

Introdução geral aos modelos mistos

Vanderly Janeiro

Dep. de Estatística - UEM

Conteúdo

4.1	Matriz de planejamento para um modelo linear sistemático	2
4.1.1	Exemplo 4.1 ANOVA bidirecional - Two way ANOVA	2
4.1.2	Exemplo 4.2 ANOVA bidirecional - Two way ANOVA (continuação) .	4
4.2	O modelo misto	6
4.2.1	Exemplo 4.3 ANOVA unidirecional com efeitos de bloco aleatório . .	6
4.2.2	Exemplo 4.4 ANOVA unidirecional com efeitos de bloco aleatório (continuação)	8
5	Função de verossimilhança e estimativa de parâmetros	10
5.1	Estimativa de máxima verossimilhança (ML)	10
5.1.1	Estimativa de máxima verossimilhança REstricted/REsidual (REML)	11
5.1.2	Predição de coeficientes de efeito aleatório	12
5.2	Parametrização e análise post hoc em ANOVA bidirecional	13
5.2.1	Análise post hoc: LSMEANS e diferenças	14
5.3	Parametrizações de modelo em R	16
5.4	Testando efeitos fixos	18
5.5	Tabelas ANOVA parciais e sequenciais	21
5.6	Intervalos de confiança de efeitos fixos	22
5.7	Teste para parâmetros de efeitos aleatórios	22
5.8	Intervalos de confiança para parâmetros de efeitos aleatórios	23
5.9	R-TUTORIAL: Testando efeitos aleatórios	24
5.10	R-TUTORIAL: Testando efeitos fixos	31
5.11	R-TUTORIAL: Estimando efeitos aleatórios com ICs	32

Obs.: Este documento é uma tradução e alguns (poucos) incrementos do arquivo orinal encontrado no endereço <http://www2.compute.dtu.dk/courses/02429/enotepdfs/eNote-4.pdf> em 03/12/2022.

Este módulo fornece uma descrição bastante detalhada da estrutura do modelo misto. Esperamos que isto proporcione ao leitor uma melhor compreensão da estrutura e natureza destes modelos, juntamente com uma melhor capacidade de interpretar os resultados destes modelos.

4.1 Matriz de planejamento para um modelo linear sistemático

4.1.1 Exemplo 4.1 ANOVA bidirecional - Two way ANOVA

Considere novamente o modelo de análise de variância bidirecional (ANOVA) sistemático (ou de efeitos fixos) descrito no primeiro módulo. Para tornar a notação mais leve neste exemplo, é utilizado o caso com dois tratamentos e três blocos. A forma usual de apresentar este modelo é:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2), \quad (1)$$

em que:

- μ denota a média geral;
- α_i o efeito do i -ésimo tratamento;
- β_j o efeito do j -ésimo bloco.

em que $i = 1, 2$, $j = 1, 2, 3$ e ε_{ijk} são independentes. Neste caso simples (com apenas 6 observações) é possível escrever exatamente o que este modelo afirma para cada observação:

$$\begin{array}{llll} y_{11} = \mu & +\alpha_1 & & +\beta_1 & +\varepsilon_{11} \\ y_{21} = \mu & & +\alpha_2 & +\beta_1 & +\varepsilon_{21} \\ y_{12} = \mu & +\alpha_1 & & +\beta_2 & +\varepsilon_{12} \\ y_{22} = \mu & & +\alpha_2 & +\beta_2 & +\varepsilon_{22} \\ y_{13} = \mu & +\alpha_1 & & +\beta_3 & +\varepsilon_{13} \\ y_{23} = \mu & & +\alpha_2 & +\beta_3 & +\varepsilon_{23} \end{array}$$

Esta expansão ilustra exatamente quais parâmetros são usados para prever cada observação. Por exemplo, a observação do bloco 3 que recebe o tratamento 1 (y_{13}) é predita adicionando o parâmetro de média comum μ , o parâmetro de tratamento para o primeiro tratamento α_1 e o parâmetro de bloco para o terceiro bloco β_3 . Esta visão expandida inspira a seguinte notação matricial para o modelo:

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

Observe como cada coluna da matriz \mathbf{X} corresponde a um parâmetro do modelo e cada linha da matriz \mathbf{X} corresponde a predição de uma única observação. Assim como na descrição expandida do modelo em (1). (Lembre-se da álgebra como \mathbf{X} multiplicado por $\boldsymbol{\beta}$ resultará em uma matriz 6×1 (uma coluna), em que o elemento na i -ésima linha é a i -ésima linha de \mathbf{X} multiplicado por $\boldsymbol{\beta}$. Por exemplo, o quinto elemento de $(\mathbf{X}\boldsymbol{\beta})$ será igual $\mu + \alpha_1 + \beta_3$, porque cada elemento de $\boldsymbol{\beta}$ é multiplicado pelo elemento correspondente na quinta linha de \mathbf{X} e depois somado.)

Um modelo linear geral de efeitos fixos pode ser apresentado em notação matricial por:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \quad (2)$$

A matriz \mathbf{X} é conhecida como matriz de delineamento. A dimensão de \mathbf{X} é $n \times p$, sendo n o número de observações no conjunto de dados e p o número de parâmetros de efeito fixo no modelo. O vetor $\boldsymbol{\beta}$ que consiste em todos os parâmetros de efeito fixo tem dimensão $p \times 1$.

Observe que a distribuição do termo de erro é descrita como:

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

Aqui, diz-se que $\boldsymbol{\epsilon}$ segue uma distribuição normal multivariada com vetor de médias $\mathbf{0}$ e matriz de covariância $\sigma^2 \mathbf{I}$. Em geral, diz-se que um vetor estocástico \mathbf{x} segue uma distribuição normal multivariada com vetor de médias $\boldsymbol{\mu}$ e matriz de covariância Σ , se cada coordenada x_i segue uma distribuição normal com média μ_i e variância Σ_{ii} , e se a covariância entre quaisquer duas coordenadas x_i e x_j é Σ_{ij} . No caso de $\boldsymbol{\epsilon}$, o vetor médio é $\mathbf{0} = (0, \dots, 0)'$, então cada coordenada tem média zero. A matriz de covariância é dada como σ^2 vezes a matriz unitária \mathbf{I} com uns na diagonal e zeros em qualquer outro lugar, então cada coordenada tem variância σ^2 e a covariância entre quaisquer duas coordenadas é zero. Se a covariância entre duas coordenadas seguindo uma distribuição normal multivariada for zero, então elas

são independentes. Esta implicação nem sempre é verdadeira para variáveis estocásticas que seguem outras distribuições, mas pode ser demonstrada para a distribuição normal multivariada.

O vetor de observações \mathbf{y} segue uma distribuição normal multivariada, com vetor de média $\mathbf{X}\boldsymbol{\beta}$, e matriz de covariância, denotada por \mathbf{V} , tal que $\mathbf{V} = \text{Var}(\mathbf{y}) = \sigma^2 \mathbf{I}$. A matriz de covariância \mathbf{V} descreve a covariância entre quaisquer duas observações, em qualquer posição i , j na matriz ela declara a covariância entre o número de observação i e o número de observação j . Algumas consequências disso é que \mathbf{V} sempre será simétrico ($V_{i,j} = V_{j,i}$), e os elementos diagonais são as variâncias das observações ($\text{Cov}(y, y) = \text{var}(y)$). Para um modelo de efeito fixo a matriz de covariância é sempre dada como $\mathbf{V} = \sigma^2 \mathbf{I}$, pois todas as observações são independentes, mas para modelos mistos aparecerão algumas estruturas mais interessantes.

É uma tarefa bastante simples construir a matriz de delineamento correspondente a um determinado modelo e a um determinado conjunto de dados. Para cada fator e covariável na formulação do modelo, as colunas relevantes são adicionadas à matriz. Para um fator, uma coluna é adicionada para cada nível, com unidades nas linhas onde a observação correspondente no vetor \mathbf{y} é desse nível e zero caso contrário. Para uma covariável, uma única coluna é adicionada com as medidas dessa covariável.

A representação matricial de um modelo linear geral de efeitos fixos é extremamente útil para muitos propósitos. Por exemplo, reduz o cálculo de estimativas e suas variações a simples manipulações de matrizes (para um computador). Aqui é fornecida uma maneira simples de identificar a sobreparametrização por meio da representação matricial.

Diz-se que um modelo está **sobreparametrizado** (ou **não identificável**) se vários valores diferentes dos parâmetros fornecem a mesma predição do modelo, caso em que não é possível encontrar um conjunto único de estimativas de parâmetros. Em geral, é uma tarefa desafiadora determinar se um modelo está superparametrizado, especialmente para modelos não lineares, mas para modelos lineares de efeitos fixos isso pode ser feito calculando a ordem da matriz de delineamento.

