

Universidade de São Paulo

Escola Superior de Agricultura Luiz de Queiroz - ESALQ

Regressão Aleatória

Aluna: Pollyane Vieira da Silva

Disciplina: LCE5801 - Regressão e Covariância

Profª Drª Taciana Villela Savian

2016

Sumário

1	Modelos mistos	3
2	Derivação das equações de modelos mistos	4
2.1	Soluções para os efeitos fixos e predições dos efeitos aleatórios	5
2.2	Algumas propriedades das soluções para os efeitos fixos	7
2.3	Algumas propriedades da predição para os efeitos aleatórios .	10
3	Discussão	12
4	Exemplo ilustrativo	14
5	Conclusão	20

Introdução

Um modelo linear que apresenta somente fatores de efeitos fixos, além do erro experimental, que é sempre aleatório, é denominado modelo fixo. Esse tipo de modelo já foi amplamente estudado, existindo inúmeros livros abordando seus aspectos teóricos e aplicados, em vários níveis de complexidade. Os modelos que apresentam apenas fatores de efeitos aleatórios, exceto a constante μ , que é sempre fixa, é denominado modelo aleatório.

Um modelo misto é aquele que apresenta tanto fatores de efeitos fixos como aleatórios, além do erro experimental e da constante μ . Quando um modelo é considerado misto, sua análise de variância apresenta algumas peculiaridades, como a composição das esperanças matemáticas dos quadrados médios, cujo conhecimento permite o estabelecimento correto dos testes de hipóteses.

Outro motivo de se adotar um modelo linear misto é a possibilidade de se fazer a predição de efeitos aleatórios, na presença de efeitos fixos, através dos BLUP's (best linear unbiased prediction) que são de grande valia em genética e melhoramentos.

1 Modelos mistos

Matricialmente, o modelo misto geral é denotado por:

$$y = X\beta + Z\gamma + e$$

em que,

${}_n y_1$ é o vetor de observações;

${}_n X_{p+1}$ é a matriz de incidência dos efeitos fixos (conhecida);

${}_{p+1} \beta_1$ é o vetor de efeitos fixos desconhecidos;

${}_n Z_q$ é a matriz de incidência dos efeitos aleatórios (conhecida);

${}_q \gamma_1$ é o vetor de efeitos aleatórios desconhecidos;

${}_n e_1$ é o vetor de erros aleatórios.

Assumindo-se que os efeitos aleatórios e os erros (resíduos) têm distribuição normal com média zero e são não correlacionados, com matrizes de variâncias e covariâncias dadas por:

$$Var(\gamma) = E(\gamma\gamma') = D$$

e

$$Var(e) = E(ee') = R$$

Deste modo, tem-se que:

$$V = Var(y) = Var(X\beta + Z\gamma + e) = ZDZ' + R$$

Assume-se ainda que V é não singular, e

$$E(y) = E(X\beta + Z\gamma + e) = X\beta,$$

assim,

$$y \sim N(X\beta; ZDZ' + R)$$

De acordo com Sheffé (1959), o modelo misto foi amplamente estudado por Fisher em 1958, com grande repercussão no estudo de genética quantitativa. Tal modelo foi denominado pelo autor de modelo de componente de variância.

O enquadramento de efeitos como fixos ou aleatórios está relacionado ao objetivo da análise: se os níveis do efeito constituem-se amostras de uma população sobre a qual se quer tirar conclusões (efeitos aleatórios) ou se são parâmetros constantes sobre os quais se quer identificar diferenças e magnitudes (efeitos fixos).

Considerar fatores como fixos ou aleatórios pode afetar a estimação e suas consequências pela presença das variâncias das variáveis aleatórias perturbadoras. Assim, ignorar as covariâncias entre as observações de um mesmo nível da variável aleatória pode, também, afetar a qualidade da predição devido à alteração na ordem dos valores das predições dos efeitos aleatórios ou pela simples mudança na magnitude desses valores.

Um dos problemas centrais no ajuste de modelos mistos a um conjunto de observações é a estimação dos componentes de variância e covariâncias dos efeitos aleatórios e qual a metodologia mais adequada para a estimação dos efeitos fixos.

