

2

A questão da restrição na aleatorização e a utilização de Modelos mistos

Neste capítulo é apresentada inicialmente a questão da restrição na aleatorização em planejamento de experimentos, e em seguida é descrito o procedimento experimental *split-plot*. Além disso, é abordada a utilização de modelos mistos como uma solução para a estimação dos parâmetros quando ocorre restrição na aleatorização durante as rodadas do experimento. Para isso, é detalhada a modelagem mista, incluindo a estimação dos parâmetros de efeitos fixos, efeitos aleatórios e a estimação dos componentes da variância.

2.1

Restrição na aleatorização

Em um experimento planejado, normalmente assume-se que as rodadas são realizadas de forma completamente aleatorizadas. Isso significa que as corridas são executadas de forma aleatória, e os níveis dos fatores são reinicializados em cada uma das rodadas do experimento. Quando pelo menos uma dessas características não ocorre, dizemos que há restrição na aleatorização.

A questão da restrição na aleatorização é natural nos experimentos que envolvem blocos. Nessas situações, os blocos surgem como componentes necessários do planejamento experimental. Porém, o uso de blocagem padrão não é a única fonte de restrição na aleatorização. Em algumas vezes, um experimento que é definido, pela sua eficiência, como completamente aleatorizado, deveria utilizar um formato de ‘restrição à aleatorização’, devido a considerações práticas. Tais considerações dizem respeito a fatores que possuem níveis difíceis de mudar,

ou difíceis de controlar, e com isso não são reinicializados nas sucessivas rodadas do experimento (Myers & Montgomery, 2002).

Segundo Webb *et al.* (2004), é muito comum, na indústria, os níveis dos fatores não serem reinicializados de uma rodada para a outra, em um experimento. As principais causas de não reinicializar os níveis dos fatores são tempo e custo. Segundo registros dos autores, pelo menos um fator não é reinicializado na maioria dos experimentos. E dessa forma, algumas vezes são utilizadas técnicas como blocagem ou *split-plot* para tratar dos fatores que não foram reinicializados.

Quanto à estimação dos parâmetros, o fato dos níveis de um ou mais fatores serem os mesmos em sucessivas rodadas pode acarretar a quebra da premissa de independência, o que faz com que o método dos mínimos quadrados ordinários produza estimadores viesados, gerando testes e inferências incorretos (Webb *et al.*, 2004). Além disso, a não reinicialização dos níveis dos fatores – em sucessivas rodadas – gera perda de precisão na estimativa dos parâmetros e variância maior do que a esperada.

Ju e Lucas (2002) estudaram experimentos com estrutura fatorial onde nem todos os fatores são reinicializados nas sucessivas rodadas do experimento. Como mencionado anteriormente, a não reinicialização acarreta uma restrição na aleatorização. Com isso, um experimento completamente aleatorizado não é atingido mesmo quando o experimento é conduzido usando-se uma ordem aleatória nas rodadas. Segundo Ju e Lucas (2002), a restrição na aleatorização mais comum ocorre quando o experimento é executado em ordem aleatória, mas nem todos os fatores são reinicializados em cada rodada. Isso acontece principalmente devido à presença de fatores cujos níveis são difíceis de mudar. Em linhas gerais, podemos considerar dois tipos de fatores em relação à facilidade de mudar os níveis. São eles: fatores fáceis de mudar e fatores difíceis de mudar.

O fator fácil de mudar é definido como sendo o fator cujos níveis são estabelecidos de forma independente em cada rodada do experimento, ou como o fator que não necessita ser reinicializado. Já o fator difícil de mudar é definido como o fator cujos níveis não são reinicializados em sucessivas rodadas que possuem o mesmo nível para o fator. Normalmente, o fator difícil de mudar leva mais tempo e consome mais recursos, sendo mais caro para ser reinicializado do que o fator fácil de mudar.

Segundo Myers & Montgomery (2002), um experimento completamente aleatorizado é, geralmente, muito difícil de ser conduzido. Torna-se praticamente inviável quando existem fatores cujos níveis são difíceis de mudar ou difíceis de controlar. Um exemplo é o fator temperatura. Em experimentos fatoriais de dois níveis, é conveniente e mais barato executar o experimento em que sejam realizadas, primeiramente, as rodadas em baixa temperatura, e posteriormente em alta temperatura. Isso eliminaria ter que por a temperatura de volta e esquentar o forno de acordo com a ordem do esquema atribuído na aleatorização. Em muitas aplicações existem fatores que são difíceis de mudar.