4.1.2 Exemplo 4.2 ANOVA bidirecional - Two way ANOVA (continuação)

Considere novamente o modelo bidirecional (ANOVA) com dois tratamentos e três blocos:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2),$$

este modelo é superparametrizado. Qualquer conjunto de estimativas de parâmetros $(\hat{\mu}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$ pode, por exemplo, ser substituído por $(\hat{\mu} + 1, \hat{\alpha}_1 - 1, \hat{\alpha}_2 - 1, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$,

e o modelo ainda daria exatamente as mesmas prediç ões de cada observação (pense!).

Este é um caso conhecido de sobreparametrização e a solução é colocar algumas restrições nos parâmetros para garantir estimativas de parâmetros únicas. A solução é colocar algumas restrições nos parâmetros para garantir estimativas exclusivas de parâmetros:

- O SAS fixa $\alpha_2 = 0$ e $\beta_3 = 0$;
- O R fixa $\alpha_1 = 0$ e $\beta_1 = 0$;

É importante perceber que isso não altera o modelo. O modelo ainda é capaz de fornecer exatamente as mesmas predições, mas o número de parâmetros do modelo foi reduzido ao mínimo necessário para descrever o modelo.

A matriz de delineamento \mathbf{X} é de ordem 4, significando que \mathbf{X} contém 4 colunas linearmente independentes. As 2 colunas restantes podem ser construídas como combinações lineares das quatro. Logo o modelo necessita apenas de quatro parâmetros, pois os dois parâmetros restantes são uma combinação linear dos quatro.

Uma maneira de identificar quais colunas são uma combinação linear das outras (ou em outras palavras, quais parâmetros devem ser definidos como zero), é simplesmente percorrer as colunas, uma de cada vez, e para cada coluna determinar se essa coluna é combinação linear das colunas anteriores. Cada vez que uma coluna é linearmente dependente das colunas independentes anteriores, ela é removida da matriz de delineamento e o parâmetro correspondente é definido como zero. Para ilustrar este procedimento para este modelo, considere as etapas:

$$\begin{array}{cccc}
 \underbrace{\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}}_{\text{passo}=1 \text{ rank}=1} &
 \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}}_{\text{passo}=2 \text{ rank}=2} &
 \underbrace{\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}}_{\text{passo}=3 \text{ rank}=2} &
 \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}}_{\text{passo}=4 \text{ rank}=3}
 \end{array}$$

$$\underbrace{\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}}_{\text{passo}=5 \text{ rank}=4} \quad \underbrace{\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{passo}=5 \text{ rank}=4}$$

Uma dependência linear é identificada sempre que o posto não é igual ao número de colunas da matriz. Neste caso, a coluna 3 e a coluna 6 da matriz de delineamento original podem ser removidas e os parâmetros correspondentes α_2 e β_3 fixados em zero. Este procedimento corresponde então à abordagem padrão do SAS. Um procedimento semelhante poderia ser definido correspondendo à abordagem padrão do R: “As primeiras colunas de cada bloco teriam que ser removidas em vez das últimas colunas”.

Observe como a tarefa complexa de identificar a parametrização excessiva se reduz à operação relativamente simples de calcular a classificação de algumas matrizes. Muitos benefícios semelhantes da notação matricial ficarão evidentes mais adiante neste curso, mas primeiro esta notação deve ser generalizada para incluir modelos mistos.

4.2 O modelo misto

A notação matricial para um modelo misto é muito semelhante à notação matricial para um modelo sistemático. A principal diferença é que, em vez de uma matriz de delineamento explicando todo o modelo, a notação matricial para um modelo misto usa duas matrizes de delineamento. Uma matriz de delineamento \mathbf{X} para descrever os efeitos fixos no modelo e uma matriz de delineamento \mathbf{Z} para descrever os efeitos aleatórios no modelo. A matriz de delineamento de efeitos fixos \mathbf{X} é construída exatamente da mesma maneira que em um modelo de efeitos fixos, mas apenas para os efeitos fixos. \mathbf{X} tem dimensão $n \times p$, sendo n é o número de observações no conjunto de dados e p é o número de parâmetros de efeito fixo no modelo. A matriz de delineamento de efeitos aleatórios \mathbf{Z} é construída da mesma maneira, mas apenas para os efeitos aleatórios. \mathbf{Z} tem dimensão $n \times q$, q é o número de coeficientes de efeito aleatório no modelo.

4.2.1 Exemplo 4.3 ANOVA unidirecional com efeitos de bloco aleatório

Considere o modelo de análise de variância unidirecional com efeito de bloco aleatório adicional. Novamente é utilizado o caso com dois tratamentos e três blocos.

$$y_{ij} = \mu + \alpha_i + b_j + \varepsilon_{ij}, \quad \text{em que } b_j \stackrel{iid}{\sim} N(0, \sigma_B^2) \text{ e } \varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2),$$

Aqui $i = 1, 2$, $j = 1, 2, 3$ e os efeitos aleatórios b_j e ε_{ij} são todos independentes. A notação matricial para este modelo é:

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

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

sendo $\mathbf{u} \sim N(0, \mathbf{G})$ e $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$. A matriz de covariância \mathbf{G} para os efeitos aleatórios é neste caso uma matriz diagonal 3×3 com elementos diagonais σ_B^2 . Observe como a representação matricial corresponde exatamente à formulação do modelo, quando as matrizes são multiplicadas.

Um modelo linear misto geral pode ser apresentado em notação matricial por:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \quad \text{sendo } \mathbf{u} \sim N(0, \mathbf{G}) \text{ e } \boldsymbol{\varepsilon} \sim N(0, \mathbf{R})$$

O vetor \mathbf{u} é a coleção de todos os coeficientes de efeito aleatório (assim como $\boldsymbol{\beta}$ para os parâmetros de efeito fixo). A matriz de covariância para os erros de medição $\mathbf{R} = \text{Var}(\boldsymbol{\varepsilon})$ tem dimensão $n \times n$. Na maioria dos exemplos $\mathbf{R} = \sigma^2 \mathbf{I}$, mas em alguns exemplos a serem descritos posteriormente neste curso, é conveniente usar um \mathbf{R} diferente. A matriz de covariância para os coeficientes de efeito aleatório $\mathbf{G} = \text{Var}(\boldsymbol{\mu})$ tem dimensão $q \times q$, onde q é o número de coeficientes de efeito aleatório. A estrutura da matriz \mathbf{G} pode ser muito simples. Se todos os coeficientes de efeito aleatório forem independentes, então \mathbf{G} é uma matriz diagonal com elementos diagonais iguais à variância dos coeficientes de efeito aleatório.

A matriz de covariância \mathbf{V} que descreve a covariância entre quaisquer duas observações no conjunto de dados pode ser calculada diretamente a partir da representação matricial do modelo da seguinte maneira, sendo o vetor de média $\boldsymbol{\mu}$ e a matriz de covariância \mathbf{V} :

$$\boldsymbol{\mu} = \mathbb{E}(\mathbf{Y}) = \mathbb{E}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}) = \mathbf{X}\boldsymbol{\beta} \quad \text{os outros termos tem média zero}$$

$$\begin{aligned} \mathbf{V} &= \text{Var}(\mathbf{Y}) = \text{Var}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}) && \text{variância de } \mathbf{y} \text{ apartir do modelo} \\ &= \text{Var}(\mathbf{X}\boldsymbol{\beta}) + \text{Var}(\mathbf{Z}\mathbf{u}) + \text{Var}(\boldsymbol{\varepsilon}) && \text{todos os termos são independentes} \\ &= \text{Var}(\mathbf{Z}\mathbf{u}) + \text{Var}(\boldsymbol{\varepsilon}) && \text{variância do efeito fixo é zero} \\ &= \mathbf{Z}\text{Var}(\mathbf{u})\mathbf{Z}^T + \text{Var}(\boldsymbol{\varepsilon}) && \mathbf{Z} \text{ é constante} \\ &= \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R} && \text{variância de } \mathbf{y} \text{ segundo modelo} \end{aligned}$$

$$\text{Assim, } \mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R})$$

Esses cálculos utilizaram algumas regras de álgebra para variáveis estocásticas, que não foram apresentadas em sua forma multivariada. Estas regras são generalizadas a partir do caso univariado e serão listadas aqui sem prova. Sejam \mathbf{x} e \mathbf{y} duas variáveis estocásticas multivariadas e seja \mathbf{A} uma matriz constante:

$$\text{Var}(\mathbf{A}\mathbf{x}) = \mathbf{A}\text{Var}(\mathbf{x})\mathbf{A}',$$

e se \mathbf{x} e \mathbf{y} são independentes, então

$$\text{Var}(\mathbf{x} + \mathbf{y}) = \text{Var}(\mathbf{x}) + \text{Var}(\mathbf{y}).$$

