known case of over-parameterization and the solution is to put some constraints on the parameters to ensure unique parameter estimates. One set of constraints that is known to work is to fix $\alpha_2 = 0$ and $\beta_3 = 0$ (SAS uses these constrains). Another set of constraints that is known to also work is to fix $\alpha_1 = 0$ and $\beta_1 = 0$ (R uses these constraints). It is important to realize that this does not change the model. The model is still able to give the exact same predictions, but the number of model parameters has been reduced to the minimum needed to describe the model.
>
> The rank of the design matrix **X** is 4, which means that **X** contains 4 linearly independent columns. The remaining 2 columns can be constructed as linear combinations of the four. This means that the model only needs four parameters, as the two remaining parameters is a linear combination of the four.
>
> One way of identifying what columns are a linear combination of the others (or in other words what parameters should be set to zero), is to simply step through the columns one at

the time, and for each column determine if that column is a linear combination of the previous columns. Every time a column is found to be linearly dependent of the previous independent columns, it is removed from the design matrix, and the corresponding parameter is set to zero. To illustrate this procedure for this model consider the steps:

$$
\underbrace{\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}}_{\text{step=1 rank=1}}
\quad
\underbrace{\begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}}_{\text{step=2 rank=2}}
\quad
\underbrace{\begin{pmatrix} 1 & 1 & \mathbf{0} \\ 1 & 0 & \mathbf{1} \\ 1 & 1 & \mathbf{0} \\ 1 & 0 & \mathbf{1} \\ 1 & 1 & \mathbf{0} \\ 1 & 0 & \mathbf{1} \end{pmatrix}}_{\text{step=3 rank=2}}
\quad
\underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}}_{\text{step=4 rank=3}}
\tag{4-3}
$$

$$
\underbrace{\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}}_{\text{step=5 rank=4}}
\quad
\underbrace{\begin{pmatrix} 1 & 1 & 1 & 0 & \mathbf{0} \\ 1 & 0 & 1 & 0 & \mathbf{0} \\ 1 & 1 & 0 & 1 & \mathbf{0} \\ 1 & 0 & 0 & 1 & \mathbf{0} \\ 1 & 1 & 0 & 0 & \mathbf{1} \\ 1 & 0 & 0 & 0 & \mathbf{1} \end{pmatrix}}_{\text{step=6 rank=4}}
\tag{4-4}
$$

A linear dependence is identified whenever the rank is not equal to the number of columns in the matrix. In this case column 3 and column 6 from the original design matrix can be removed and the corresponding parameters α_2 and β_3 fixed to zero. This procedure then corresponds to the SAS default approach. A similar procedure could be defined corresponding to the R default approach: "First columns in each block would have to be removed rather than last columns".

Notice how the complex task of identifying over-parameterization reduces to the relative simple operation of computing the rank of a few matrices. A lot of similar benefits of the matrix notation will become evident later in this course, but first this notation must be generalized to include mixed models.

4.2 The mixed model

The matrix notation for a mixed model is very similar to the matrix notation for a systematic model. The main difference is, that instead of one design matrix explaining the entire model, the matrix notation for a mixed model uses two design matrices. One design matrix $\mathbf{X}$ to describe the fixed effects in the model, and one design matrix $\mathbf{Z}$ to describe the random effects in the model. The fixed effects design matrix $\mathbf{X}$ is constructed in exactly the same way as in a fixed effects model, but for the fixed effects

only. $\mathbf{X}$ has dimension $n \times p$, where n is the number of observations in the data set and p is the number of fixed effect parameters in the model. The random effects design matrix $\mathbf{Z}$ is constructed in the same way, but for the random effects only. $\mathbf{Z}$ has dimension $n \times q$, where q is the number of random effect coefficients in the model.

‖‖‖ **Example 4.3   One way ANOVA with random block effects**

Consider the one way analysis of variance model with additional random block effect. Again the case with two treatments and three blocks is used.

$$y_{ij} = \mu + \alpha_i + b_j + \varepsilon_{ij}, \quad \text{where } b_j \sim N(0, \sigma_B^2) \text{ and } \varepsilon_{ij} \sim N(0, \sigma^2).$$

Here $i = 1, 2$, $j = 1, 2, 3$ and the random effects b_j and ε_{ij} are all independent. The matrix notation for this model is:

$$
\underbrace{\begin{pmatrix} y_{11} \\ y_{21} \\ y_{12} \\ y_{22} \\ y_{13} \\ y_{23} \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}}_{\mathbf{Z}} \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}}_{\mathbf{u}} + \underbrace{\begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{13} \\ \varepsilon_{23} \end{pmatrix}}_{\varepsilon},
$$

where $\mathbf{u} \sim N(0, \mathbf{G})$ and $\varepsilon \sim N(0, \sigma^2 \mathbf{I})$. The covariance matrix $\mathbf{G}$ for the random effects is in this case a 3×3 diagonal matrix with diagonal elements σ_B^2. Notice how the matrix representation exactly correspond to model formulation, when the matrices are multiplied.

A general linear mixed model can be presented in matrix notation by:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \varepsilon, \quad \text{where } \mathbf{u} \sim N(0, \mathbf{G}) \text{ and } \varepsilon \sim N(0, \mathbf{R}).$$

The vector $\mathbf{u}$ is the collection of all the random effect coefficients (just like β for the fixed effect parameters). The covariance matrix for the measurement errors $\mathbf{R} = \text{var}(\varepsilon)$ has dimension $n \times n$. In most examples $\mathbf{R} = \sigma^2 \mathbf{I}$, but in some examples to be described later in this course, it is convenient to use a different $\mathbf{R}$. The covariance matrix for the random effect coefficients $\mathbf{G} = \text{var}(\mathbf{u})$ has dimension $q \times q$, where q is the number of random effect coefficients. The structure of the $\mathbf{G}$ matrix can be very simple. If all the random effect coefficients are independent, then $\mathbf{G}$ is a diagonal matrix with diagonal elements equal to the variance of the random effect coefficients.