2.1.1

Estruturas Split-Plot

Existem ocasiões em que a dificuldade em mudar um ou mais fatores requer o uso da estrutura *split-plot* (Myers & Montgomery, 2002). Tal estrutura envolve fatores cujos níveis são atribuídos aleatoriamente aos *whole plots* (*ou main plots*), e outros fatores com seus níveis atribuídos aleatoriamente aos *subplots*. Os *subplots* ficam aninhados no interior dos *whole plots*, tal que um *whole plot* consiste em um cluster de *subplots*, e um nível de um desses fatores do *whole plot* é aplicado ao cluster inteiro. Os fatores difíceis de mudar têm seus níveis atribuídos aleatoriamente aos *whole plots*. Os fatores que não são difíceis de mudar têm seus níveis atribuídos aleatoriamente às unidades experimentais do *subplot*. Como resultado, cada *whole plot* recebe somente um nível do fator difícil de mudar, mas todos os níveis dos fatores que não são difíceis de controlar. Como tal estrutura fatorial envolve restrição na aleatorização, não se trata de um experimento completamente aleatorizado.

Em uma estrutura *split-plot* com replicação é necessário, inicialmente, testar os efeitos do *whole plot* e do *subplot*. Para isso, as duas médias quadráticas dos erros são usadas separadamente. Se a replicação for uma opção prática, então os testes de significância das variáveis podem ser realizados pelo uso de dois termos de erro (erro do *whole plot* e do *subplot*). E com isso, a estimativa dos coeficientes da regressão é afetada pela existência de dois componentes da

variância (σ^2 e σ_δ^2), onde σ^2 é a variância do erro do *sub-plot* e σ_δ^2 é variância do erro do *whole plot*. Para ilustrar os problemas associados com a estimativa do modelo sob a estrutura *split-plot*, considere o modelo da equação na forma matricial, a seguir:

$$\mathbf{y} \sim N(\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\beta}), \mathbf{V}) \quad (2.1)$$

onde $\mathbf{V} = \text{var}(\mathbf{y})$ é a matriz de variância-covariância da resposta \mathbf{y} . Vale ressaltar que quaisquer duas observações no mesmo *whole plot* não são independentes. Elas têm covariância σ_δ^2 porque compartilham o mesmo componente de erro. E quaisquer duas observações em diferentes *whole plots* são independentes. A forma da matriz \mathbf{V} é dada por:

$$\mathbf{V} = \begin{bmatrix} \mathbf{T}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_2 & \cdots & \mathbf{0} \\ \mathbf{0} & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{T}_a \end{bmatrix} \quad (2.2)$$

onde $\mathbf{T}_i = \sigma^2 \mathbf{I}_{b \times b} + \sigma_\delta^2 \mathbf{1}_{b \times 1} \mathbf{1}'_{1 \times b}$, onde a é o número de *whole plots*, \mathbf{b} representa os níveis do *sub-plot*, e:

σ^2 = variância do erro *sub-plot*

σ_δ^2 = variância do erro *whole plot*.

Então, a estimativa dos parâmetros é dada por:

$$\mathbf{b} = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y} \quad (2.3)$$

onde \mathbf{X} é a matriz das variáveis regressoras e \mathbf{V} é matriz de variância-covariância na Equação 2.1. Os estimadores na Equação 2.3 são estimadores de mínimos quadrados generalizados que são apropriados quando $\mathbf{V} \neq \sigma^2 \mathbf{I}$. Infelizmente, a Equação 2.3 não pode ser usada diretamente, pois o analista não tem conhecimento de σ^2 e σ_δ^2 e por isso não pode calcular \mathbf{V} . Ou seja, enquanto que a estrutura *split-plot* resulta em um conveniente meio de experimentação, os aspectos da análise da regressão se tornam complicados devido à presença de observações correlacionadas. Sendo assim, o princípio da estimativa por mínimos quadrados generalizados, descrito na Equação 2.3, sugere o uso da abordagem de modelos mistos para estimar σ^2 e σ_δ^2 .

2.2

Modelos lineares mistos (LMM)

Os modelos mistos são utilizados para descrever dados de experimentos cuja estrutura de tratamentos envolve fatores fixos e aleatórios. Ou seja, são modelos que permitem que se trabalhe com ambos os efeitos fixos e aleatórios em um modelo linear que possui uma variável resposta contínua.