4.2.2 Exemplo 4.4 ANOVA unidirecional com efeitos de bloco aleatório (continuação)

A matriz de covariância \mathbf{V} agora pode ser calculada para o modelo em nosso exemplo:

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R} = \mathbf{Z}(\sigma_b^2 \mathbf{I})\mathbf{Z}^T + \sigma^2 \mathbf{I}$$

$$\mathbf{V} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sigma_b^2 & 0 & 0 \\ 0 & \sigma_b^2 & 0 \\ 0 & 0 & \sigma_b^2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} + \begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

$$\mathbf{V} = \begin{pmatrix} \sigma_b^2 & 0 & 0 \\ \sigma_b^2 & 0 & 0 \\ 0 & \sigma_b^2 & 0 \\ 0 & \sigma_b^2 & 0 \\ 0 & 0 & \sigma_b^2 \\ 0 & 0 & \sigma_b^2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} + \begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

$$\mathbf{V} = \begin{pmatrix} \sigma_b^2 & \sigma_b^2 & 0 & 0 & 0 & 0 \\ \sigma_b^2 & \sigma_b^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_b^2 & \sigma_b^2 & 0 & 0 \\ 0 & 0 & \sigma_b^2 & \sigma_b^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_b^2 & \sigma_b^2 \\ 0 & 0 & 0 & 0 & \sigma_b^2 & \sigma_b^2 \end{pmatrix} + \begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

$$\mathbf{V} = \begin{pmatrix} \sigma^2 + \sigma_b^2 & \sigma_b^2 & 0 & 0 & 0 & 0 \\ \sigma_b^2 & \sigma^2 + \sigma_b^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 + \sigma_b^2 & \sigma_b^2 & 0 & 0 \\ 0 & 0 & \sigma_b^2 & \sigma^2 + \sigma_b^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 + \sigma_b^2 & \sigma_b^2 \\ 0 & 0 & 0 & 0 & \sigma_b^2 & \sigma^2 + \sigma_b^2 \end{pmatrix}$$

Note que observações no mesmo bloco são correlacionadas.

Esta matriz de covariância \mathbf{V} ilustra a estrutura de variância do modelo. A primeira observação y_{11} corresponde à primeira coluna (ou linha) em \mathbf{V} . A variância de y_{11} é $V_{11} = \sigma^2 + \sigma_b^2$ (a covariância consigo mesma). A covariância entre y_{11} e y_{21} é $V_{12}(= V_{21}) = \sigma_b^2$. Essas duas observações vêm do mesmo bloco, portanto isso corresponde bem ao modelo. A covariância entre y_{11} e todas as outras observações é zero. Da matriz \mathbf{V} podem ser encontradas todas as covariâncias possíveis.

5 Função de verossimilhança e estimativa de parâmetros

A *função de verossimilhança* L é uma função das observações e dos parâmetros do modelo. Ele retorna uma medida da probabilidade de observar uma observação específica \mathbf{y} , dado um conjunto de parâmetros do modelo $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$. Aqui $\boldsymbol{\beta}$ é o vetor dos parâmetros de efeito fixo e $\boldsymbol{\gamma}$ é o vetor de parâmetros usados nas duas matrizes de covariância \mathbf{G} e \mathbf{R} e, portanto, em \mathbf{V} . Em vez da própria função de verossimilhança L , muitas vezes é mais conveniente trabalhar com negativo função de log de verossimilhança, denotada por ℓ . A função de log de verossimilhança negativa para um modelo misto é dada por:

$$\begin{aligned}\ell(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}) &= \frac{1}{2} \left\{ n \log(2\pi) + \log |\mathbf{V}(\boldsymbol{\gamma})| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{V}(\boldsymbol{\gamma}))^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \\ &\propto \frac{1}{2} \left\{ \log |\mathbf{V}(\boldsymbol{\gamma})| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{V}(\boldsymbol{\gamma}))^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}\end{aligned}\quad (3)$$

O símbolo “ \propto ” diz “proporcional a” e é usado aqui para indicar que apenas uma constante aditiva (constante em relação aos parâmetros do modelo) foi deixada de fora.

5.1 Estimativa de máxima verossimilhança (ML)

Um método natural e frequentemente utilizado para estimar parâmetros do modelo é o método de máxima verossimilhança. O método de máxima verossimilhança “pega as observações reais e escolhe os parâmetros que tornam essas observações mais prováveis”. Em outras palavras, as estimativas dos parâmetros são encontradas por:

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = \arg \min_{(\boldsymbol{\beta}, \boldsymbol{\gamma})} \ell(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma})$$

Na prática esse mínimo é encontrado em três etapas.

- 1) A estimativa dos parâmetros de efeito fixo $\boldsymbol{\beta}$ é expressa como uma função dos parâmetros de efeito aleatório $\boldsymbol{\gamma}$, pois verifica-se que não importa qual seja o valor dos parâmetros do modelo que minimizam ℓ , então $\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = (\mathbf{X}'(\mathbf{V}(\boldsymbol{\gamma}))^{-1} \mathbf{X})^{-1} \mathbf{X}'(\mathbf{V}(\boldsymbol{\gamma}))^{-1} \mathbf{y}$.
- 2) A estimativa dos parâmetros do efeito aleatório é encontrada minimizando $\ell(\mathbf{y}, \hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}), \boldsymbol{\gamma})$ em função de $\boldsymbol{\gamma}$.
- 3) Os parâmetros de efeito fixo são calculados por $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\gamma}})$.

O método de máxima verossimilhança é amplamente utilizado para obter estimativas de parâmetros em modelos estatísticos, pois possui diversas propriedades interessantes. Uma propriedade interessante do estimador de máxima verossimilhança é a “invariância funcional”, o que significa que para qualquer função f , o estimador de máxima verossimilhança de $f(\psi)$ é $f(\hat{\psi})$, no qual $\hat{\psi}$ é o estimador de máxima verossimilhança de ψ . No entanto, para modelos mistos, o método da máxima verossimilhança tende, em média, a subestimar os parâmetros do efeito aleatório ou, por outras palavras, o estimador é viesado para baixo.

Um exemplo bem conhecido desse viés está em uma amostra aleatória simples. Considere uma amostra aleatória $\mathbf{x} = (x_1, \dots, x_N)$ de uma distribuição normal com média μ e variância σ^2 . O parâmetro médio é estimado pela média $\hat{\mu} = (1/n) \sum x_i$. A estimativa de máxima verossimilhança do parâmetro de variância é $\hat{\sigma}^2 = (1/n) \sum (x_i - \hat{\mu})^2$. Esta estimativa não é frequentemente utilizada porque é conhecida por ser tendenciosa. Em vez disso, a estimativa imparcial $\hat{\sigma}^2 = (1/(n-1)) \sum (x_i - \hat{\mu})^2$ é mais frequentemente usada. Essa estimativa é conhecida como estimativa de verossimilhança máxima restrita ou residual.

5.1.1 Estimativa de máxima verossimilhança REstricted/REsidual (REML)

O método de máxima verossimilhança restrita (também conhecido como residual) é uma modificação do método de máxima verossimilhança. Em vez de minimizar a função logarítmica de verossimilhança negativa ℓ no passo 2), a função ℓ_{re} é minimizada, dado por:

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

As outras duas etapas 1) e 3) são exatamente as mesmas.

A intuição por trás do método de máxima verossimilhança bruta é que deve retornar as estimativas que tornam as observações reais mais prováveis. A intuição por trás do método de verossimilhança restrita é quase a mesma, mas em vez de otimizar diretamente a probabilidade das observações, otimiza a probabilidade dos *resíduos completos*. Os resíduos completos são definidos como as observações menos a parte dos efeitos fixos do modelo. Este foco nos resíduos completos pode ser justificado teoricamente, pois estes resíduos completos contêm todas as informações sobre os componentes de variância.

Esta modificação garante, pelo menos em casos equilibrados, que os parâmetros de efeitos aleatórios sejam estimados sem viés, e por esta razão o estimador REML é geralmente preferido em modelos mistos.