2 Derivação das equações de modelos mistos

A derivação das equações de modelos mistos pode ser feita pela minimização da soma de quadrados dos resíduos ou pela maximização da função densidade de probabilidade conjunta de y e γ . Aqui será adotada a segunda forma, considerando que a distribuição seja normal. A função densidade de probabilidade de y é dada por:

$$f(y, \gamma) = \frac{1}{(2\pi)^{n/2} (ZDZ' + R)^{1/2}} \exp[-1/2[(y - X\beta)'(ZDZ' + R)^{-1}(y - X\beta)]$$

A função densidade de probabilidade conjunta de y e γ pode ser escrita como o produto entre a função densidade condicional de y , dado γ , e a função densidade de probabilidade de γ .

$$f(y, \gamma) = f(y|\gamma)f(\gamma)$$

$$f(y, \gamma) = \frac{1}{(2\pi)^{n/2}(R)^{1/2}} \exp[-1/2[(y-X\beta-Z\gamma)'(R)^{-1}(y-X\beta-Z\gamma)]] \frac{1}{(2\pi)^{n/2}(D)^{1/2}} \exp[-1/2[(\gamma-\beta)'$$

Para se proceder à maximização de $f(y, \gamma)$, pode-se usar o artifício da transformação por logaritmo. Isso é possível, visto que, sendo $f(y, \gamma)$ e $\log f(y, \gamma)$ funções contínuas e crescentes no espaço \mathbb{R}^+ , seus pontos de máximo são coincidentes dentro do espaço de $[\beta' \gamma']$ e $ZDZ' + R$. Assim, fazendo-se $L = \log f(y, \gamma)$, tem-se:

$$L = \frac{1}{2} 2n \log(2\pi) - \frac{1}{2} (\log R + \log D) - \frac{1}{2} (y'R^{-1}y - 2y'R^{-1}X\beta - 2y'R^{-1}Z\gamma + 2\beta'X'R^{-1}Z\gamma + \beta'X'R^{-1}X\beta + \gamma'D^{-1}\gamma)$$

Derivando L em relação a β e γ , e tornando-se tais derivadas identicamente nulas, obtêm-se:

$$\begin{bmatrix} \frac{\partial L}{\partial \beta} \\ \frac{\partial L}{\partial \gamma} \end{bmatrix} = \begin{bmatrix} -X'R^{-1}y + X'R^{-1}X\beta_0 + X'R^{-1}Z\hat{\gamma} \\ -ZR^{-1}y + Z'R^{-1}X\beta_0 + Z'R^{-1}Z\hat{\gamma} + D^{-1}\hat{\gamma} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} X'R^{-1}X\beta_0 + X'R^{-1}Z\hat{\gamma} \\ -ZR^{-1}X\beta_0 + Z'R^{-1}Z\hat{\gamma} + D^{-1}\hat{\gamma} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ -ZR^{-1}X & Z'R^{-1}Z\hat{\gamma} + D^{-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

Essas são as equações de modelos mistos (MME), que permitem obter soluções para os efeitos fixos (β_0) e predições para os efeitos aleatórios ($\hat{\gamma}$).

2.1 Soluções para os efeitos fixos e predições dos efeitos aleatórios

A solução do sistema de equações de modelos mistos pode ser obtida por absorção ou por obtenção da matriz inversa por partição. Em ambos os casos, os resultados serão:

$$\beta_0 = \{X'[R^{-1} - R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}]X\}^{-1} \quad (1)$$

$$X'[R^{-1} - R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}]y \quad (2)$$

$$\hat{\gamma} = (ZR^{-1}Z + D^{-1})^{-1}Z'R^{-1}(y - X\beta_0)$$

Outra alternativa para se obter soluções para os efeitos fixos é pelo uso de um modelo linear generalizado, ignorando-se os efeitos aleatórios, como a seguir:

Dado o modelo

$$y = X\beta + Z\gamma + e$$

anteriormente descrito, e com $Var(y) = ZDZ' + R$, tem-se que o sistema de equações normais generalizada é dado por:

$$X'V^{-1}X\beta_0 = X'V^{-1}y$$

cuja solução é:

$$\beta_0 = (X'V^{-1}X)^{-1}X'V^{-1}y$$

e a predição de γ seria obtida por:

$$\hat{\gamma} = DZ'V^{-1}(y - X\beta_0)$$

$$V^{-1} = R^{-1} - R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}$$

A desvantagem de se utilizar a segunda opção, que envolve o cálculo de V^{-1} é de ordem computacional, uma vez que a dimensão de V é igual ao número de observações, que muitas das vezes, principalmente na área de melhoramento genético, chega a ser de algumas centenas. No caso de modelos fixos, V usualmente assume a forma $I_n\sigma^2$ ou, é pelo menos diagonal. Nesse caso a obtenção de V^{-1} é simples. Mas em geral, $V = ZDZ' + R$ não é diagonal, e deste modo a obtenção de V^{-1} não é fácil.