Um modelo linear padrão é construído para tratar de efeitos fixos, no qual os níveis do fator representam todos os níveis possíveis para aquele fator, ou todos os níveis em que inferências serão feitas. Por outro lado, os efeitos são considerados aleatórios se os níveis dos fatores em um estudo são aleatoriamente selecionados de uma população de possíveis níveis daquele fator. Nesse caso, a população de possíveis níveis de um efeito aleatório tem uma distribuição de probabilidade, com determinada média e variância. Ao modelar ambos os fatores, fixos e aleatórios, os modelos mistos fornecem a flexibilidade de não modelar somente a média, como em um modelo linear padrão, mas também as variâncias e covariâncias.

Além disso, os modelos lineares mistos trabalham com dados onde as observações não são independentes, como ocorre quando há medidas repetidas. Ou seja, o LMM corrige modelos com erros correlacionados.

Existem duas vertentes distintas no desenvolvimento de modelos mistos. A primeira ocorreu em planejamento de experimentos, onde a introdução da estrutura *split-plot* conduziu a modelos com alguns componentes de erro. Nesse caso, o maior interesse está na inferência sobre médias, ou seja, nos efeitos dos tratamentos (Lee *et al.*, 2006). A segunda vertente surgiu em modelos de componentes da variância, onde os dados são desbalanceados, e o interesse não está centrado nas variâncias dos efeitos aleatórios, porém na estimação dos próprios efeitos aleatórios.

2.2.1

Representação matricial dos modelos mistos

Seja \mathbf{y} um vetor de observações referente à variável resposta, de dimensão $N \times 1$, e seja \mathbf{X} matriz dos regressores de efeitos fixos, conhecida, de dimensão $N \times p$, e \mathbf{Z} matriz de regressores de efeitos aleatórios, conhecida, de dimensão $N \times q$. Seja ainda $\boldsymbol{\beta}$ o vetor de efeitos fixos, desconhecido, de dimensão $p \times 1$ e \mathbf{v} vetor de efeitos aleatórios, desconhecido, de dimensão $q \times 1$.

O modelo linear misto padrão pode ser representado por:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e} \quad (2.4)$$

onde $\mathbf{e} \sim N(0, \mathbf{R})$, $\mathbf{v} \sim N(0, \mathbf{D})$, e \mathbf{v} e \mathbf{e} são independentes. Sendo que \mathbf{e} é vetor de erros aleatórios não observáveis, de dimensão $N \times 1$.

Podemos escrever também:

$$\mathbf{E}[\mathbf{y} | \mathbf{v}] = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v},$$

significando que para o \mathbf{v} realizado, a equação acima representa a média condicional (McCulloch & Searle, 2001).

Para especificar $\text{var}(\mathbf{y})$, usamos $\text{var}(\mathbf{v}) = \mathbf{D}$ e definimos

$$\text{var}(\mathbf{y} | \mathbf{v}) = \mathbf{R}$$

Com isso, temos:

$$\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R}),$$

mostrando que os efeitos fixos entram apenas na média enquanto que a matriz de regressores de efeitos aleatórios e matriz de covariância dos efeitos aleatórios entram na variância de \mathbf{y} . Ou seja, em modelos lineares mistos, os efeitos fixos são usados para modelar a média de \mathbf{y} , enquanto que os efeitos aleatórios governam a estrutura de variância-covariância de \mathbf{y} . De fato, a principal razão de se usar efeitos aleatórios é simplificar a dificuldade de especificação de $N(N+1)/2$ elementos distintos da $\text{var}(\mathbf{y}_{NX1})$. Sem usar efeitos aleatórios, teríamos que lidar com os elementos de $\text{var}(\mathbf{y})$ com uma variedade de formas. Porém, com o uso de efeitos aleatórios podemos lidar com variâncias e covariâncias atribuíveis a fatores reconhecidos em estarem afetando os dados. Uma vez que os dois tipos de efeitos (fixos e aleatórios) são diferentes e são tratados de forma diferente quando os dados são analisados, devemos decidir, para cada fator, quando ele deve ser considerado fixo ou aleatório.

Nos modelos mistos, três aspectos são fundamentais: estimação e testes de hipóteses dos efeitos fixos, predição dos efeitos aleatórios e estimação dos componentes de variância.