5.1.2 Predição de coeficientes de efeito aleatório

Um modelo com efeitos aleatórios é normalmente usado em situações em que os sujeitos (ou blocos) do estudo devem ser considerados representantes de uma população maior. Como tal, o interesse principal geralmente não está nos níveis reais dos assuntos selecionados aleatoriamente, mas na variação entre eles. No entanto, é possível obter previsões dos níveis de disciplina individuais \mathbf{u} . A fórmula é dada em notação de matriz como:

$$\hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

Se \mathbf{G} e \mathbf{R} forem conhecidos, pode-se mostrar que $\hat{\mathbf{u}}$ é “o melhor preditor linear imparcial” (BLUP) de \mathbf{u} . Aqui, “melhor” significa erro quadrático médio mínimo. Em aplicações reais, \mathbf{G} e \mathbf{R} são sempre estimados, mas o preditor ainda é conhecido como BLUP.

5.1.2.1 Exemplo 4.5 - Experimento de alimentação Em (um subconjunto de) um experimento de alimentação, o rendimento de seis vacas foi medido em um período anterior ao experimento, então três das vacas (selecionadas aleatoriamente) receberam um tipo diferente de forragem e o rendimento foi medido novamente. Os seguintes dados foram obtidos:

Tipo de forragem	Rendimento anterior	Rendimento
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

Um modelo relevante para avaliar a diferença entre os dois tipos de forragens poderia ser:

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

de forma que o rendimento anterior é incluído como covariável e o fator α tem um nível para cada um dos dois tipos de forragem. A matriz de delineamento para este modelo é:

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

Aqui a primeira coluna corresponde ao parâmetro médio geral μ , a segunda coluna corresponde ao primeiro nível de α (o primeiro tipo de forragem), a terceira coluna corresponde ao segundo de α (o segundo tipo de forragem) e a coluna final corresponde para a covariável (rendimento anterior).

É bastante fácil ver que esta matriz possui apenas três colunas linearmente independentes. Por exemplo, a terceira coluna é igual à primeira menos a segunda. Isto significa que este modelo na sua forma atual está sobreparametrizado. Para obter parametrização exclusiva, R fixará α_1 em zero.

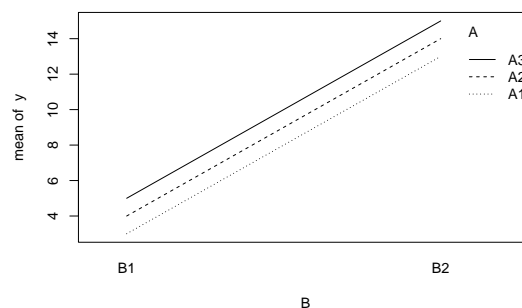
5.2 Parametrização e análise post hoc em ANOVA bidirecional

Considere os seguintes dados bidirecionais artificiais com o fator A em 3 níveis e o fator B em dois níveis e duas observações em cada nível:

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

As médias relevantes que mostram o que está acontecendo nesses dados são:

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



Observe que os dados são construídos de forma que não haja interação presente. No modelo aditivo bidirecional:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}, \quad \begin{cases} i = 1, 2, 3 \\ j = 1, 2 \\ k = 1, 2 \end{cases}$$

vemos todos juntos 6 parâmetros diferentes: $(\mu, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2)$. Mas se o modelo se ajustar a esses dados usando o método de restrição de parâmetros empregado neste material do curso, o software fornecerá os seguintes 4 números: (3, 1, 2, 10):

(Intercept)	AA2	AA3	BB2
3	1	2	10

como o R usa a restrição $\alpha_1 = \beta_1 = 0$, temos que os coeficientes são referentes a:

$\hat{\mu} = \hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_1$	$\hat{\alpha}_2 - \hat{\alpha}_1$	$\hat{\alpha}_3 - \hat{\alpha}_1$	$\hat{\beta}_2 - \hat{\beta}_1$
3	1	2	10

5.2.1 Análise post hoc: LSMEANS e diferenças

Na ANOVA, fixa e mista, devemos resumir os resultados dos efeitos fixos significativos nos dados. Isto é muitas vezes feito estimando/computando o valor esperado para os diferentes níveis de tratamento (LSMEANS) e as diferenças esperadas entre os níveis de tratamento. Já ilustramos no Módulo 1 como poderíamos fazer essas coisas em R. Embora possamos fazer análises post hoc mais relevantes de forma mais ou menos automática, pode de fato ser útil ser realmente capaz de especificar/compreender e trabalhar com contrastes em geral.

Suponha que gostaríamos dos quatro contrastes a seguir:

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

poderíamos configurar quatro combinações lineares diferentes dos 4 números fornecidos pelo software:

Contraste	$\mu = \mu + \alpha_1 + \beta_1$	$\alpha_2 - \alpha_1$	$\alpha_3 - \alpha_1$	$\beta_2 - \beta_1$
$\alpha_2 - \alpha_1$	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

É evidente que três deles já estão diretamente disponíveis. Tais combinações lineares de parâmetros (originais) são chamadas de contrastes e podem ser calculadas juntamente com erros padrão/intervalos de confiança por R por meio da função `estimable` no pacote `gmodels`.

Os chamados LSMEANS podem ser um pouco mais difíceis de entender - uma definição vaga é dada aqui:

LSMEAN para nível i do fator $A = \mu + \alpha_i +$ Média de todos os demais efeitos do modelo
Na ANOVA bidirecional exemplificada aqui, a forma exata então se torna:

$$\text{LSMEAN para nível } i \text{ do fator } A = \mu + \alpha_i + \frac{1}{2}(\beta_1 + \beta_2)$$

Para o exemplo acima, onde “outros efeitos” incluem uma covariável \times equivaleria a inserir a média da covariável:

$$\text{LSMEAN para nível } i \text{ do fator } A = \mu + \alpha_i + \beta \bar{x}$$

Note-se que esta abordagem é necessária para poder “contar a história” sobre os níveis A de uma forma significativa, quando outros efeitos estão presentes. No caso da ANOVA bidirecional, APENAS fornecer os valores de, digamos, $\mu + \alpha_i : (3, 3 + 1, 3 + 2) = (3, 4, 5)$ iria, como visto, implicitamente “contar A história” dentro do primeiro nível do fator B. Para a situação de covariável, estes valores seriam os valores do modelo para a covariável assumida como zero – estes raramente são números relevantes para fornecer. No exemplo ANOVA, a média dos parâmetros B é:

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

Assim, os LSMEANS para os níveis A tornam-se:

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

da mesma forma para os níveis B, uma vez que o nível A médio é $1/3(0 + 1 + 2) = 1$:

Nível	LSMEAN
B1 3+0+1	4
B2 3+10+1	14

o que neste caso completamente balanceado corresponde exatamente às médias brutas. As combinações lineares (contrastes) necessárias para obtê-los a partir dos 4 valores originais dos parâmetros fornecidos pelo software são:

Contraste	$\mu = \mu + \alpha_1 + \beta_1$	$\alpha_2 - \alpha_1$	$\alpha_3 - \alpha_1$	$\beta_2 - \beta_1$
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

Na função `estimable`, esses números são usados diretamente conforme fornecido aqui. Observe como os detalhes se ajustam, por exemplo. para o A3 LSMEAN:

$$\begin{aligned}
 \text{A3} &\rightarrow 1(\mu + \alpha_1 + \beta_1) + 0(\alpha_2 - \alpha_1) + 1(\alpha_3 - \alpha_1) + 1/2(\beta_2 - \beta_1) \\
 &= \mu + \alpha_1 + \beta_1 + \alpha_3 - \alpha_1 + 1/2\beta_2 - 1/2\beta_1 \\
 &= \mu + \alpha_3 + 1/2\beta_1 + 1/2\beta_2
 \end{aligned}$$

5.3 Parametrizações de modelo em R

Como padrão o R utiliza uma parametrização (oposta ao SAS) no sentido de que os valores dos parâmetros para os primeiros níveis de cada fator são definidos como zero. “Primeiro” em ordem alfanumérica dos nomes/valores dos níveis. Existe uma opção de fazer a abordagem em R assim como no SAS: (O último definido como zero):

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

A configuração de contraste padrão do R é `contr.treatment`, portanto, o seguinte código irá configurá-lo de volta ao padrão:

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

```
fm<- aov(y ~A+B, data=two.way)
coefficients(fm)
```

```
(Intercept)      AA2      AA3      BB2
              3         1         2        10
```

```
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")
```

```
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")
```

```
library(gmodels)
```

```
estimable(fm,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

```
estimable(fm,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

```
library(lsmeans)
```

```
lsmeans(fm, pairwise ~ A)
```

```
$lsmeans
```