Obter $R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R_{-1}$ é mais simples. Pois R^{-1} pode ser facilmente obtida por $I \otimes R_0^{-1}$, onde R_0 é a matriz de variância e covariância residual qxq , entre as q médias que compõem uma observação. D^{-1} por $A_{-1} \otimes D_0^{-1}$, em que D_0 é a matriz de variância e covariância, qxq , entre os efeitos aleatórios nas q medidas que compõem uma observação, e A é a matriz de correlação, nxn , entre os efeitos aleatórios das n observações. Apesar da matriz A não possuir estrutura simples, como ocorre na maioria das vezes, para aplicações em melhoramento animal, existem algoritmos eficientes para obtenção direta de A^{-1} . Mesmo assim persiste a necessidade de se obter $(Z'R^{-1}Z + D^{-1})^{-1}$, que, a despeito de possuir as mesmas dimensões de V , pode ser obtida por processos iterativos com a vantagem de rápida convergência em razão da dominância dos elementos da diagonal causada pela adição de D^{-1} a $Z'R^{-1}Z$. Nos casos de distribuição multivariada, elementos dominantes podem estar fora da diagonal. Nesses casos, processos que usam iteração em blocos garantem a rápida convergência, porque os elementos dominantes passarão a estar nos blocos.

2.2 Algumas propriedades das soluções para os efeitos fixos

- a) A solução β_0 , obtida pelas MME é também uma solução de mínimos quadrados generalizados (GLS), utilizando o modelo que ignora os efeitos aleatórios.

Demonstração:

Foi visto que uma solução de mínimos quadrados generalizados para $y = X\beta + e$:

$$\beta_0 = (X'V^{-1}X)^{-1}X'V^{-1}y$$

Das equações de modelos mistos (MME),

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ -ZR^{-1}X & Z'R^{-1}Z\hat{\gamma} + D^{-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

tem-se que:

$$\beta_0 = X'[R^{-1} - R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}]X^{-1}X'[R^{-1} - R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}]y$$

$$\hat{\gamma} = (Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}(y - X\beta_0)$$

substituindo $\hat{\gamma}$ em:

$$X'R^{-1}X\beta_0 + X'R^{-1}Z\hat{\gamma} = X'R^{-1}y$$

,

tem-se:

$$X'R^{-1}X\beta_0 + X'R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}(y - X\beta_0) = X'R^{-1}y$$

$$X'R^{-1}X\beta_0 + X'R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}y - X'R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}X\beta_0 = X'R^{-1}y$$

$$[X'R^{-1}X - X'R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}X]\beta_0 = [X'R^{-1} - X'R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}]y$$

$$X'[R^{-1} - R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}]X\beta_0 = X'[R^{-1} - R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}]y$$

$$\beta_0 = \{X'[R^{-1} - R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}]\}y$$

então,

$$\beta_0 = (X'V^{-1}X)^{-}X'V^{-1}y$$

b) A variância de β_0 é dada por:

$$\begin{aligned} Var[\beta_0] &= Var[(X'V^{-1}X)^{-}X'V^{-1}y] \\ Var[\beta_0] &= (X'V^{-1}X)^{-}X'V^{-1}Var[y]V^{-1}X(X'V^{-1}X)^{-} \\ Var[\beta_0] &= (X'V^{-1}X)^{-}X'V^{-1}VV^{-1}X(X'V^{-1}X)^{-} \\ Var[\beta_0] &= (X'V^{-1}X)^{-}X'V^{-1}X(X'V^{-1}X)^{-} \end{aligned}$$