2.2.2

Estimação dos efeitos fixos

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

o logaritmo da verossimilhança é

$$\ell = -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})' \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) - \frac{1}{2} \log |\mathbf{V}| - \frac{N}{2} \log 2\pi. \quad (2.5)$$

Tomando a derivada do log da verossimilhança em relação a $\boldsymbol{\theta}$, temos:

$$\frac{\partial \ell}{\partial \theta} = \frac{\partial \boldsymbol{\mu}'}{\partial \theta} \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}),$$

com $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ e $\boldsymbol{\theta} = \boldsymbol{\beta}$. Substituindo na derivada e igualando a zero, com $\boldsymbol{\beta}$ representado por $\boldsymbol{\beta}^0$ tem-se

$$\mathbf{X}' \mathbf{V}^{-1} \mathbf{X} \boldsymbol{\beta}^0 = \mathbf{X}' \mathbf{V}^{-1} \mathbf{y}$$

tal que

$$\boldsymbol{\beta}^0 = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y} \quad (2.6).$$

Como $\boldsymbol{\beta}^0$ varia com $(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-}$, limitamos a atenção em $\mathbf{X}\boldsymbol{\beta}^0$ que é invariante, uma vez que $(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-} \mathbf{X}' \mathbf{V}^{-1}$ o é. Dessa forma,

$$\text{ML}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y} \quad (2.7)$$

é o estimador de máxima verossimilhança de $\mathbf{X}\boldsymbol{\beta}$, e assim, $\boldsymbol{\lambda}' \mathbf{X}\boldsymbol{\beta}^0$ é o estimador de máxima verossimilhança de $\boldsymbol{\lambda}' \mathbf{X}\boldsymbol{\beta}$ para qualquer $\boldsymbol{\lambda}$.

Com $\text{var}(\mathbf{y}) = \mathbf{V}$ pode-se ver que

$$\text{var}(\mathbf{X}\boldsymbol{\beta}^0) = \mathbf{X}(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-} \mathbf{X}' \mathbf{V}^{-1} \mathbf{X}(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-} \mathbf{X}'.$$

Uma vez que $(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-}$ é uma inversa generalizada de $(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})$, e também por causa da propriedade da invariância mencionada anteriormente,

$$\text{var}(\mathbf{X}\boldsymbol{\beta}^0) = \mathbf{X}(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-} \mathbf{X}'.$$

Para testar a hipótese nula $H_0 : \mathbf{S}' \mathbf{X} \boldsymbol{\beta} = \mathbf{m}$, onde \mathbf{S}' possui *rank* completo ($r_s \leq r_X$), a estatística de qui-quadrado pode ser obtida usando-se

$$X^2 = (\mathbf{S}' \mathbf{X} \boldsymbol{\beta}^0 - \mathbf{m})' [\mathbf{S}' \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{S}]^{-1} (\mathbf{S}' \mathbf{X} \boldsymbol{\beta}^0 - \mathbf{m}).$$

Sob H_0 , X^2 tem distribuição χ^2 com $r_s = \text{rank}(S)$ graus de liberdade.

Mais tipicamente, \mathbf{V} é conhecida por um múltiplo escalar. Para simplificar a notação, escrevemos \mathbf{V} em termos de uma matriz de pesos \mathbf{W} , que é a inversa de \mathbf{V} de acordo com o múltiplo escalar, ou seja, $\mathbf{V} = \sigma^2 \mathbf{W}^{-1}$, onde \mathbf{W} é conhecido. Nesse caso, a estatística a seguir pode ser obtida como um teste de razão de verossimilhança:

$$F = \frac{(\mathbf{S}' \mathbf{X} \boldsymbol{\beta}^0 - \mathbf{m})' [\mathbf{S}' \mathbf{X} (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}' \mathbf{S}]^{-1} (\mathbf{S}' \mathbf{X} \boldsymbol{\beta}^0 - \mathbf{m})}{r_s \hat{\sigma}^2}$$

onde

$$\hat{\sigma}^2 = \frac{\mathbf{y}' [\mathbf{W} - \mathbf{W} \mathbf{X} (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W}] \mathbf{y}}{N - r_X}$$

Sob a hipótese nula, F tem uma distribuição F com r_s e $N - r_X$ graus de liberdade.

Quando \mathbf{V} for desconhecido, a função log de verossimilhança ℓ de 2.5 tem que ser maximizada em relação aos elementos de $\boldsymbol{\mu}$ e \mathbf{V} . Para $\boldsymbol{\mu} = \mathbf{X} \boldsymbol{\beta}$ e $\boldsymbol{\theta} = \boldsymbol{\beta}$, fixando $\partial \ell / \partial \theta$ igual a zero, chegaremos ao mesmo resultado encontrado para $\boldsymbol{\beta}^0$ em (2.6), somente quando \mathbf{V} for estimado por $\hat{\mathbf{V}}$ a partir da maximização ℓ em relação aos parâmetros em \mathbf{V} . Dessa forma, o estimador de $\mathbf{X} \boldsymbol{\beta}$ será