A	lsmean	SE	df	lower.CL	upper.CL
A1	8 0.612	8	6.59	9.41	
A2	9 0.612	8	7.59	10.41	

```
A3      10 0.612  8      8.59    11.41
```

Results are averaged over the levels of: B

Confidence level used: 0.95

```
$contrasts
```

contrast	estimate	SE	df	t.ratio	p.value
A1 - A2	-1	0.866	8	-1.155	0.5102
A1 - A3	-2	0.866	8	-2.309	0.1118
A2 - A3	-1	0.866	8	-1.155	0.5102

Results are averaged over the levels of: B

P value adjustment: tukey method for comparing a family of 3 estimates

```
lsmeans(fm, pairwise ~ B)
```

```
$lsmeans
```

B	lsmean	SE	df	lower.CL	upper.CL
B1	4	0.5	8	2.85	5.15
B2	14	0.5	8	12.85	15.15

Results are averaged over the levels of: A

Confidence level used: 0.95

```
$contrasts
```

contrast	estimate	SE	df	t.ratio	p.value
B1 - B2	-10	0.707	8	-14.142	<.0001

Results are averaged over the levels of: A

5.4 Testando efeitos fixos

Normalmente, a hipótese de interesse pode ser expressa como alguma combinação linear dos parâmetros do modelo:

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

em que \mathbf{L} é uma matriz ou um vetor coluna com o mesmo número de linhas que há elementos

em β . c é uma constante e muitas vezes zero. Considere o seguinte exemplo

Exemplo 4.6

Em um modelo ANOVA unidirecional com três tratamentos, o vetor de parâmetros de efeitos fixos seria $\beta = (\mu, \alpha_1, \alpha_2, \alpha_3)'$. O teste para efeito igualdade entre o tratamento 1 e o tratamento 2 pode ser expresso como:

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

que é o mesmo que $\alpha_1 - \alpha_2 = 0$. A hipótese de que todos os três tratamentos têm o mesmo efeito pode ser expressa de forma semelhante como:

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

aqui a matriz L expressa que $\alpha_1 - \alpha_2 = 0$ e $\alpha_1 - \alpha_3 = 0$, que é o mesmo que todos os três sendo iguais.

Nem todas as hipóteses que podem ser expressas como uma combinação linear dos parâmetros são significativas.

Exemplo 4.7

Considere novamente o exemplo da ANOVA unilateral com parâmetros $\beta = (\mu, \alpha_1, \alpha_2, \alpha_3)'$. A hipótese $\alpha_1 = 0$ não é significativa para este modelo. Isso não é óbvio de imediato, mas considere a parte fixa do modelo com α_1 arbitrário e com $\alpha_1 = 0$:

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

O modelo com zero no lugar de α_1 pode fornecer exatamente as mesmas predições em cada grupo de tratamento que o modelo com α_1 arbitrário. Se, por exemplo, $\alpha_1 = 3$ no primeiro caso, então definindo $\tilde{\mu} = \mu + 3$, $\tilde{\alpha}_2 = \alpha_2 - 3$ e $\tilde{\alpha}_3 = \alpha_3 - 3$ darão as mesmas predições no

segundo caso. Em outras palavras, os dois modelos são idênticos e compará-los com um teste estatístico não faz sentido.

Para evitar tais situações, é dada a seguinte definição:

Definição: Uma combinação linear dos parâmetros $\mathbf{L}'\boldsymbol{\beta}$ do modelo de efeitos fixos é considerada estimável se e somente se existe um vetor $\boldsymbol{\lambda}$ tal que $\boldsymbol{\lambda}'\mathbf{X} = \mathbf{L}'$.

A seguir, assume-se que a hipótese em questão é estimável. Isto não é uma restrição, pois todas as hipóteses significativas são estimáveis.

A estimativa da combinação linear dos parâmetros do modelo $\mathbf{L}'\boldsymbol{\beta}$ é $\mathbf{L}'\hat{\boldsymbol{\beta}}$. A estimativa de $\boldsymbol{\beta}$ é conhecida desde o primeiro módulo teórico, então:

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

Aplicando a regra $\text{Cov}(\mathbf{A}\mathbf{x}) = \mathbf{A}\text{Cov}(\mathbf{x})\mathbf{A}'$ do primeiro módulo de teoria, e fazendo alguns cálculos matriciais mostram que a covariância de $\mathbf{L}'\hat{\boldsymbol{\beta}}$ é $\mathbf{L}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{L}$, e a média é $\mathbf{L}'\boldsymbol{\beta}$. Tudo isso equivale a:

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

Se a hipótese $\mathbf{L}'\boldsymbol{\beta} = c$ for verdadeira, então:

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

Agora a distribuição está descrita, e o chamado teste de Wald pode ser construído por:

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

O teste de Wald pode ser pensado como a diferença quadrada da hipótese dividida pela sua variância. W tem uma distribuição aproximada $\chi^2_{gl_1}$ com graus de liberdade gl_1 igual ao número de parâmetros “eliminados” pela hipótese, que é o mesmo que a classificação de \mathbf{L} . Este resultado assintótico é baseado na suposição de que a variância \mathbf{V} é conhecido sem erro, mas \mathbf{V} é estimado a partir das observações e não é conhecido. Uma melhor aproximação pode ser obtida usando o teste F de Wald:

$$F = \frac{W}{gl_1}$$

em combinação com a chamada aproximação de *Satterthwaite*. Neste caso, a aproximação de Satterthwaite fornece uma estimativa dos graus de liberdade do denominador gl_2 (assumindo que F é F_{gl_1, gl_2} -distribuído). O valor P para a hipótese $\mathbf{L}'\boldsymbol{\beta} = c$ é calculado como:

$$P_{\mathbf{L}'\boldsymbol{\beta}=c} = P(F_{gl_1, gl_2} \geq F)$$

Os testes F baseados em Satterthwaite estão disponíveis no R, no pacote `lmerTest`. Em resumo, estimativa de uma única componente de variância seguirá distribuição χ^2 . É nisso que consiste o denominador no teste F de modelos lineares simples. No teste F de modelos lineares mistos, o denominador é um tipo de combinação linear complexa de diferentes estimativas de componentes de variância. A distribuição de tais combinações lineares de diferentes distribuições χ^2 não é teoricamente uma χ^2 , MAS acontece que pode ser bem aproximada por uma, se o GL for escolhido adequadamente (para corresponder à média e à variância da combinação linear).

5.5 Tabelas ANOVA parciais e sequenciais

- Uma tabela ANOVA: Uma lista de valores SQ para cada efeito/termo no modelo
- Um valor SQ expressa a quantidade (bruta) de variação Y explicada pelo efeito.
- Construído sucessivamente (sequencialmente) (Em SAS: TIPO I, Em R: dado por “anova”)
- Parcialmente construído (Em SAS: TIPO III, Em R: dado por `drop1` OU `Anova(car)`)
- (Também Tipo II e IV - mas MAIS sutil - e NÃO tão importante)
- Geralmente: Eles (podem) testar diferentes hipóteses - definidas pela estrutura de dados (por exemplo, contagens de células)
- Construído sucessivamente (sequencialmente) (Tipo I)
 - Cada SQ é corrigido SOMENTE para os efeitos listados ANTES do efeito atual (dada alguma ordem dos efeitos)
 - A ordem vem da forma como o modelo é expresso na expressão real do modelo R/SAS.
 - Esses valores SQ somam-se ao valor SQ total dos dados.
 - E portanto: Eles fornecem uma decomposição real da variabilidade total dos dados.
 - O último SQ da tabela é também o SQ Tipo III para esse efeito
- Parcialmente construído (Tipo III)
 - Cada SQ é corrigido para TODOS os efeitos no modelo

- NÃO dependem da forma como o modelo é expresso
- Esses valores SQ geralmente NÃO serão somados ao valor SQ total dos dados.
- E portanto: NÃO fornecem uma decomposição real da variabilidade total dos dados.
- SOMENTE se os dados estiverem balanceados (por exemplo, todas as contagens de células n_{ij} são iguais para uma configuração $A \times B$ bidirecional):
 - O Tipo I e o Tipo III são iguais para todos os efeitos!
- De outra forma:
 - O Tipo I e o Tipo III NÃO são iguais!
- Geralmente: preferimos o Tipo III
- O tipo I às vezes é uma maneira conveniente de observar certas decomposições (por exemplo, em situações puramente aninhadas)
- Quando olhamos para o “resumo do modelo” em R: os testes correspondem aos testes do “Tipo III” (tendo sido corrigidos para todo o resto)