Uma vez que $(X'V^{-1}X)$ é uma matriz simétrica, a escolha, apropriada de uma inversa generalizada também simétrica, leva a igualdade

$$Var[\beta_0] = (X'V^{-1}X)^{-}X'V^{-1}X(X'V^{-1}X)^{-} = (X'V^{-1}X)^{-}$$

e assim,

$$\begin{aligned} Var[\beta_0] &= (X'V^{-1}X)^{-} = \{X'[R^{-1} - R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}]X\}^{-} \\ Var[\beta_0] &= [X'R^{-1} - X'R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}X]^{-}. \end{aligned}$$

c) Para um dado conjunto p de funções estimáveis, linearmente independentes, estabelecidas por uma matriz conhecida λ , a variância de $\lambda'\beta_0$, (BLUE) de $\lambda'\beta$ é dada por:

$$\begin{aligned} Var[\lambda'\beta_0] &= \lambda'Var[\beta_0]\lambda \\ Var[\lambda'\beta_0] &= \lambda'[X'R^{-1} - X'R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}X]^{-}\lambda \end{aligned}$$

2.3 Algumas propriedades da predição para os efeitos aleatórios

a) O preditos $\hat{\gamma}$ é o Melhor Preditor Linear Não -viesado (BLUP) de γ .

O termo predição refere-se a fatores aleatórios e a Melhor Predição Linear Não - Viesada pode ser, resumidamente, definida como resultado da regressão dos efeitos de um fator aleatório (γ) em função das observações (y) corrigidas para os efeitos dos fatores fixos ($X\beta$), como dado na seguinte expressão;

$$\hat{\gamma} = DZ'(ZDZ' + R)^{-1}(y - X\beta_0) = DZ'V^{-1}(y - X\beta_0)$$

Observa-se que o termo $DZ'(ZDZ' + R)^{-1}$ é o conjunto de coeficientes de regressão de γ em função de y , uma vez que DZ' é a matriz de covariâncias entre γ e y .

$(ZDZ' + R)^{-1}$ é a inversa da matriz de variância de y , enquanto o termo $(y - X\beta_0)$ contém os valores das observações, y , corrigidas para os efeitos fixos $X\beta$.

Pelas MME, $\hat{\gamma}$ é dado por

$$\hat{\gamma} = (Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}(y - X\beta_0).$$

Então, se a igualdade:

$$DZ'(ZDZ' + R)^{-1} = (Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}$$

for verdadeira, $\hat{\gamma}$, obtido pela MME, é o BLUP de γ .

b) A variância de $\hat{\gamma}$ é dada por

$$Var[\hat{\gamma}] = Var[DZ'V^{-1}(y - X\beta_0)]$$

$$Var[\hat{\gamma}] = DZ'V^{-1}Var(y - X\beta_0)V^{-1}ZD'$$

$$Var[\hat{\gamma}] = DZ'V^{-1}[Var(y) - 2Cov(y, \beta_0'X') + Var(X\beta_0)]V^{-1}ZD'$$

mas $Cov(y, \beta_0'X') = Var(X\beta_0)$;então,

$$Var[\hat{\gamma}] = DZ'V^{-1}[Var(y) + Var(X\beta_0)]V^{-1}ZD'$$

$$Var[\hat{\gamma}] = DZ'V^{-1}[V - X(X'V^{-1}X)^{-1}X']V^{-1}ZD'$$

$$Var[\hat{\gamma}] = DZ'[V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}V^{-1}]ZD'$$

Pode-se notar que a expressão

$$V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}V^{-1}$$

é o complemento ortogonal de y no espaço coluna de X , o que significa que

$$[V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}V^{-1}]y = y - X\beta_0$$

obs. $V^{-1} = R^{-1} - R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}$

c) A variância do erro de predição é dada por:

$$Var[\gamma - \hat{\gamma}] = Var(\gamma) - 2Cov(\gamma, \hat{\gamma}') + Var(\hat{\gamma}),$$

mas, $Cov(\gamma, \hat{\gamma}') = Var(\hat{\gamma})$, então $Var[\gamma - \hat{\gamma}] = Var(\gamma) - Var(\hat{\gamma})$

$$= D - DZ'[V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}]ZD'$$

d) A correlação entre os valores reais e preditos é máxima.

Dentre uma classe de funções lineares que gera predições não viciadas, o BLUP maximiza a correlação $(\gamma - \hat{\gamma})$.

3 Discussão

As avaliações genéticas, tradicionalmente, têm sido realizadas por meio do modelo animal, em análise uni ou multicaracterísticas, sob a metodologia de modelos mistos. Os modelos de regressão aleatória (MRA), têm sido propostos como alternativa para avaliação genética de características que são medidas repetidamente na vida do animal ou são consideradas como função de qualquer variável contínua.