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

Uma dificuldade é que (2.7) requer $\mathbf{V} = \text{var}(\mathbf{y})$, ou seja, envolvem os componentes de variâncias. Em muitas situações práticas os componentes de variância não são conhecidos. Nesses casos, uma estratégia interessante e conveniente consiste em obter estimativas dos componentes de variância, que serão utilizadas em lugar dos componentes em \mathbf{V} . Então, substitui-se \mathbf{V} por $\hat{\mathbf{V}}$ e assim tem-se:

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

Ou seja, na análise do modelo linear misto tem-se, em geral, interesse na estimativa e testes de hipóteses dos efeitos fixos. Entretanto, para a estimativa de

uma função estimável dos parâmetros de efeitos fixos, no modelo misto, é necessário o conhecimento das estimativas dos componentes de variância. Assim as estimativas dos parâmetros de efeitos fixos, no modelo misto, dependem diretamente dos métodos utilizados na obtenção das estimativas dos componentes de variância (McCulloch & Searle, 2001).

2.2.3

Predição dos efeitos aleatórios

As suposições acerca dos efeitos aleatórios diferem das suposições dos efeitos fixos, e com isso o tratamento dos dois tipos de efeitos não é o mesmo. Um efeito fixo é considerado uma constante, que queremos estimar. Mas um efeito aleatório é considerado como apenas um efeito vindo de uma população de efeitos. Para enfatizar essa diferença, usamos o termo ‘predição de efeitos aleatórios’ ao invés de ‘estimação’ (McCulloch & Searle, 2001).

Usando a suposição que os efeitos aleatórios realizados, que determinam os dados, são apenas uma seleção aleatória da população de efeitos aleatórios, a melhor predição (a que minimiza a média do quadrado dos erros de predição) para um efeito aleatório é a média condicional $E[\mathbf{v} | \mathbf{y}]$.

Voltando ao caso geral onde,

$$\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta}, \mathbf{V}), \text{ para } \mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R},$$

e assumindo que \mathbf{y} e \mathbf{v} seguem uma distribuição Normal conjunta, temos a esperança condicional $E[\mathbf{v} | \mathbf{y}]$,

$$E[\mathbf{v} | \mathbf{y}] = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Substituindo $\boldsymbol{\beta}$ por $\boldsymbol{\beta}^0 = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$, temos o melhor preditor linear não viesado (BLUP) de \mathbf{v} sob normalidade. Escrevemos então $\tilde{\mathbf{v}} = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0)$

$$\tilde{\mathbf{v}}^0 = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0) = \mathbf{D}\mathbf{Z}'\mathbf{P}\mathbf{y}.$$

E temos $\text{var}(\tilde{\mathbf{v}}) = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{D}$ e $\text{var}(\tilde{\mathbf{v}}^0) = \mathbf{D}\mathbf{Z}'\mathbf{P}\mathbf{Z}\mathbf{D}$.

Quando \mathbf{D} e \mathbf{V} são conhecidos, o cálculo de $\tilde{\mathbf{v}}$ não apresenta dificuldades e pode ser encontrado utilizando-se as equações do modelo misto.

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \sigma_e^2 \mathbf{D}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{pmatrix}$$

Resolvendo o sistema de equações, obtém-se as soluções para os efeitos fixos $\boldsymbol{\beta}^0$ e predições para os efeitos aleatórios $\tilde{\mathbf{v}}$. Para tal desenvolvimento assume-se que \mathbf{V} é conhecida. Quando isso não ocorre, \mathbf{D} e \mathbf{V} são estimadas por $\hat{\mathbf{D}}$ e $\hat{\mathbf{V}}$, utilizando-se um dos métodos disponíveis, como o método da máxima verossimilhança (ML), ou o método da máxima verossimilhança restrita (REML). E as soluções podem ser escritas como:

$$\begin{pmatrix} \boldsymbol{\beta}^0 \\ \tilde{\mathbf{v}} \end{pmatrix} = \begin{pmatrix} (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y} \\ \mathbf{D}\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0) \end{pmatrix} \quad (2.8)$$

Assim, utilizando-se as expressões apresentadas em (2.8) obtém-se as estimativas dos efeitos fixos e as predições dos efeitos aleatórios, $\boldsymbol{\beta}^0$ e $\tilde{\mathbf{v}}$, respectivamente. No caso de $\tilde{\mathbf{v}}$, o resultado é o chamado melhor preditor estimado não viesado (BLUP), representado por:

$$\hat{\mathbf{v}} = \hat{\mathbf{D}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \hat{\mathbf{D}}\mathbf{Z}'\hat{\mathbf{P}}\mathbf{y}.$$

Um aspecto interessante das equações de modelo misto é que elas podem ser utilizadas em procedimentos iterativos para os cálculos das estimativas de ML, e REML, dos componentes de variância.