The covariance matrix $\mathbf{V}$ describing the covariance between any two observations in the data set, can be calculated directly from the matrix representation of the model in the

following way:

$$
\begin{aligned}
\mathbf{V} &= \operatorname{var}(\mathbf{y}) = \operatorname{var}(\mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \varepsilon) \quad \text{[from model]} \\
&= \operatorname{var}(\mathbf{X}\beta) + \operatorname{var}(\mathbf{Z}\mathbf{u}) + \operatorname{var}(\varepsilon) \quad \text{[all terms are independent]} \\
&= \operatorname{var}(\mathbf{Z}\mathbf{u}) + \operatorname{var}(\varepsilon) \quad \text{[variance of fixed effects is zero]} \\
&= \mathbf{Z}\operatorname{var}(\mathbf{u})\mathbf{Z}' + \operatorname{var}(\varepsilon) \quad \text{[$\mathbf{Z}$ is constant]} \\
&= \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} \quad \text{[from model]}
\end{aligned}
$$

These calculations used a few rules of calculus for stochastic variables, which has not been presented in their multivariate form. These rules are generalized from the univariate case, and will here be listed without proof. Let $\mathbf{x}$ and $\mathbf{y}$ be two multivariate stochastic variables, and let $\mathbf{A}$ be a constant matrix:

$$
\operatorname{var}(\mathbf{A}\mathbf{x}) = \mathbf{A}\operatorname{var}(\mathbf{x})\mathbf{A}',
$$

and if $\mathbf{x}$ and $\mathbf{y}$ are independent, then

$$
\operatorname{var}(\mathbf{x} + \mathbf{y}) = \operatorname{var}(\mathbf{x}) + \operatorname{var}(\mathbf{y}).
$$

||||| **Example 4.4   One way ANOVA with random block effects (continued)**

The covariance matrix $\mathbf{V}$ can now be calculated for the model in our example:

$$
\begin{aligned}
\mathbf{V} &= \operatorname{var}(\mathbf{y}) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \sigma^2\mathbf{I} \\
&=
\begin{pmatrix}
\sigma^2 + \sigma_B^2 & \sigma_B^2 & 0 & 0 & 0 & 0 \\
\sigma_B^2 & \sigma^2 + \sigma_B^2 & 0 & 0 & 0 & 0 \\
0 & 0 & \sigma^2 + \sigma_B^2 & \sigma_B^2 & 0 & 0 \\
0 & 0 & \sigma_B^2 & \sigma^2 + \sigma_B^2 & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma^2 + \sigma_B^2 & \sigma_B^2 \\
0 & 0 & 0 & 0 & \sigma_B^2 & \sigma^2 + \sigma_B^2
\end{pmatrix}
\end{aligned}
$$

The covariance matrix $\mathbf{V}$ illustrates the variance structure of the model. The first observation y_{11} correspond to the first column (or row) in $\mathbf{V}$. The variance of y_{11} is $V_{11} = \sigma^2 + \sigma_B^2$ (the covariance with itself). The covariance between y_{11} and y_{21} is $V_{12}(= V_{21}) = \sigma_B^2$. These two observations comes from the same block, so this correspond well with the intuition behind the model. The covariance between y_{11} and all other observations is zero. From the matrix $\mathbf{V}$ all possible covariances can be found.

4.3 Likelihood function and parameter estimation

The *likelihood function L* is a function of the observations and the model parameters. It returns a measure of the probability of observing a particular observation $\mathbf{y}$, given a set of model parameters β and γ. Here β is the vector of the fixed effect parameters, and γ is the vector of parameters used in the two covariance matrices $\mathbf{G}$ and $\mathbf{R}$, and hence in $\mathbf{V}$. Instead of the likelihood function L itself, it is often more convenient to work with *negative log likelihood function*, denoted ℓ. The negative log likelihood function for a mixed model is given by:

$$
\begin{aligned}
\ell(\mathbf{y}, \beta, \gamma) &= \frac{1}{2}\left\{ n\log(2\pi) + \log|\mathbf{V}(\gamma)| + (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{V}(\gamma))^{-1}(\mathbf{y} - \mathbf{X}\beta) \right\} \\
&\propto \frac{1}{2}\left\{ \log|\mathbf{V}(\gamma)| + (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{V}(\gamma))^{-1}(\mathbf{y} - \mathbf{X}\beta) \right\}
\end{aligned}
\tag{4-5}
$$

The symbol '$\propto$' reads 'proportional to', and is used here to indicate that only an additive constant (constant with respect to the model parameters) have been left out.

4.3.1 Maximum Likelihood estimation (ML)

A natural and often used method for estimating model parameters, is the maximum likelihood method. The maximum likelihood method take the actual observations and chooses the parameters which make those observations most likely. In other words, the parameter estimates are found by:

$$
(\hat{\beta}, \hat{\gamma}) = \underset{(\beta, \gamma)}{\operatorname{argmin}} \, \ell(\mathbf{y}, \beta, \gamma)
$$

In practice this minimum is found in three steps. 1) The estimate of the fixed effect parameters β is expressed as a function of the random effect parameters γ, as it turns out that no matter what value of the model parameters that minimize ℓ, then $\hat{\beta}(\gamma) = (\mathbf{X}'(\mathbf{V}(\gamma))^{-1}\mathbf{X})^{-1}\mathbf{X}'(\mathbf{V}(\gamma))^{-1}\mathbf{y}$. 2) The estimate of the random effect parameters is found by minimizing $\ell(\mathbf{y}, \hat{\beta}(\gamma), \gamma)$ as a function of γ. 3) The fixed effect parameters are calculated by $\hat{\beta} = \hat{\beta}(\hat{\gamma})$.

The maximum likelihood method is widely used to obtain parameter estimates in statistical models, because it has several nice properties. One nice property of the maximum likelihood estimator is "functional invariance", which means that for any function f, the maximum likelihood estimator of $f(\psi)$ is $f(\hat{\psi})$, where $\hat{\psi}$ is the maximum likelihood

estimator of ψ. For mixed models however, the maximum likelihood method on average tends to underestimate the random effect parameters, or in other words the estimator is *biased* downwards.