A principal diferença entre os MRA e os modelos de repetibilidade e multicaracterística está na inclusão de colunas de covariáveis na matriz de incidência dos efeitos aleatórios, em substituição as colunas de zero ou um. Dessa forma, considera-se que as variâncias podem ser alteradas com o tempo, além de permitirem a estimação de parâmetros individuais que consideram o parentesco entre os animais.

O MRA pressupõe que todos os registros, de um mesmo animal, se referem a uma mesma característica, porém assumindo diferentes correlações genéticas entre esses registros, considerando que o caráter pode estar mudando continua e gradualmente com o passar do tempo. Os MRA são considerados casos especiais de funções de covariância, os quais permitem estimar diretamente estes coeficientes pelo método da máxima verossimilhança restrita. As funções de covariância, as quais inicialmente foram desenvolvidas para o melhoramento de bovinos leiteiros, são uma forma de modelar variâncias e covariâncias ao longo do tempo e são descritas como regressões aleatórias em que as covariáveis são expressões padronizadas do tempo variando de -1 a 1. As funções de covariância podem ser obtidas por meio das matrizes de covariância entre os coeficientes de regressão estimadas pelos MRA.

Os MRA admitem a análise de registros repetidos de características que mudam gradativamente ao longo do tempo e não necessitam pressuposições quanto a constância das variâncias e correlações. Estes modelos permitem a modelagem dos efeitos aleatórios, utilizando funções polinomiais lineares e, também, consideram heterogeneidade de variâncias residuais em cada idade.

Inicialmente os modelos de regressão aleatória consideravam a estrutura de variância residuais homogêneas, que juntamente associada a problemas de ajuste das variâncias de efeito de ambiente permanente, superestimavam as variâncias genéticas aditivas.

O ajuste dos modelos, ao considerar as variâncias residuais heterogêneas, tende a melhorar a partição da variância total, nas variâncias atribuídas aos efeitos aleatórios incluídos no modelo de análise. No entanto, aumenta o número de parâmetros a ser estimado, o que dificulta a convergência na estimação dos componentes de variância. Contudo, modelos com estruturas de variâncias residuais distintas e número reduzido de classes de variâncias, advindas do agrupamento de classes que contém variações semelhantes, ou por meio de funções de variância residuais, utilizando polinômios ortogonais de Legendre ou ordinários tem sido propostos para melhorar estas análises.

A utilização do modelo de regressão aleatória exige a escolha de uma função para a descrição das trajetórias dos efeitos fixos e aleatórios, estas funções podem ser ortogonais, paramétricas ou de covariância. Dentre estas, os polinômios ortogonais de Legendre, tem sido utilizados. Os polinômios de Legendre são regressões ortogonais, portanto possuem a vantagem computacional de reduzir a correlação entre os coeficientes estimados, facilitando a convergência.

Os polinômios ortogonais de Legendre são os mais adequados para serem usados como covariáveis no ajuste desses modelos, pois evitam a superestimação das variâncias genéticas e herdabilidades nos termos da trajetória ou período de lactação. Estes polinômios, não requerem qualquer pressuposição sobre a forma da curva ou trajetória da característica com o tempo,

são flexíveis e tem se mostrado capazes de modelar as mudanças das médias e variâncias, para as características que se repetem. As funções de Legendre estão na tabela a seguir:

Ordem do polinômio	Expressões
1ª ordem	$\phi_0(t) = \sqrt{(1/2)}$
2ª ordem	$\phi_1(t) = \sqrt{(3/2)}t$
3ª ordem	$\phi_2(t) = 1/2\sqrt{(5/2)}(3t^2 - 1)$
4ª ordem	$\phi(t)_3 = 1/2\sqrt{(7/2)}(5t^2 - 3t)$

4 Exemplo ilustrativo

Vamos considerar o exemplo da apostila retirado do livro de Montgomery.

O experimento estuda a variabilidade de lotes e fornecedores na pureza da matéria prima. A análise assume que os fornecedores são um efeito fixo enquanto que lotes são efeitos aleatórios.