5.6 Intervalos de confiança de efeitos fixos

Intervalos de confiança baseados na distribuição t -aproximada podem ser aplicados para combinações lineares dos efeitos fixos. Quando um único parâmetro de efeito fixo ou uma única combinação linear estimável de parâmetros de efeito fixo é considerado, a matriz \mathbf{L} tem apenas uma coluna, e o intervalo de confiança de 95% torna-se:

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

Aqui a matriz de covariância \mathbf{V} não é conhecida, mas é baseada em estimativas de parâmetros de variância. O único problema que resta é determinar os graus de liberdade apropriados *g.l.*. Mais uma vez recomenda-se a aproximação de Satterthwaite. (Uma alternativa à aproximação de Satterthwaite é o método Kenward-Rogers, que é usado pela função Anova do pacote `car` e pelo pacote `lsmeans`)

5.7 Teste para parâmetros de efeitos aleatórios

O teste da razão de verossimilhança restrita/residual pode ser usado para testar a significância de um parâmetro de efeitos aleatórios. O teste da razão de verossimilhança é usado para comparar dois modelos A e B , sendo B um submodelo de A . Aqui, o modelo incluindo algum parâmetro de variância (modelo A) e o modelo sem este parâmetro de variância (modelo B) devem ser comparados.

- 1) Calcular os dois valores negativos de log-verossimilhança restrito/residual ($\ell_{re}^{(A)}$ e $\ell_{re}^{(B)}$) executando ambos os modelos.
- 2) Calcule a estatística de teste:

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

Teoricamente, tais estatísticas baseadas em log-verossimilhança seguirão assintoticamente $G_{A \rightarrow B}$ uma distribuição χ_1^2 . (Um grau de liberdade, porque um parâmetro de variância é testado ao comparar A com B). No entanto, o fato de que a mudança do modelo mais geral para o modelo mais específico envolve definir a variância de certos componentes dos efeitos aleatórios para zero, o que é o limite da região do parâmetro, os resultados assintóticos para LRT (LRT segue assintoticamente a distribuição com um grau de liberdade) devem ser ajustados para condições de contorno. Seguindo Self & Liang (1987); Stram & Lee (1994) o LRT segue mais de perto uma mistura igual de distribuições χ^2 com zero graus de liberdade (uma distribuição de massa pontual ou degenerada) e um grau de liberdade, isto é $0,5\chi_0^2 + 0,5\chi_1^2$ (ver Pinheiro e Bates, 2000, p. 86). O valor- p deste teste pode ser obtido reduzindo-o pela metade.

Observação 4.8

É importante, ao fazer testes baseados em REML como este, que a parte fixa dos dois modelos aninhados seja a mesma - caso contrário, as probabilidades de REML não são comparáveis.

5.8 Intervalos de confiança para parâmetros de efeitos aleatórios

O intervalo de confiança para um determinado parâmetro de variância é baseado na suposição de que a estimativa do parâmetro de variância $\hat{\sigma}_b^2$ é aproximadamente $\frac{\sigma^2}{gl} \chi_{gl}^2$ distribuída. Isto é verdade em casos balanceados (e outros “legais”). Uma consequência disso é que o intervalo de confiança assume a forma:

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

mas com os graus de liberdade $g.l.$ ainda indeterminados.

A tarefa é escolher o $g.l.$ tal que a distribuição χ_{gl}^2 correspondente corresponda à distribuição da estimativa. A variância (teórica) de $\frac{\sigma^2}{gl} \chi_{gl}^2$ é:

$$\text{Var}\left(\frac{\sigma^2}{gl} \chi_{gl}^2\right) = \frac{2\sigma^4}{gl}$$

A variância real (assintótica) do parâmetro pode ser estimada a partir da curvatura da função de log de verossimilhança negativa ℓ (pela teoria geral de máxima verossimilhança). Combinando a variância real estimada do estimador com a variância da distribuição desejada e resolvendo a equação:

$$\text{Var}(\hat{\sigma}_b^2) = \frac{2\sigma^4}{gl}$$

a seguinte estimativa dos graus de liberdade é obtida, após inserir a variância estimada:

$$\hat{gl} = \frac{2\hat{\sigma}^4}{\text{Var}(\hat{\sigma}_b^2)}$$

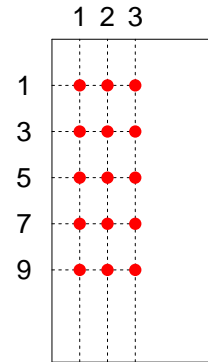
Esta forma de aproximar os graus de liberdade é um caso especial da aproximação de Satterthwaite.

5.9 R-TUTORIAL: Testando efeitos aleatórios

Algumas observações sobre como testar efeitos fixos e aleatórios em R são fornecidas. Como normalmente a estrutura aleatória deve ser determinada antes de olhar para a estrutura fixa, começamos com os efeitos aleatórios. Os dados das pranchas dos Módulos 3 e 6 são usados como exemplo:

Para investigar o efeito do método de secagem na porcentagem de umidade em madeira de faia, foi realizado o seguinte experimento. Um total de 20 tábuas foram secas em um determinado período de tempo. Em seguida, a porcentagem de umidade foi medida em 5 profundidades e 3 larguras para cada tábua:

- profundidade 1: perto do topo
- profundidade 5: no centro
- profundidade 9: perto do fundo
- profundidade 3: entre 1 e 5
- profundidade 7: entre 5 e 9
- largura 1: próximo ao lado
- largura 3: no centro
- largura 2: entre 1 e 3



Portanto, existem $3 \times 5 = 15$ medições para cada prancha e, no total, 300 observações. Os dados podem ser encontrados em pranchas e são reproduzidos na tabela a seguir.

	plank	width	depth	humidity	loghum
1	1	1	1	3.4	1.22
2	1	1	3	4.9	1.59
...	<NA>	<NA>	<NA>
297	20	3	3	6	1.79
298	20	3	5	6	1.79
299	20	3	7	6.3	1.84
300	20	3	9	4.2	1.44

Variável	Descrição
prancha	Numerado de 1 a 20
largura	Numerado 1,2,3
profundidade	Numerado 1,3,5,7,9
umidade	porcentagem de umidade

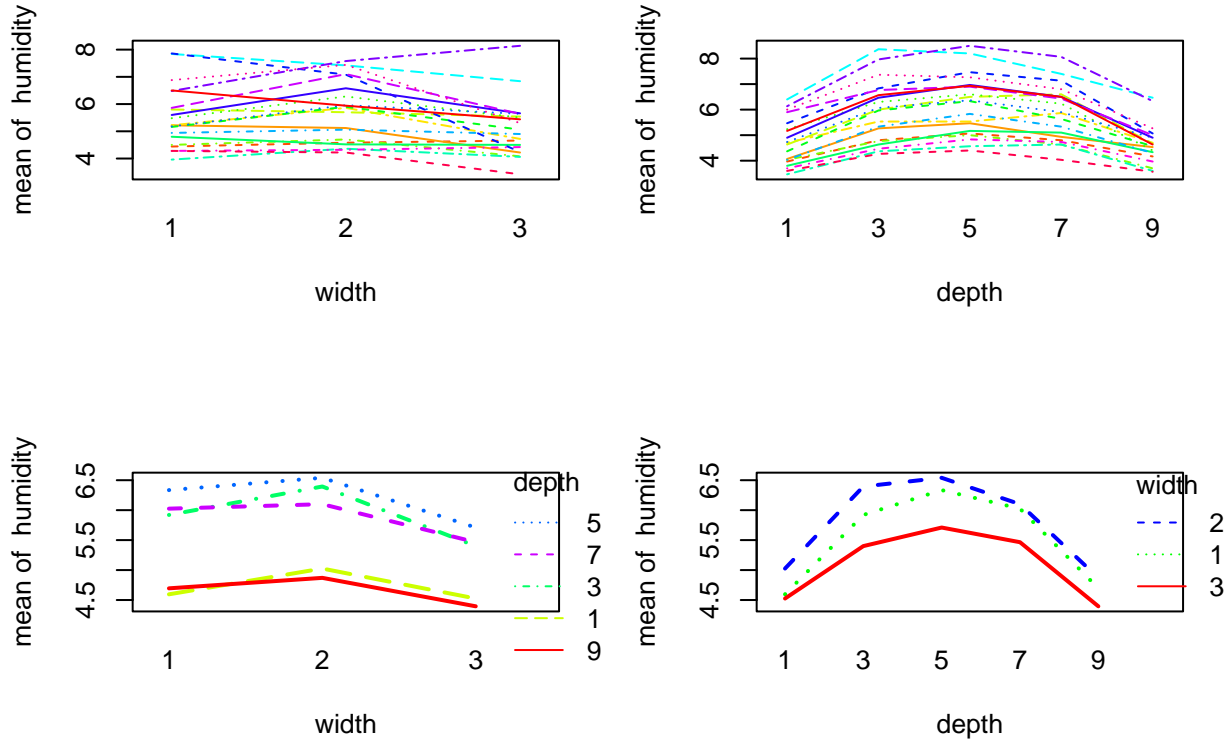


Figura 1: Quatro perfis de umidade média.