A well known example of this bias is in a simple random sample. Consider a random sample $\mathbf{x} = (x_1, \ldots, x_N)$ from a normal distribution with mean μ and variance σ^2. The mean parameter is estimated by the average $\hat{\mu} = (1/n) \sum x_i$. The maximum likelihood estimate of the variance parameter is $\hat{\sigma}^2 = (1/n) \sum (x_i - \hat{\mu})^2$. This estimate is not often used, because it is known to be biased. Instead the unbiased estimate $\hat{\sigma}^2 = (1/(n-1)) \sum (x_i - \hat{\mu})^2$ is most often used. This estimate is known as the *Restricted or Residual maximum likelihood estimate*.

4.3.2 REstricted/REsidual Maximum Likelihood estimation (REML)

The restricted (also known as residual) maximum likelihood method is a modification of the maximum likelihood method. Instead of minimizing the negative log likelihood function ℓ in step 2), the function ℓ_{re} is minimized, where ℓ_{re} is given by:

$$\frac{1}{2} \left\{ \log |\mathbf{V}(\gamma)| + (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{V}(\gamma))^{-1}(\mathbf{y} - \mathbf{X}\beta) + \log |\mathbf{X}'(\mathbf{V}(\gamma))^{-1}\mathbf{X}| \right\}$$

The two other steps 1) and 3) are exactly the same.

The intuition behind the raw maximum likelihood method, is that it should return the estimates that makes the actual observations most likely. The intuition behind the restricted likelihood method is almost the same, but instead of optimizing the likelihood of the observations directly, it optimizes the likelihood of the *full residuals*. The full residuals are defined as the observations minus the fixed effects part of the model. This focus on the full residuals can be theoretically justified, as it turns out, that these full residuals contain all information about the variance parameters.

This modification ensures, at least in balanced cases, that the random effect parameters are estimated without bias, and for this reason the REML estimator is generally preferred in mixed models.

4.3.3 Prediction of random effect coefficients

A model with random effects is typically used in situations where the subjects (or blocks) in the study are to be considered as representatives from a greater population. As

such, the main interest is usually not in the actual levels of the randomly selected subjects, but in the variation among them. It is however possible to obtain predictions of the individual subject levels $\mathbf{u}$. The formula is given in matrix notation as:

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

If $\mathbf{G}$ and $\mathbf{R}$ are known, $\hat{\mathbf{u}}$ can be shown to be 'the best linear unbiased predictor' (BLUP) of $\mathbf{u}$. Here, 'best' means minimum mean squared error. In real applications $\mathbf{G}$ and $\mathbf{R}$ are always estimated, but the predictor is still referred to as the BLUP.

▏▏▏▏ **Example 4.5**

Feeding experiment In (a subset of) a feeding experiment, the yield from six cows were measured in a period prior to the experiment, then three of the cows (randomly selected) were given a different type of fodder, and the yield were measured again. The following data were obtained:

Fodder type	Prior yield	Yield
1	34.32	24.09
1	32.00	26.59
1	29.41	25.64
2	32.32	30.71
2	33.67	21.96
2	29.80	27.00

A relevant model for evaluating the difference between the two types of fodder, could be:

$$y_{ij} = \mu + \alpha_i + \beta \cdot x_{ij} + \varepsilon_{ij},$$

where the prior yield is included as a covariate, and the factor α has a level for each of the two fodder types.

The design matrix for this model is:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 34.32 \\ 1 & 1 & 0 & 32.00 \\ 1 & 1 & 0 & 29.41 \\ 1 & 0 & 1 & 32.32 \\ 1 & 0 & 1 & 33.67 \\ 1 & 0 & 1 & 29.80 \end{pmatrix}$$

Here the first column corresponds to the overall mean parameter μ, the second column corresponds to the first level of α (the first fodder type), the third column corresponds to the

> second of α (the second fodder type), and the final column corresponds to the covariate (prior yield).
>
> It is fairly easy to see that this matrix only has three linearly independent columns. For instance the third column is equal to the first minus the second. This means that this model in its present form is over-parametrized. To get unique parameterization R will fix α_1 to zero.

4.4 Parametrization and post hoc analysis in two-way ANOVA

Consider the following artificial two-way data with factor A on 3 levels and factor B on two levels and two observations on each level:

	B1	B2
A1	2,4	12,14
A2	3,5	13,15
A3	4,6	14,16

The relevant averages showing what is going on in these data are:

	B1	B2		
A1	3	13	8	-1
A2	4	14	9	0
A3	5	15	10	+1
	4	14	9	
	-5	+5		0

Note that the data is constructed such that there is no interaction present. In the additive two-way model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}, \; i = 1,2,3, \; j = 1,2, \; k = 1,2$$

we see all together 6 different parameters: $(\mu, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2)$. But if the model is fit to this data using the parameter restriction method employed throughout in this course material, then the software would give you the following 4 numbers: $(3, 1, 2, 10)$:

```
A <- factor(c(1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3))
B <- factor(c(1, 1, 2, 2, 1, 1, 2, 2, 1, 1, 2, 2))
y=c(2, 4, 12, 14, 3, 5, 13, 15, 4, 6, 14, 16)
lm1 <- lm(y ~ A + B)
coef(lm1)


(Intercept)          A2          A3          B2
          3           1           2          10
```

and somehow let you know that the meaning of these numbers are, that $\alpha_1 = \beta_1 = 0$ and:

$$\begin{array}{cccc} \mu = \mu + \alpha_1 + \beta_1 & \alpha_2 - \alpha_1 & \alpha_3 - \alpha_1 & \beta_2 - \beta_1 \\ 3 & 1 & 2 & 10 \end{array}$$

Compare with the table of averages! The convention to set the first level of each factor at zero is default by R. In SAS, the default approach would be to set the last level of each factor at zero. It is possible to force R to use the SAS approach, if wanted for some reason, see below.

4.4.1 Post hoc analysis: LSMEANS and differences

In ANOVA, fixed as well as mixed, we must summarize the results of the significant fixed effects in the data. This is often done by estimating/computing the expected value for the different treatment levels (LSMEANS) and the expected differences between treatment levels. We already illustrated in Module 1 how we could do these things in R. Even though we can do most relevant post hoc analysis more or less automatically, it can indeed be helpful to really be able to specify/understand and work with contrasts in general.