2.2.3.1

Quando devemos usar efeitos aleatórios?

O modelo $y_{ij} = \mu + \nu_i + e_{ij}$, $i = 1, \dots, q$, $j = 1, \dots, n_i$ será o mesmo se assumirmos os ν_i sendo parâmetros fixos ou aleatórios. Então, quando devemos assumir os efeitos como aleatórios?

Uma regra comum é que os efeitos são assumidos como fixos se o interesse está na inferência de valores específicos dos efeitos. Porém, acredita-se que essa é uma visão errada, pois implica que não é importante estimar os efeitos aleatórios. De fato, existem muitas aplicações onde existem interesses nos efeitos aleatórios. Existem, também, aplicações onde acreditamos que a resposta dependa de alguns fatores, mas nem todos são conhecidos ou mensuráveis. Essas variáveis

desconhecidas são normalmente modeladas como efeitos aleatórios. Quando medidas repetidas são obtidas de uma observação, o efeito aleatório é uma variável comum não observável para cada observação e é responsável pela criação de uma dependência entre as medições repetidas. Esses efeitos aleatórios devem ser considerados como uma amostra de alguma distribuição populacional convenientemente definida (Lee & Nelder, 2006).

2.2.4

Estimação dos componentes da variância

Na análise de modelos lineares mistos, a estimação dos componentes de variância é de fundamental importância, pois tanto a predição dos efeitos aleatórios quanto à estimação dos efeitos fixos depende dessa estimação. Ou seja, para a obtenção do estimador (BLUE) de $\mathbf{X}\beta$ e do BLUP de \mathbf{v} , é necessária a estimativa dos componentes de variância.

O componente da variância é estimado para o fator, somente quando ele é considerado efeito aleatório. Isto reflete uma das diferenças entre efeitos fixos e aleatórios. Para os efeitos aleatórios nós queremos não apenas prevê-los, como também estimar suas variâncias. A variação nos níveis de um fator aleatório é assumida como sendo representativa da variação de toda a população de possíveis níveis. Dessa forma, variações nos níveis dos fatores aleatórios podem ser usadas para estimar a variação na população.

A variação nos níveis dos fatores fixos é considerada arbitrariamente determinada pelo experimentador, ou seja, o experimentador pode fazer os níveis de um fator fixo variarem muito ou pouco como desejado. Sendo assim, a variação do fator fixo não pode ser usada para estimar a variância da população. Essas duas distinções básicas entre efeitos fixos e efeitos aleatórios são importantes para se conhecer melhor as propriedades dos componentes da variância.

Diversos métodos têm sido propostos para estimar os componentes de variância, destacando-se o método da máxima verossimilhança (*Maximum Likelihood*: ML) devido a Hartley & Rao (1967); o método da estimação quadrática não-viesada de variância mínima (*Minimum Variance Quadratic*

Unbiased Estimation: MIVQUE) descrito em Rao (1971), e o método da máxima verossimilhança restrita (*Restricted Maximum Likelihood*: REML) descrito por Patterson & Thompson (1971).

A idéia básica por trás da estimação por REML e ML é encontrar um conjunto de pesos para os efeitos aleatórios no modelo, que maximize o valor negativo do logaritmo natural da verossimilhança dos dados (a verossimilhança dos dados pode variar de 0 a 1, então minimizando o valor negativo do logaritmo natural vezes a verossimilhança dos dados significa maximizar a probabilidade, ou verossimilhança, dos dados).

A estimação dos componentes da variância pelos vários métodos mencionados é similar. Componentes da variância REML e ML são estimados otimizando-se iterativamente a estimativa dos parâmetros para os efeitos no modelo. Os métodos de estimação dos componentes da variância REML e ML estão relacionados ao MIVQUE. De fato, nos programas, REML e ML utilizam as estimativas do MIVQUE como valores iniciais para uma solução iterativa para os componentes da variância (Iemma & Perri, 1999).

O método REML difere do ML pelo fato da verossimilhança dos dados ser maximizada somente para os efeitos aleatórios. Sendo assim, REML é chamado de solução restrita. Em ambos, REML e ML, uma solução iterativa é encontrada para os pesos dos efeitos aleatórios, no modelo, que maximiza a verossimilhança dos dados. A seguir, serão tratados, com um maior detalhamento, os métodos ML e REML.