Fica imediatamente claro que existe grande variação no nível de umidade de prancha para prancha. Na Fig. 1.a, não está evidente que há efeito largura. Já na Fig. 1.b está bem evidente que a umidade é alta no centro (profundidade=5) e baixa na parte superior (profundidade=1) e na parte inferior (profundidade=9), efeito observado quando se calcula a média para as três larguras, Fig. 1.d.

Pode acontecer que o efeito de profundidade seja diferente para larguras próximas da lateral da prancha (largura=1) e para larguras no centro (largura=3), Fig. 1.d. Em outras palavras, poderia haver um efeito de interação prancha*largura, que não encontraríamos nos gráficos acima. A Fig. 1.c, confirma a diferença de umidade na profundidades 1 e 9 em relação as demais. Além disso, o padrão de umidade em profundidade parece ser aproximadamente o mesmo para as três larguras.

O modelo correspondente, expresso formalmente, será:

$$\log y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + d_k + f_{ik} + g_{jk} + \varepsilon_{ijk}, \quad \begin{cases} d_k \sim N(0, \sigma_{plank}^2) \\ f_{ik} \sim N(0, \sigma_{plank*with}^2) \\ g_{jk} \sim N(0, \sigma_{plank*debth}^2) \\ \varepsilon_{ijk} \sim N(0, \sigma^2) \end{cases}$$

em que:

- μ denota a média geral;
- α_i o efeito da i -ésima largura (**width**);
- β_j o efeito da j -ésima profundidade (**depth**);
- γ_{ij} o efeito da ij -ésima interação de largura:profundidade (**width*depth**);
- d_k o efeito da k -ésima **aleatório** de prancha (**plank**);
- f_{ik} o efeito da ik -ésima **aleatório** de interação prancha:largura (**plank*width**);
- g_{jk} o efeito da jk -ésima **aleatório** de interação prancha:profundidade (**plank*debth**);
- $d_k, f_{ik}, g_{jk}, \varepsilon_{ijk}$ independentes.

é ajustado como segue:

```
library(lme4)

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

Linear mixed model fit by REML ['lmerMod']

Formula: loghum ~ depth * width + (1 | plank) + (1 | depth:plank) + (1 | plank:width)

Data: planks

REML criterion at convergence: -464.5

Scaled residuals:

	Min	1Q	Median	3Q	Max
	-2.7463	-0.5217	-0.0058	0.5665	2.0883

Random effects:

Groups	Name	Variance	Std.Dev.
depth:plank	(Intercept)	0.000607	0.0246
plank:width	(Intercept)	0.008184	0.0905

```

plank      (Intercept) 0.028834 0.1698
Residual                    0.004933 0.0702
Number of obs: 300, groups:  depth:plank, 100; plank:width, 60; plank, 20

```

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1.50390	0.04613	32.60
depth3	0.25135	0.02354	10.68
depth5	0.31888	0.02354	13.55
depth7	0.27316	0.02354	11.61
depth9	0.02965	0.02354	1.26
width2	0.08757	0.03622	2.42
width3	-0.01280	0.03622	-0.35
depth3:width2	-0.00909	0.03141	-0.29
depth5:width2	-0.05288	0.03141	-1.68
depth7:width2	-0.07758	0.03141	-2.47
depth9:width2	-0.05349	0.03141	-1.70
depth3:width3	-0.07917	0.03141	-2.52
depth5:width3	-0.08960	0.03141	-2.85
depth7:width3	-0.09002	0.03141	-2.87
depth9:width3	-0.05907	0.03141	-1.88

A matriz estimada de componentes de variância dos efeitos aleatórios será:

$$\hat{\mathbf{G}} = \begin{pmatrix} \hat{\sigma}_{plank}^2 & 0 & 0 \\ 0 & \hat{\sigma}_{depth:plank}^2 & 0 \\ 0 & 0 & \hat{\sigma}_{plank:width}^2 \end{pmatrix} = \begin{pmatrix} 0.0288^2 & 0 & 0 \\ 0 & (6.0687 \times 10^{-4})^2 & 0 \\ 0 & 0 & 0.0082^2 \end{pmatrix}$$

enquanto a matriz de variância covariância dos resíduos é dada por:

$$\hat{\mathbf{R}} = \begin{pmatrix} \hat{\sigma}^2 & 0 & \dots & 0 \\ 0 & \hat{\sigma}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} 0.0702^2 & 0 & \dots & 0 \\ 0 & 0.0702^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0.0702^2 \end{pmatrix}$$

O teste REML para a hipótese

$$\begin{cases} H_0 : \sigma_{plank*depth}^2 = 0 \\ H_1 : \sigma_{plank*depth}^2 > 0 \end{cases}$$

é realizado ajustando o submodelo e comparando-o com o modelo completo:

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

Data: planks

Models:

```
model4.1: loghum ~ depth * width + (1 | plank) + (1 | plank:width)
model4: loghum ~ depth * width + (1 | plank) + (1 | depth:plank) + (1 | plank:width)
      npar  AIC   BIC logLik deviance Chisq Df Pr(>Chisq)
model4.1   18 -426 -359    231     -462
model4     19 -427 -356    232     -465  2.58  1      0.11
```

```
valor.p<- 0.5*pchisq(tb[2,6], df = 0) + 0.5*pchisq(tb[2,6], df = 1)
```

Note que neste caso esta sendo considerado que a distribuição da estatística de teste é χ_1^2 e não como recomendado por Pinheiro e Bates (2000). Para a mistura de distribuições $0,5\chi_0^2 + 0,5\chi_1^2$ teremos:

```
(valor.p<- 0.5*pchisq(tb[2,6], df = 0, lower.tail = FALSE) +
0.5*pchisq(tb[2,6], df = 1, lower.tail = FALSE))
```

[1] 0.05413

Como mencionado anteriormente, essa mistura de distribuições reduz o valor-p pela metade, nesse caso específico.

E da mesma forma para o outro componente aleatório de interação:

$$\begin{cases} H_0 : \sigma_{plank*width}^2 = 0 \\ H_1 : \sigma_{plank*width}^2 > 0 \end{cases}$$

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

Data: planks

Models:

```
model4.2: loghum ~ depth * width + (1 | plank) + (1 | depth:plank)
model4: loghum ~ depth * width + (1 | plank) + (1 | depth:plank) + (1 | plank:width)
      npar  AIC  BIC logLik deviance Chisq Df Pr(>Chisq)
model4.2   18 -317 -250    176     -353
model4     19 -427 -356    232     -465   112  1    <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Considerando mistura de distribuições:
```

```
(valor.p<- 0.5*pchisq(tb[2,6], df = 0, lower.tail = FALSE) +
          0.5*pchisq(tb[2,6], df = 1, lower.tail = FALSE))
```

```
[1] 1.932e-26
```

Observe a opção lógica `refit` indicando se os modelos devem ser reajustados com ML antes de comparar os modelos. O padrão é TRUE para evitar o erro comum de comparar inadequadamente modelos ajustados por REML com diferentes efeitos fixos, cujas probabilidades não são diretamente comparáveis.

Agora compare com a tabela ANOVA da análise de efeitos fixos também fornecida na seção de exemplos do módulo 06:

```
      Df Sum Sq Mean Sq F value Pr(>F)
depth    4   4.28   1.071  217.14 <2e-16 ***
width    2   0.80   0.399   80.86 <2e-16 ***
plank    19   9.12   0.480   97.34 <2e-16 ***
depth:width  8   0.09   0.011    2.25  0.027 *
depth:plank 76   0.51   0.007    1.37  0.052 .
width:plank 38   1.74   0.046    9.29 <2e-16 ***
Residuals 152   0.75   0.005
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As estatísticas F (e valores P) fazem sentido somente para os testes de interação de ordem. Aqui também obtemos testes para os dois efeitos de interação aleatórios (com resultados semelhantes, mas não 100% de resultados equivalentes). Essas estatísticas F test são formalmente testes no modelo fixo, MAS pode-se argumentar que funcionam como testes também no modelo aleatório. Elas são baseados na distribuição F , que é uma propriedade distributiva EXATA (se o modelo estiver correto), enquanto os testes χ^2 do modelo aleatório/misto realizados acima são baseados em resultados distribucionais APROXIMADOS. Observe que

é perfeitamente correto que os testes χ^2 tenham apenas $GL = 1$, já que apenas um único parâmetro é testado nesses testes! Então, em resumo: A abordagem aproximada do teste χ^2 dada neste módulo teórico pode ser aplicada para qualquer (conjunto de) parâmetros de estrutura aleatória em qualquer modelo misto e é, portanto, um método de aplicação geral. Em ALGUNS casos, podemos encontrar/construir estatísticas F com propriedades distributivas potencialmente melhores.