Inicialmente vamos ler (importar) os dados para R com o comando `read.table`. A seguir vamos examinar o objeto que contém os dados.

```
> ex <- read.table("exemplo.txt", header=T)
> ex

> dim(ex)
[1] 36  3

> names(ex)
[1] "forn" "lot"  "resp"

> is.factor(ex$forn)
[1] FALSE
> is.factor(ex$lot)
[1] FALSE

> ex$forn <- as.factor(ex$forn)
> ex$lot <- as.factor(ex$lot)

> is.factor(ex$resp)
```

```

[1] FALSE
> is.numeric(ex$resp)
[1] TRUE

> summary(ex)
forn   lot       resp
1:12   1:9   Min.    : -4.0000
2:12   2:9   1st Qu.: -1.0000
3:12   3:9   Median  :  0.0000
4:9    Mean   :  0.3611
3rd Qu.:  2.0000
Max.    :  4.0000

```

Nos comandos acima verificamos que o objeto `ex` possui 36 linhas correspondentes às observações e 3 colunas que correspondem às variáveis `forn` (fornecedor), `lot` (lote) e `resp` (a variável resposta).

A seguir verificamos que `forn` e `lot` não foram lidas como fatores. NÃO podemos seguir a análise desta forma pois o R leria os valores destas variáveis como quantidades numéricas e não como indicadores dos níveis dos fatores. Para corrigir isto usamos o comando `as.factor` para indicar ao R que estas variáveis são fatores.

Finalmente verificamos que a variável resposta é numérica e produzimos um rápido resumo dos dados.

Na sequência deveríamos fazer uma análise exploratória, alguns gráficos descritivos etc, como na análise dos experimentos mostrados anteriormente. A notação para indicar efeitos aninhados no modelo é `/`. Desta forma poderíamos ajustar o modelo da seguinte forma:

```

> ex.av <- aov(resp ~ forn/lot, data=ex)
> summary(ex.av)
Df Sum Sq Mean Sq F value Pr(>F)
forn          2 15.056   7.528  2.8526 0.07736 .
forn:lot       9 69.917   7.769  2.9439 0.01667 *
Residuals    24 63.333   2.639

```

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

Embora os elementos do quadro de análise de variância estejam corretos o teste F para efeito dos fornecedores está ERRADO. A análise acima considerou todos os efeitos como fixos e portanto dividiu os quadrados médios dos efeitos pelo quadrado médio do resíduo. Como lotes é um efeito aleatório deveríamos dividir o quadrado médio do termo lot pelo quadrado médio de forn:lot

Uma forma de indicar a estrutura hierárquica ao R é especificar o modelo de forma que o termo de resíduo seja dividido de maneira adequada. Veja o resultado abaixo.

```
> ex.av1 <- aov(resp ~ forn/lot + Error(forn) , data=ex)
> summary(ex.av1)
```

```
Error: forn
Df Sum Sq Mean Sq
forn 2 15.0556 7.5278
```

```
Error: Within
Df Sum Sq Mean Sq F value Pr(>F)
forn:lot 9 69.917 7.769 2.9439 0.01667 *
Residuals 24 63.333 2.639
```

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

Agora o teste F errado não é mais mostrado, mas o teste correto também não foi feito! Isto não é problema porque podemos extrair os elementos que nos interessam e fazer o teste desejado. Primeiro vamos verificar que o comando anova produz uma lista que tem entre seus elementos os graus de

liberdade Df e os quadrados médios (Mean Sq). A partir destes elementos podemos obter o valor da estatística F e o valor P associado.

```
> ex.anova <- anova(ex.av)
> is.list(ex.anova)
[1] TRUE

> names(ex.anova)
[1] "Df"      "Sum Sq"  "Mean Sq" "F value" "Pr(>F)"

> ex.anova$Df
1  2
2  9 24
> ex06.anova$Mean
1      2
7.527778 7.768519 2.638889
> Fcalc <- ex.anova$Mean[1]/ex.anova$Mean[2]
> Fcalc
1
0.9690107
> pvalor <- 1 - pf(Fcalc, ex.anova$Df[1], ex.anova$Df[2])
> pvalor
1
0.4157831
```

Uma outra possível e elegante solução no R para este problema é utilizar a função lme do pacote nlme.