2.2.5

Método da máxima verossimilhança: ML

Hartley & Rao (1967) aplicaram o método da máxima verossimilhança ao modelo misto geral. Tal método consiste em maximizar a função de verossimilhança em relação aos efeitos fixos e aos componentes de variância. Assim, para o modelo misto (2.4), assumindo $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, a função de verossimilhança é:

$$L = (2\pi)^{-\frac{n}{2}} |\mathbf{V}|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\mathbf{y} - \mathbf{X}\beta)' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta) \right]$$

, onde $|\mathbf{V}|$ é o determinante da matriz \mathbf{V} .

O logaritmo da função de verossimilhança é dado por:

$$\ell = \log L = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\beta)' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta) \quad (2.9)$$

Assim, - 2 log da função de verossimilhança (2.9) é

$$-2 \log L = \log |\mathbf{V}| + (\mathbf{y} - \mathbf{X}\beta)' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta) + n \log(2\pi) \quad (2.10)$$

Minimizando essa expressão com respeito a β , tem-se:

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

com $\hat{\beta}$ e $\hat{\mathbf{V}}$ representando as estimativas ML de β e \mathbf{V} , respectivamente.

Substituindo $\hat{\beta}$ na expressão (2.10), tem-se

$$-2 \log L = n \log(2\pi) + \log |\hat{\mathbf{V}}| + (\mathbf{y} - \mathbf{X}\hat{\beta})' \hat{\mathbf{V}}^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta}) \quad (2.11)$$

Minimizando essa função em relação à todos os parâmetros desconhecidos, obtém-se um sistema de equações cuja solução fornece as estimativas ML.

Essas equações são não lineares e resolvidas numericamente, em geral, por processos iterativos como o algoritmo de Newton-Raphson. O processo é repetido até que o critério de convergência adotado seja satisfeito. Assim, o método da máxima verossimilhança supõe normalidade dos dados, é iterativo e fornece estimativas não-negativas de componentes de variância, mas estas são viesadas, pois o método não considera a perda de graus de liberdade resultante da estimação dos efeitos fixos do modelo (Iemma & Perri, 1999).

2.2.6

Método da máxima verossimilhança restrita: REML

Patterson & Thompson (1971) propuseram uma modificação do método da máxima verossimilhança para modelos mistos.

Os estimadores REML são obtidos maximizando-se a parte da função de verossimilhança que é invariante ao parâmetro de locação, isto é, em termos do modelo misto $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\upsilon} + \boldsymbol{\epsilon}$, invariante a $\mathbf{X}\boldsymbol{\beta}$. Em outras palavras, os estimadores REML maximizam a função de verossimilhança de um vetor de combinações lineares das observações que são invariantes a $\mathbf{X}\boldsymbol{\beta}$. Seja $\mathbf{K}'\mathbf{y}$ esse vetor. Então $\mathbf{K}'\mathbf{y} = \mathbf{K}'\mathbf{X}\boldsymbol{\beta} + \mathbf{K}'\mathbf{Z}\boldsymbol{\upsilon} + \mathbf{K}'\boldsymbol{\epsilon}$ é invariante a $\mathbf{X}\boldsymbol{\beta}$ se e somente se $\mathbf{K}'\mathbf{X} = 0$.

Com $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, tem-se que para $\mathbf{K}'\mathbf{X} = 0$, $\mathbf{K}'\mathbf{y} \sim N(0, \mathbf{K}'\mathbf{V}\mathbf{K})$. As equações REML também podem ser deduzidas das equações ML substituindo-se: \mathbf{y} por $\mathbf{K}'\mathbf{y}$, \mathbf{X} por $\mathbf{K}'\mathbf{X} = 0$, \mathbf{Z} por $\mathbf{K}'\mathbf{Z}$, \mathbf{V} por $\mathbf{K}'\mathbf{V}\mathbf{K}$.

O método REML é implementado construindo-se a função $-2 \log L_R$. Assim, para a estimação REML, -2 log da função de verossimilhança restrita é

$$-2 \log L_R = \log|\mathbf{V}| + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (n - k) \log(2\pi) \quad (2.12)$$

onde k é o *rank* da matriz \mathbf{X} e $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}$, com $\hat{\boldsymbol{\beta}}$ e $\hat{\mathbf{V}}$ representando as estimativas REML de $\boldsymbol{\beta}$ e \mathbf{V} , respectivamente.

O método REML tem sido considerado o preferido para estimar componentes de variância de dados desbalanceados. As razões para essa preferência são justificadas pelas propriedades desses estimadores.

O método REML supõe normalidade dos dados, é iterativo e fornece sempre estimativas não negativas dos componentes de variância, assim como o método ML. No entanto, considera a perda de graus de liberdade devido aos efeitos fixos, fornecendo estimadores não viesados e de variância mínima para dados balanceados.