Cuidado: Aqui o autor está dizendo que quando assumimos uma única distribuição χ^2 para a estatística dos testes realizados para parâmetros na fronteira de seu espaço paramétrico, estamos fazendo uma aproximação!!!

Um recurso interessante sobre o pacote `lmerTest` é que ele oferece uma investigação automatizada passo a passo da importância de todos os efeitos aleatórios:

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

Backward reduced random-effect table:

	Eliminated	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
<none>		19	232	-427			
(1 depth:plank)	1	18	231	-426	2.6	1	0.11
(1 plank)	0	17	213	-391	36.6	1	1.5e-09 ***
(1 plank:width)	0	17	176	-319	109.3	1	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

5.10 R-TUTORIAL: Testando efeitos fixos

O teste de efeitos fixos, dada alguma estrutura de modelo aleatório em um modelo misto, é um tópico de discussão e existem várias abordagens. Em primeiro lugar, já vimos no Módulo 1 e acima como poderíamos usar a função `estimável` para fornecer comparações post hoc no modelo misto. Estes são exemplos dos testes de Wald apresentados neste módulo. Também vimos o uso de `lsmeans` e do pacote `multcomp`. Finalmente, vimos que se quisermos os testes globais para interação ou efeitos principais, como geralmente dados na tabela ANOVA de um modelo puramente fixo, poderíamos obtê-los, por exemplo, de a função `Anova` do pacote `car` ou a função `anova` do pacote `lmerTest`:

```
anova(model4)
```

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
depth	3.130	0.782	4	76	158.61	< 2e-16 ***
width	0.086	0.043	2	38	8.70	0.00077 ***
depth:width	0.089	0.011	8	152	2.25	0.02678 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Se compararmos novamente com a tabela ANOVA fixa, seremos capazes de encontrar os dois efeitos principais do teste F neste caso como:

$$F_{depth} = \frac{QM_{depth}}{QM_{depth*plank}} \quad \text{e} \quad F_{width} = \frac{QM_{width}}{QM_{width*plank}}$$

e a interação diretamente como:

$$F_{depth:width} = \frac{QM_{depth:width}}{QM_{Error}}$$

Este é um exemplo de como a abordagem geral baseada na verossimilhança REML levará a algo simples que poderíamos ter percebido, se estivéssemos familiarizados com o chamado método baseado em Henderson e a teoria mista ANOVA da Esperança do Quadrado Médio do Erro - Expected Mean Square (EMS). A força da abordagem geral do modelo misto é que ela fará a coisa certa mesmo que percamos o balanceamento/completude dos dados. E nesses casos simples TAMBÉM dá certo.

5.11 R-TUTORIAL: Estimando efeitos aleatórios com ICs

Fazemos isso muito brevemente. O método baseado em Satterthwaite descrito acima requer que possamos extrair a chamada matriz de variância-covariância assintótica para os parâmetros aleatórios. A teoria da verossimilhança nos diz que isso é encontrado a partir das segundas derivadas (Hessiana) da função log-verossimilhança (RE)ML com os parâmetros estimados inseridos. Infelizmente, no R isso não é facilmente extraído de um objeto resultante de lmer.

Em vez disso, pode-se obter facilmente os chamados intervalos de confiança baseados no perfil de probabilidade:


```
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

```
library(parameters)
```

```
ci(model4)
```

	Parameter	CI	CI_low	CI_high
1	(Intercept)	0.95	1.41309	1.594700
2	depth3	0.95	0.20502	0.297682
3	depth5	0.95	0.27255	0.365217
4	depth7	0.95	0.22683	0.319495
5	depth9	0.95	-0.01669	0.075980
6	width2	0.95	0.01627	0.158858
7	width3	0.95	-0.08409	0.058494
8	depth3:width2	0.95	-0.07092	0.052739
9	depth5:width2	0.95	-0.11471	0.008950
10	depth7:width2	0.95	-0.13942	-0.015753
11	depth9:width2	0.95	-0.11532	0.008345
12	depth3:width3	0.95	-0.14100	-0.017339
13	depth5:width3	0.95	-0.15143	-0.027764
14	depth7:width3	0.95	-0.15185	-0.028184
15	depth9:width3	0.95	-0.12091	0.002757

```
model_parameters(model4, df_method = "satterthwaite")
```

Fixed Effects

Parameter	Coefficient	SE	95% CI	t(281)	p
(Intercept)	1.50	0.05	[1.41, 1.59]	32.60	< .001
depth [3]	0.25	0.02	[0.21, 0.30]	10.68	< .001
depth [5]	0.32	0.02	[0.27, 0.37]	13.55	< .001
depth [7]	0.27	0.02	[0.23, 0.32]	11.61	< .001
depth [9]	0.03	0.02	[-0.02, 0.08]	1.26	0.209
width [2]	0.09	0.04	[0.02, 0.16]	2.42	0.016
width [3]	-0.01	0.04	[-0.08, 0.06]	-0.35	0.724
depth [3] × width [2]	-9.09e-03	0.03	[-0.07, 0.05]	-0.29	0.772
depth [5] × width [2]	-0.05	0.03	[-0.11, 0.01]	-1.68	0.093
depth [7] × width [2]	-0.08	0.03	[-0.14, -0.02]	-2.47	0.014
depth [9] × width [2]	-0.05	0.03	[-0.12, 0.01]	-1.70	0.090
depth [3] × width [3]	-0.08	0.03	[-0.14, -0.02]	-2.52	0.012
depth [5] × width [3]	-0.09	0.03	[-0.15, -0.03]	-2.85	0.005
depth [7] × width [3]	-0.09	0.03	[-0.15, -0.03]	-2.87	0.004
depth [9] × width [3]	-0.06	0.03	[-0.12, 0.00]	-1.88	0.061

Random Effects

Parameter	Coefficient
SD (Intercept: depth:plank)	0.02
SD (Intercept: plank:width)	0.09
SD (Intercept: plank)	0.17
SD (Residual)	0.07

```
# library(lmeresampler)
# system.time(
#   bb3 <- bootstrap(model4, .f = f1, type = "reb", reb_type = 0, B=1000)
# ) ## 21 seconds
# ## reb = "random effects block bootstrap"
# ## read the vignette("lmeresampler") for details on reb_type
#
# confint(bb3, type = "perc")
```

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \delta_k + \varepsilon_{ijk}, \quad \begin{cases} i = 1, 2, 3 = a \\ j = 1, \dots, 5 = J \\ k = 1, \dots, 20 = K \end{cases}$$

Tabela 1: Resumo das informações para comparação

F.V.	SQ	$G.L.$	$\mathbb{E}(QM)$	
			δ fixo	δ aleatório
Planks	$\frac{1}{aJ} \sum_k^K y_{..k}^2 - \frac{y_{...}^2}{aJK}$	$K - 1$	$\sigma^2 + \frac{aJ \sum \delta_k^2}{K-1}$	$\sigma^2 + aJ\sigma_\delta^2$
width	$\frac{1}{JK} \sum_i^a y_{i..}^2 - \frac{y_{...}^2}{aJK}$	$a - 1$	$\sigma^2 + \frac{JK \sum \alpha_i^2}{a-1}$	
depth	$\frac{1}{aK} \sum_j^J y_{.j.}^2 - \frac{y_{...}^2}{aJK}$	$J - 1$	$\sigma^2 + \frac{aK \sum \beta_j^2}{a-1}$	
width:depth	$\frac{1}{K} \sum_i^a \sum_j^J y_{ij.}^2 - \frac{y_{...}^2}{aJK}$	$(a-1)(J-1)$	$\sigma^2 + \frac{K \sum \sum \gamma_{ij}^2}{(a-1)(J-1)}$	
Erro	subtração	$(aJ-1)(k-1)$	σ^2	