Assume we would like the following four contrasts:

$$\alpha_2 - \alpha_1, \alpha_3 - \alpha_1, \alpha_3 - \alpha_2, \beta_2 - \beta_1$$

we could set up four different linear combinations of the 4 numbers provided by the software:

Contrast	$\mu = \mu + \alpha_1 + \beta_1$	$\alpha_2 - \alpha_1$	$\alpha_3 - \alpha_1$	$\beta_2 - \beta_1$
$\alpha_2 - \alpha_1$	0	1	0	0
$\alpha_3 - \alpha_1$	0	0	1	0
$\alpha_3 - \alpha_2$	0	-1	1	0
$\beta_2 - \beta_1$	0	0	0	1

Clearly three of them are already directly available. Such linear combinations of (original) parameters are called contrasts and can be computed together with standard errors/confidence intervals by R by us of the `estimable` function in the `gregmisc` package, or as indicated in Module 1 by the `glht` package.

The so-called LSMEANS may be somewhat more challenging to understand - a loose definition is given here:

LSMEAN for level i of factor A $= \mu + \alpha_i +$ Average of all other effects in the model

In the two-way ANOVA exemplified here there, the exact form then becomes:

$$\text{LSMEAN for level } i \text{ of factor A} = \mu + \alpha_i + \frac{1}{2}(\beta_1 + \beta_2)$$

For the example above, where "other effects" include a covariate x it would amount to inserting the average of the covariate:

$$\text{LSMEAN for level } i \text{ of factor A} = \mu + \alpha_i + \beta \bar{x}$$

Note that this approach is needed to be able to "tell the story" about the A-levels in a meaningful way, when other effects are present. In the two-way ANOVA case, JUST providing the values of, say, $\mu + \alpha_i$: $(3, 3 + 1, 3 + 2) = (3, 4, 5)$ would, as seen, implicitly "tell the A-story" within the first level of the B-factor. For the covariate situation, these values would be the model values for the covariate assumed to be zero - these are rarely relevant figures to provide. In the ANOVA example, the average of the B-parameters is:

$$\frac{1}{2}(0 + 10) = 5$$

So, the LSMEANs for the A-levels become:

Level		LSMEAN
A1	3+0+5	8
A2	3+1+5	9
A3	3+2+5	10

for similarly for the B-levels, since the average A-level is $1/3(0 + 1 + 2) = 1$:

Level		LSMEAN
B1	3+0+1	4
B2	3+10+1	14

which in this completely balanced case corresponds exactly to the raw averages. The linear combinations (contrasts) needed to obtain these from the original 4 parameter values given by the software are:

LSMEAN	$\mu = \mu + \alpha_1 + \beta_1$	$\alpha_2 - \alpha_1$	$\alpha_3 - \alpha_1$	$\beta_2 - \beta_1$
A1	1	0	0	1/2
A2	1	1	0	1/2
A3	1	0	1	1/2
B1	1	1/3	1/3	0
B2	1	1/3	1/3	1

In `estimable` in R these numbers are used directly as given here Note how the details fit, e.g. for the A3 LSMEAN:

$$1(\mu + \alpha_1 + \beta_1) + 0(\alpha_2 - \alpha_1) + 1(\alpha_3 - \alpha_1) + 1/2(\beta_1 - \beta_2)$$

$$= \mu + \alpha_1 + \beta_1 + \alpha_3 - \alpha_1 + 1/2\beta_1 - 1/2\beta_2 = \mu + \alpha_3 + 1/2\beta_1 + 1/2\beta_2$$

4.5 Model parametrizations in R

As default R uses a parametrization (opposite to SAS) in the sense that the parameter values for the first levels of each factor are set to zero. "First"in a alpha-numerical order of the level names/values. There is an option to make the approach in R just like in SAS: (The last set to zero instead):

```
options(contrasts=c(unordered="contr.SAS",ordered="contr.poly"))
```

There is no reason not just to use the R-default, however, if you like the SAS-way better, this would be the way to make it happen. The name of the default contrast setting in R is `contr.treatment`, so the following will set them back to default:

```
options(contrasts=c(unordered="contr.treatment",ordered="contr.poly"))
```

4.5.1 R code for the 2-way ANOVA example above

The artificial data:

```
A <- factor(c(1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3))
B <- factor(c(1, 1, 2, 2, 1, 1, 2, 2, 1, 1, 2, 2))
y=c(2, 4, 12, 14, 3, 5, 13, 15, 4, 6, 14, 16)
```

The four parameters are seen by running (we use "lm"but everything holds for using the "lmer"also)

```
lm1 <- lm(y ~ A + B)
coef(lm1)
```

```
(Intercept)          A2          A3          B2
          3           1           2          10
```

The four wanted differences are found by:

```
diffmatrix <- matrix(c(
0, 1, 0, 0,
0, 0, 1, 0,
0, -1, 1, 0,
0, 0, 0, -1),ncol=4,byrow=T)

rownames(diffmatrix)=c("A2-A1","A3-A1","A3-A2","B2-B1")
## Using package gregmisc:
library(gregmisc)
estimable(lm1,diffmatrix)
```

```
      Estimate Std. Error t value DF  Pr(>|t|)
A2-A1        1     0.8660   1.155  8 2.815e-01
A3-A1        2     0.8660   2.309  8 4.974e-02
A3-A2        1     0.8660   1.155  8 2.815e-01
B2-B1      -10     0.7071 -14.142  8 6.078e-07
```

```
## Using package glht:
summary(glht(lmres,linfct=diffmatrix))
```

```
 Simultaneous Tests for General Linear Hypotheses


Fit: lm(formula = y ~ A + B)


Linear Hypotheses:
           Estimate Std. Error t value Pr(>|t|)
A2-A1 == 0    1.000      0.866    1.15     0.63
A3-A1 == 0    2.000      0.866    2.31     0.15
A3-A2 == 0    1.000      0.866    1.15     0.63
B2-B1 == 0  -10.000      0.707  -14.14   <0.001 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

And the LSMEANS for BOTH A and B are found by:

```
lsmeansmatrix <- matrix(c(
1, 0, 0, 1/2,
1, 1, 0, 1/2,
1, 0, 1, 1/2,
1, 1/3, 1/3, 0,
1, 1/3, 1/3, 1), ncol = 4, byrow = T)

rownames(lsmeansmatrix) <- c("LSMEAN A1","LSMEAN A2",
                             "LSMEAN A3","LSMEAN B1","LSMEAN B2")

estimable(lm1,lsmeansmatrix)