Inicialmente temos que carregar o pacote nlme com o comando require. A seguir criamos uma variável para indicar o efeito aleatório que neste exemplo chamamos de `ex$fa` utilizando a função getGroups. Feito isto podemos rodar a função lme que faz o ajuste do modelo.

```

> require(nlme)
[1] TRUE
> ex$fa <- getGroups(ex, ~ 1|forn/lot, level=2)
> ex.av <- lme(resp ~ forn, random=~1|fa, data=ex)
> ex.av
Linear mixed-effects model fit by REML
Data: ex06
Log-restricted-likelihood: -71.42198
Fixed: resp ~ forn
(Intercept)      forn2      forn3
-0.4166667      0.7500000      1.5833333

Random effects:
Formula: ~1 | fa
(Intercept) Residual
StdDev:      1.307561 1.624483

Number of Observations: 36
Number of Groups: 12

```

Este modelo tem a variável forn como efeito fixo e a variável lot como efeito aleatório com o componente de variância σ_{lote}^2 . Além disto temos a variância residual σ^2 . A saída acima mostra as estimativas destes componentes da variância como sendo $\hat{\sigma}_{lote}^2 = (1.307)^2 = 1.71$ e $\hat{\sigma}^2 = (1.624)^2 = 2.64$.

O comando anova vai mostrar a análise de variância com apenas os efeitos principais. O fato do programa não incluir o efeito aleatório de lotes na saída não causa problema algum. O comando intervals mostra os intervalos de confiança para os componentes de variância. Portanto para verificar a significância do efeito de lotes basta ver se o intervalo para este componente de variância exclui o valor 0, o que é o caso neste exemplo conforme vemos abaixo.

```

> anova(ex.av)
numDF denDF  F-value p-value
(Intercept)    1    24 0.6043242 0.4445
forn           2     9 0.9690643 0.4158
> intervals(ex.av)
Approximate 95% confidence intervals

```

```

Fixed effects :
lower      est.      upper
(Intercept) -2.0772277 -0.4166667 1.243894
forn2      -1.8239746  0.7500000 3.323975
forn3      -0.9906412  1.5833333 4.157308

```

```

Random Effects :
Level: fa
lower      est.      upper
sd((Intercept)) 0.6397003 1.307561 2.672682

```

```

Within-group standard error :
lower      est.      upper
1.224202  1.624483  2.155644

```

Finalmente uma versão mais detalhada dos resultados pode ser obtida com o comando `summary`.

```

> summary(ex.av)
Linear mixed-effects model fit by REML
Data: ex06
AIC      BIC      logLik
152.8440 160.3265 -71.42198

```

```

Random effects :
Formula: ~1 | fa
(Intercept) Residual
StdDev:      1.307561 1.624483

```

```

Fixed effects: resp ~ forn
Value Std.Error DF   t-value p-value
(Intercept) -0.4166667 0.8045749 24 -0.5178718 0.6093
forn2        0.7500000 1.1378407  9  0.6591432 0.5263
forn3        1.5833333 1.1378407  9  1.3915246 0.1975
Correlation:
(Intr) forn2
forn2 -0.707
forn3 -0.707  0.500

```

```

Standardized Within-Group Residuals :

```

Min	Q1	Med	Q3	Max
-1.4751376	-0.7500844	0.0812409	0.7060895	1.8720268

Number of Observations: 36

Number of Groups: 12

5 Conclusão

No uso de modelos fixos, devem-se estimar os próprios efeitos fixos, enquanto que os modelos aleatórios prestam-se para estimar os componentes de variância (das variáveis aleatórias), bem como a predição das próprias variáveis aleatórias. Dessa forma, os modelos mistos podem servir para a estimação das médias de um modo mais preciso, uma vez que deve-se levar em conta a influência dos componentes de variância que podem ser estimados pelo modelo.

Referências

- [1] Nota metodológica sobre Modelos Lineares Mistos. Professor Jomar Antonio Camarinha Filho. UFPR - 2003.
- [2] Rovadoscki, G. A. - Modelos de curvas de crescimento e regressão aleatória em linhagem nacionais de frango caipira. Dissertação de mestrado(2013) - ESALQ/USP.
- [3] Lucambio, F. - Modelos Lineares Mistos. UFPR (2008).
- [4] Macedo, O. J. - Modelos de regressão aleatória usando como bases as funções polinomiais de Legendre, de Jacobi modificadas e trigonométrica, com uma aplicação na análise genética dos pesos dos bovinos da raça Nelore. Tese -ESALQ/USP(2007).

- [5] Camarinha filho, J. A. - Modelos lineares mistos: estruturas de matrizes de variância e covariâncias e seleção de modelos. Tese - ESALQ/USP(2002).