A principal diferença entre os métodos ML e REML é que o ML usa a função de verossimilhança de \mathbf{y} ou o logaritmo desta função, enquanto o REML adota a função de verossimilhança de $\mathbf{K}'\mathbf{y}$, um vetor de combinações lineares das observações (com esperança nula) que representa efetivamente as observações ajustadas para os efeitos fixos (Iemma & Perri, 1999).

Vale ressaltar que para ambos os métodos de estimação utilizados, a estimação dos componentes de variância irá depender da estrutura de covariâncias adotada.

2.2.7

Estruturas gerais de covariâncias

Conforme abordado anteriormente, a análise de modelos mistos envolve a análise da parte fixa e a análise da parte aleatória. Se existirem fatores aleatórios ou medidas repetidas, o pesquisador deve especificar o tipo de estrutura de covariância. O tipo de estrutura de covariância especificado é usado como um ponto de partida nos algoritmos REML e ML, para a estimação dos parâmetros. Em outras palavras, a estimação dos componentes de variância depende da estrutura da matriz **D** e do método de estimação utilizado. Especificando o tipo de estrutura de covariância, o pesquisador estará informando ao algoritmo a forma da matriz de covariância existente entre os termos aleatórios. Por exemplo, uma matriz de covariância onde apenas a diagonal principal é diferente de zero significa que as observações são independentes.

Várias estruturas de covariâncias podem ser especificadas para a matriz **D**, no software S-Plus, como por exemplo: identidade, diagonal, simetria geral, simetria composta, dentre outras. A seguir são detalhadas algumas dessas estruturas.

Alguns tipos de estrutura de covariância (da matriz **D**):

- Identidade. Também chamada de ‘componentes da variância’, ‘estrutura simples’ ou ‘modelo independente’, porque a variância dos resíduos é independente das variâncias do efeito. Dessa forma, estrutura de componentes da variância assume que a correlação de qualquer par de medidas repetidas é igual não importando o quão distante estão. Esse tipo de estrutura associa uma matriz identidade com cada efeito aleatório.

$$\sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- Diagonal. Esta estrutura de covariância apresenta variâncias diferentes (o que não ocorre com a estrutura ‘componentes da variância’) e correlação zero entre os elementos. Com isso, o pesquisador está assumindo que diferentes categorias da variável de efeito aleatório têm diferentes variâncias na variável dependente.

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

Para escolher o melhor tipo de estrutura de covariância, podem ser utilizadas as estatísticas de qualidade do ajuste (AIC e BIC, por exemplo).

O software S-Plus além de oferecer várias opções para a estrutura de covariâncias da matriz **D**, também permite uma especificação geral da matriz de covariâncias dos erros. Isso é feito em duas etapas:

- (1) especificando a função de variância, onde o objetivo é modelar a heterocedasticidade dos termos de erro dentro do grupo;
- (2) determinando a estrutura de correlação, no qual o objetivo é modelar a correlação dentro do grupo, não capturada pelos efeitos aleatórios. Dessa forma, tal modelagem permite que os componentes de erro sejam correlacionados.

As funções disponíveis para os tipos de função de variância dentro do grupo no software S-Plus são, dentre outras:

- Identidade: variâncias diferentes por níveis de um fator;
- *Fixed*: pesos fixos determinados por uma covariante.
- Exponencial: exponencial de uma covariante;
- *Combination*: combinação de funções de variância.

Quanto às estruturas de correlação disponíveis, temos, dentre outras:

- AR(1): estrutura auto-regressiva de primeira ordem com variância homogênea;

- Simetria composta: corresponde a uma correlação constante;
- Simetria geral: matriz de correlação geral, sem estrutura adicional;
- Exponencial: correlação exponencial espacial;
- Linear: correlação linear espacial.

Assim como na estrutura de covariância, para selecionar a melhor estrutura de correlação, podem ser utilizadas as estatísticas de qualidade do ajuste (AIC e BIC).

Neste capítulo foi abordada a questão da restrição na aleatorização, devido à presença de fatores difíceis de mudar, e os problemas que isso acarreta, como, por exemplo, a dificuldade de estimação pelos métodos tradicionais (mínimos quadrados ordinários). Foi então introduzida a análise de modelos lineares mistos como uma solução a essa questão da modelagem na presença de dados provenientes de experimentos que não foram realizados de forma completamente aleatorizada. No próximo capítulo será apresentado o estudo de caso utilizado neste trabalho, com o objetivo de modelar a média e a variância da variável resposta utilizando modelos mistos.