          Estimate Std. Error t value DF  Pr(>|t|)
LSMEAN A1        8     0.6124   13.06  8 1.119e-06
LSMEAN A2        9     0.6124   14.70  8 4.514e-07
LSMEAN A3       10     0.6124   16.33  8 1.991e-07
LSMEAN B1        4     0.5000    8.00  8 4.367e-05
LSMEAN B2       14     0.5000   28.00  8 2.858e-09


## Or more easily:
 library(lsmeans)
 lsmeans(lm1, pairwise ~ A)
```

```
$'A lsmeans'
 A lsmean    SE df lower.CL upper.CL
 1      8 0.6124  8    6.588    9.412
 2      9 0.6124  8    7.588   10.412
 3     10 0.6124  8    8.588   11.412

$'A pairwise differences'
      estimate    SE df t.ratio p.value
1 - 2       -1 0.866  8  -1.155  0.5102
1 - 3       -2 0.866  8  -2.309  0.1118
2 - 3       -1 0.866  8  -1.155  0.5102
    p values are adjusted using the tukey method for 3 means



 lsmeans(lm1, pairwise ~ B)


$'B lsmeans'
 B lsmean  SE df lower.CL upper.CL
 1      4 0.5  8    2.847    5.153
 2     14 0.5  8   12.847   15.153

$'B pairwise differences'
      estimate    SE df t.ratio p.value
1 - 2      -10 0.7071  8  -14.14       0
    p values are adjusted using the tukey method for 2 means
```

4.6   Testing fixed effects

Typically the hypothesis of interest can be expressed as some linear combination of the model parameters:

$$\mathbf{L}'\beta = c$$

where $\mathbf{L}$ is a matrix, or a column vector with the same number of rows as there are elements in β. c is a constant and quite often zero. Consider the following example:

||||| **Example 4.6**

In a one way ANOVA model with three treatments the fixed effects parameter vector would be $\beta = (\mu, \alpha_1, \alpha_2, \alpha_3)'$. The test for similar effect of treatment 1 and treatment 2 can be expressed as:

$$\underbrace{\left(\begin{array}{cccc} 0 & 1 & -1 & 0 \end{array}\right)}_{\mathbf{L}'} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = 0$$

which is the same as $\alpha_1 - \alpha_2 = 0$. The hypothesis that all three treatments have the same effect can similarly be expressed as:

$$\underbrace{\begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}}_{\mathbf{L}'} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \mathbf{0}$$

where the **L**–matrix express that $\alpha_1 - \alpha_2 = 0$ and $\alpha_1 - \alpha_3 = 0$, which is the same as all three being equal.

Not every hypothesis that can be expressed as a linear combination of the parameters are meaningful.

||||| **Example 4.7**

Consider again the one way ANOVA example with parameters $\beta = (\mu, \alpha_1, \alpha_2, \alpha_3)'$. The hypothesis $\alpha_1 = 0$ is not meaningful for this model. This is not obvious right away, but consider the fixed part of the model with arbitrary α_1, and with $\alpha_1 = 0$:

$$E(y) = \mu + \begin{cases} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{cases} \quad \text{and} \quad E(y) = \tilde{\mu} + \begin{cases} 0 \\ \tilde{\alpha}_2 \\ \tilde{\alpha}_3 \end{cases}$$

The model with zero in place of α_1 can provide exactly the same predictions in each treatment group, as the model with arbitrary α_1. If for instance $\alpha_1 = 3$ in the first case, then setting $\tilde{\mu} = \mu + 3$, $\tilde{\alpha}_2 = \alpha_2 - 3$ and $\tilde{\alpha}_3 = \alpha_3 - 3$ will give the same predictions in the second case. In other words the two models are identical and comparing them with a statistical test is meaningless.

To avoid such situations the following definition is given:

Definition: A linear combination of the fixed effects model parameters $\mathbf{L}'\beta$ is said to be *estimable* if and only if there is a vector λ such that $\lambda'\mathbf{X} = \mathbf{L}'$.

In the following it is assumed that the hypothesis in question is estimable. This is not a restriction as all meaningful hypothesis are estimable.

The estimate of the linear combination of model parameters $\mathbf{L}'\beta$ is $\mathbf{L}'\hat{\beta}$. The estimate of β is known from the first theory module, so:

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

Applying the rule $\mathrm{cov}(\mathbf{A}\mathbf{x}) = \mathbf{A}\mathrm{cov}(\mathbf{x})\mathbf{A}'$ from the fist theory module, and doing few matrix calculations show that the covariance of $\mathbf{L}'\hat{\beta}$ is $\mathbf{L}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{L}$, and the mean is $\mathbf{L}\beta$. This all amounts to:

$$\mathbf{L}'\hat{\beta} \sim N(\mathbf{L}'\beta, \mathbf{L}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{L})$$

If the hypothesis $\mathbf{L}'\beta = c$ is true, then:

$$(\mathbf{L}'\hat{\beta} - c) \sim N(0, \mathbf{L}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{L})$$

Now the distribution is described, and the so called Wald test can be constructed by:

$$W = (\mathbf{L}'\hat{\beta} - c)'(\mathbf{L}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{L})^{-1}(\mathbf{L}'\hat{\beta} - c)'$$

The Wald test can be thought of as the squared difference from the hypothesis divided by its variance. W has an approximate $\chi^2_{df_1}$–distribution with degrees of freedom df_1 equal to the number of parameters "eliminated" by the hypothesis, which is the same as the rank of $\mathbf{L}$. This asymptotic result is based on the assumption that the variance $\mathbf{V}$ is known without error, but $\mathbf{V}$ is estimated from the observations, and not known.

A better approximation can be attained by using the Wald F–test:

$$F = \frac{W}{df_1}$$

in combination with the so-called Satterthwaite's approximation. In this case Satterthwaite's approximation supply an estimate of the denominator degrees of freedom df_2 (assuming that F is F_{df_1,df_2}–distributed). The P–value for the hypothesis $\mathbf{L}'\beta = c$ is computed as:

$$P_{\mathbf{L}'\beta=c} = P(F_{df_1,df_2} \geq F)$$

We will learn more about the Satterthwaite's approximation principle later. SAS users will have seen an option for this in the way SAS offers tests for fixed effects. Until recently these Satterthwaite based F-tests were not available in R. But now they are part of the lmerTest-package. In brief the following is the point: A single simple variance estimate will follow a χ^2-distribution. This is what the denominator consists of in simple

linear models F-testing. In mixed linear models F-testing, the denominator is some kind of complex linear combination of different variance component estimates. The distribution of such linear combinations of different χ^2-distributions are not theoretically a χ^2, BUT it turns out that it can be well approximated by one, if the DF is chosen properly (to match the mean and variance of the linear combination.)

4.7 Partial and sequential ANOVA Tables

- An ANOVA table: A list of SS-values for each effect/term in the model

- An SS-value expresses the (raw) amount of Y-variation explained by the effect.

- Successively(sequentially) constructed (In SAS: TYPE I, In R: given by "anova")

- Partially constructed (In SAS: TYPE III, In R: given by "drop1" OR "Anova(car)")

- (Also Type II and IV - but MORE subtle - and NOT so important)

- Generally: They (may) test different hypotheses - defined by the data structure (e.g. cell counts)

- Successively(sequentially) constructed (Type I)

  - Each SS is corrected ONLY for the effects listed PRIOR to the current effect (given some order of the effects)
  - The order comes from the way the model is expressed in the actual R/SAS model expression.
  - These SS-values sums together to the total SS-value of the data.
  - And hence: They give an actual decomposition of the total variability of the data.
  - The last SS in the table is also the Type III SS for that effect

- Partially constructed (Type III)

  - Each SS is corrected for ALL the effects in the model
  - These do NOT depend on the way the model is expressed
  - These SS-values will NOT generally sum together to the total SS-value of the data.
  - And hence: They do NOT give an actual decomposition of the total variability of the data.

- ONLY if the data is balanced (e.g. all cell counts n_{ij} are equal for a 2-way $A \times B$ setting):

    – The Type I and Type III are the SAME for all effects!

- Otherwise:

    – The Type I and Type III are NOT the SAME!

- Generally: We prefer Type III

- Type I is just sometimes a convenient way of looking at certain decompositions (e.g. in purely nested situations)

- When we look at "model summary"in R: the tests correspond to "Type III"tests (having been corrected for everything else)

4.8 Confidence intervals of fixed effects

Confidence intervals based on the approximative t–distribution, can be applied for linear combinations of the fixed effects. When a single fixed effect parameter or a single estimable linear combination of fixed effect parameters is considered, the **L** matrix has only one column, and the 95% confidence interval become:

$$\mathbf{L}'\beta = \mathbf{L}'\hat{\beta} \pm t_{0.975,df}\sqrt{\mathbf{L}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{L}}$$

Here the covariance matrix **V** is not known, but based variance parameter estimates. The only problem remaining is to determine the appropriate degrees of freedom df. Once again Satterthwaite's approximation is recommended. (An alternative to the is Satterthwaite's approximation is the Kenward-Rogers method, which is used by the `Anova` function of the car package and the `lsmeans` package)

4.9 Test for random effects parameters

The *restricted/residual likelihood ratio test* can be used to test the significance a random effects parameter. The likelihood ratio test is used to compare two models A and B, where B is a sub–model of A. Here the model including some variance parameter (model A), and the model without this variance parameter (model B) is to be compared. Using the test consists of two steps: 1) Compute the two negative restricted/residual

log-likelihood values ($\ell_{re}^{(A)}$ and $\ell_{re}^{(B)}$) by running both models. 2) Compute the test statistic:

$$G_{A \to B} = 2\ell_{re}^{(B)} - 2\ell_{re}^{(A)}$$

Clasically, such log-likelihood based statistics will asymptotically $G_{A \to B}$ follow a χ_1^2–distribution. (One degree of freedom, because one variance parameter is tested when comparing A to B).However the fact that changing from the more general model to the more specific model involves setting the variance of certain components of the random-effects to zero, which is on the boundary of the parameter region, the asymptotic results for LRT (LRT asymptotically follows distribution with one degree of freedom) have to be adjusted for boundary conditions. Following Self & Liang (1987); Stram & Lee (1994) the LRT more closely follows an equal mixture of χ^2-distributions with zero degrees of freedom (a point mass distribution) and one degree of freedom. The p-value from this test can be obtained by halving the p-value.

‖‖‖ **Remark 4.8**

It is important, when doing REML-based testing like this that the fixed part of the two nested models is the same - otherwise the REML likelihoods are not comoparable.

4.10   Confidence intervals for random effects parameters

The confidence interval for a given variance parameter is based on the assumption that the estimate of the variance parameter $\hat{\sigma}_b^2$ is approximately $\frac{\sigma_b^2}{df}\chi_{df}^2$–distributed. This is true in balanced (and other "nice") cases. A consequence of this is that the confidence interval takes the form:

$$\frac{(df)\hat{\sigma}_b^2}{\chi_{0.025;df}^2} < \sigma_b^2 < \frac{(df)\hat{\sigma}_b^2}{\chi_{0.975;df}^2} \, ,$$

but with the degrees of freedom df still undetermined.

The task is to choose the df such that the corresponding χ^2–distribution matches the distribution of the estimate. The (theoretical) variance of $\frac{\sigma_b^2}{df}\chi_{df}^2$ is:

$$\mathrm{var}\left(\frac{\sigma_b^2}{df}\chi_{df}^2\right) = \frac{2\sigma_b^4}{df}$$

The actual (asymptotic) variance of the parameter can be estimated from the curvature of the negative log likelihood function ℓ (by general maximum likelihood theory). By matching the estimated actual variance of the estimator to the variance of the desired distribution, and solving the equation:

$$\text{var}(\hat{\sigma}_b^2) = \frac{2\sigma_b^4}{df}$$

the following estimate of the degrees of freedom is obtained, after plugging in the estimated variance:

$$\hat{df} = \frac{2\hat{\sigma}_b^4}{\text{var}(\hat{\sigma}_b^2)}$$

This way of approximating the degrees of freedom is a special case of Satterthwaite's approximation.

4.11 R-TUTORIAL: Testing random effects

Some remarks on how to test fixed and random effects in R is given. Since typically the random structure should be determined before looking at the fixed structure, we start with the random effects. The planks data from Modules 3 and 6 is used as an example:

```
planks <- read.table("planks.txt", header = TRUE, sep = ",")
planks$plank<-factor(planks$plank)
planks$depth<-factor(planks$depth)
planks$width<-factor(planks$width)
planks$loghum <- log(planks$humidity)
library(lme4)
```
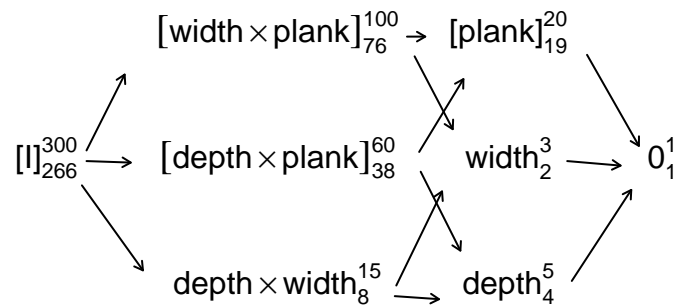
The model corresponding to:

or expressed formally:

$$\log Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + d_k + f_{ik} + g_{jk} + \epsilon_{ijk} \tag{4-6}$$

where

$$d_k \sim N(0, \sigma_{Plank}^2), \quad f_{ik} \sim N(0, \sigma_{Plank*width}^2)$$

$$g_{jk} \sim N(0, \sigma_{Plank*debth}^2), \quad \epsilon_{ijk} \sim N(0, \sigma^2).$$

is fitted by the following:

$$[\text{width} \times \text{plank}]_{76}^{100} \rightarrow [\text{plank}]_{19}^{20}$$

$$[\text{I}]_{266}^{300} \longrightarrow [\text{depth} \times \text{plank}]_{38}^{60} \qquad \text{width}_2^3 \longrightarrow 0_1^1$$

$$\text{depth} \times \text{width}_8^{15} \longrightarrow \text{depth}_4^5$$

Figur 4.1: The factor structure diagram

```
model4<-lmer(loghum~depth*width + (1|plank)  + (1|depth:plank)
           + (1|plank:width), data = planks)
```

The REML test for the hypothesis

$$\sigma^2_{Plank*depth} = 0$$

is carried out by fitting the submodel and comparing this with the full model:

```
model4.1 <- lmer(loghum~depth*width + (1|plank)  + (1|plank:width),
           data = planks)
anova(model4, model4.1, refit = FALSE)

Data: planks
Models:
..1: loghum ~ depth * width + (1 | plank) + (1 | plank:width)
object: loghum ~ depth * width + (1 | plank) + (1 | depth:plank) + (1 |
object:     plank:width)
       Df  AIC  BIC logLik deviance Chisq Chi Df Pr(>Chisq)
..1    18 -426 -359    231     -462
object 19 -427 -356    232     -465  2.58      1       0.11
```

And similarly for the other interaction random component:

```
model4.2 <- lmer(loghum~depth*width + (1|plank)  + (1|depth:plank),
           data = planks)
anova(model4, model4.2, refit = FALSE)


Data: planks
Models:
..1: loghum ~ depth * width + (1 | plank) + (1 | depth:plank)
object: loghum ~ depth * width + (1 | plank) + (1 | depth:plank) + (1 |
object:     plank:width)
       Df  AIC  BIC logLik deviance Chisq Chi Df Pr(>Chisq)
..1    18 -317 -250    176     -353
object 19 -427 -356    232     -465   112      1     <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note the logical option `refit` indicating if the models should be refitted with ML before comparing models. The default is TRUE to prevent the common mistake of inappropriately comparing REML-fitted models with different fixed effects, whose likelihoods are not directly comparable.

Now compare with the ANOVA table from the fixed effects analysis also given in the examples section of module 06:

Source of variation	DF	Sums of squares	Mean squares	F	P-value
depth	4	4.28493467	1.07123367	(217.14)	($<$ .0001)
width	2	0.79785186	0.39892593	(80.86)	($<$ .0001)
depth*width	8	0.08877653	0.01109707	2.25	0.0268
plank	19	9.12355684	0.48018720	(97.34)	($<$ .0001)
depth*plank	76	0.51331023	0.00675408	1.37	0.0521
width*plank	38	1.74239118	0.04585240	9.29	$<$ .0001
Error	152	0.74986837	0.00493334		

(The F-statistics (and P-values) NOT to be used are put in parentheses. Only the highest order interaction tests wil generally make good sense.) Here we also get tests for the two random interaction effects (with similar but not 100% equivalent results). These F-test statistics are formally tests in the fixed model, BUT can be argued to work as tests

also in the random model. They are based on the F-distribution, which is an EXACT distributional property (if the model is correct) whereas the χ^2-tests of the random/mixed model carried out above are based on APPROXIMATE distributional results. Note that it is perfectly correct that the χ^2-tests only have DF=1, since only one single parameter is tested in these tests! So in summary: The approximate χ^2-test approach given in this theory module can be applied for any (set of) random structure parameter in any mixed model, and is hence a general applicable method. In SOME cases, we may be able to find/construct F-statistics with potentially better distributional properties.

Please also note the discussion about devideing the P-value with 2 above.

A nice feature about the lmerTest-package is that it offers an automated step-wise investigation of the significance of all the random effects:

```
library(lmerTest)
model4<-lmer(loghum~depth*width + (1|plank)  + (1|depth:plank)
             + (1|plank:width), data = planks)
step4 <- step(model4, reduce.fixed = FALSE)
step4$rand.table


            Chi.sq Chi.DF elim.num   p.value
plank        38.219      1     kept 6.324e-10
depth:plank   2.715      1     kept 9.939e-02
plank:width 117.732      1     kept 1.985e-27
```

4.12 R-TUTORIAL: Testing fixed effects

Testing of fixed effects, given some random model structure in a mixed model is a topic of discussion and various approaches exist. First of all, we already saw in Module 1 and above how we could use the `estimable` function to provide post hoc comparisons in the mixed model. These are examples of the Wald tests as presented in this module. We have also seen the use of `lsmeans` and the multcomp-package. Finally, we also saw that if we want the overall tests for interaction or main effects as usually given in the ANOVA table of a purely fixed model, we could get it from e.g. the `Anova` function of the car package or the `anova` function of the lmerTest package:

```
anova(model4)


Analysis of Variance Table of type 3  with  Satterthwaite
```

```
approximation for degrees of freedom
          Sum Sq Mean Sq NumDF DenDF F.value  Pr(>F)
depth      3.130   0.782     4    76   158.6 < 2e-16 ***
width      0.086   0.043     2    38     8.7 0.00077 ***
depth:width 0.089  0.011     8   152     2.2 0.02678 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If we again compare this with the fixed ANOVA table, we will be able to find the two main effect F-test in this case as:

$$F_{depth} = \frac{\text{MS}_{depth}}{\text{MS}_{depth*plank}} \text{ and } F_{width} = \frac{\text{MS}_{width}}{\text{MS}_{width*plank}}$$

and the interaction directly as:

$$F_{depth:width} = \frac{\text{MS}_{depth:width}}{\text{MS}_{Error}}$$

This is an example of how the general REML-likelihood based approach will lead to something simple that we could have realized, if we are familiar with this so-called Henderson-based method and the mixed ANOVA Expected Mean Square (EMS) theory. The strength of the general mixed model approach is that it will do the right thing even though we loose the balance/completeness of data. And in these simple cases it ALSO gives the right thing.

4.13 R-TUTORIAL: Estimating random effects with CIs

We do this very brief. The Satterthwaite based method described above requires that we can extract so-called asymptotic variance-covariance matrix for the random parameters. Likelihood theory tells us that this is found from the second derivatives (Hessian) of the (RE)ML log-likelihood function with the estimated parameters inserted. Unfortunately, in R this is not easily extracted from an lmer-result object.

Instead one can easily get the so-called likelihood profile based confidence intervals:

```
summary(model4)$varcor
```

```
 Groups      Name        Std.Dev.
```

```
 depth:plank (Intercept) 0.0246
 plank:width (Intercept) 0.0905
 plank       (Intercept) 0.1698
 Residual                0.0702
```

```
m4prof <- profile(model4,which=1:4)
confint(m4prof)
```

```
        2.5 %  97.5 %
.sig01 0.00000 0.03814
.sig02 0.06960 0.11421
.sig03 0.11924 0.24019
.sigma 0.06159 0.07670
```

4.14 Exercises

ⅠⅠⅠⅠ Exercise 1   Two-way ANOVA model

a) Consider the two way ANOVA model with interaction:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad i = 1,2, \; j = 1,2, \; k = 1,2$$

Write down the design matrix $\mathbf{X}$. (Hint: The interaction term is a factor with one level for each combination of the two factors).

b) Is this model over-parametrized?

c) Try to see how R deals with this model by running the following code:

```
sex <- factor(c(rep("female", 4), rep("male", 4)))
tmt <- factor(c(0, 0, 1, 1, 0, 0, 1, 1))
y <- c(-9.27, -1.28, 3.98, 7.06, 1.02, -1.79, 3.64, 1.94)
summary(lm(y ~ sex*tmt))


Call:
lm(formula = y ~ sex * tmt)

Residuals:
    1     2     3     4     5     6     7     8
-3.99  3.99 -1.54  1.54  1.40 -1.40  0.85 -0.85

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)     -5.27       2.29   -2.30    0.083 .
sexmale          4.89       3.24    1.51    0.206
tmt1            10.79       3.24    3.33    0.029 *
sexmale:tmt1    -7.62       4.59   -1.66    0.172
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.24 on 4 degrees of freedom
Multiple R-squared:  0.754,Adjusted R-squared:  0.57
F-statistic: 4.09 on 3 and 4 DF,  p-value: 0.104
```

Notice what parameters are fixed to zero in R output. Now try the SAS way:

```
options(contrasts=c(unordered="contr.SAS",ordered="contr.poly"))
summary(lm(y ~ sex*tmt))


Call:
lm(formula = y ~ sex * tmt)

Residuals:
    1     2     3     4     5     6     7     8
-3.99  3.99 -1.54  1.54  1.40 -1.41  0.85 -0.85

Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)      2.79        2.29    1.22     0.29
sexfemale        2.73        3.24    0.84     0.45
tmt0            -3.17        3.24   -0.98     0.38
sexfemale:tmt0  -7.62        4.59   -1.66     0.17


Residual standard error: 3.24 on 4 degrees of freedom
Multiple R-squared:  0.754,Adjusted R-squared:  0.57
F-statistic: 4.09 on 3 and 4 DF,  p-value: 0.104



options(contrasts=c(unordered="contr.treatment",ordered="contr.poly"))
```

||||| Exercise 2   Design of specific tests

a) Consider the situation where we wish to compare four treatments. The first three are "real" treatments with three different drugs (A, B, C), and the fourth is a placebo treatment (P). The observations are collected from different blocks (farms).

A natural model for this experiment is the one–way ANOVA model with random block effect.

$$y_i = \mu + \alpha(\texttt{treatment}_i) + b(\texttt{block}_i) + \varepsilon_i$$

The basic analysis of this model can be carried out in R by writing:

```
lmer(y ~ treatment + (1|block))
```

The test in the `treatment` row of the ANOVA table is the test for all four treatments being equal $\alpha_A = \alpha_B = \alpha_C = \alpha_P$. This test can be used to see if at least one of the treatments have an effect, but often we would like to set up more specific tests.

For the rest of this exercise we will assume that the fixed effects parameter vector for this model is organized as: $\beta = (\mu, \alpha_A, \alpha_B, \alpha_C, \alpha_P)'$. Set up the **L**–matrix (with one column and five rows) to test it treatment A is no better than placebo (P).

b) Set up the **L**–matrix to test if all three real treatments are equally good $\alpha_A = \alpha_B = \alpha_C$.

c) Sketch how to do the tests from a) and b) using R. (We have no real data for this exercises)