

Recensões

Brady T. West, Kathleen B. Welch,
Andrzej T. Galecki
*Linear Mixed Models: A Practical Guide
Using Statistical Software*
Chapman & Hall, 353 págs., Preço: 75,58€
em 2007



O presente livro faz uma abordagem à temática dos modelos lineares mistos (MLM), sendo objectivo dos seus autores a modelação estatística de bases de dados, representativas de casos tipo, fundamentada num adequado substrato de natureza teórica e aplicada. O livro pressupõe algum conhecimento prévio das temáticas relacionadas com álgebra matricial, análise numérica ou modelos lineares de efeitos fixos e aleatórios, abordadas em obras como "Matrix Algebra Useful for Statistics", and "Linear Models", de S.R. Searle; "Variance Components" de G. Casella, e C.E. McCulloch; ou "Linear Mixed Models for Longitudinal Data" de G. Verbeke e G. Molenberghs.

A diferenciação deste texto, relativamente a essas obras de natureza mais fundamental, baseia-se assim no ajustamento de MLM a cinco bases de dados de complexidade crescente,

utilizando cinco pacotes informáticos - SAS, R, SPSS, Stata e HLM. Os programas informáticos utilizados são apresentados e discutidos com um nível de pedagogia e profundidade suficientes para habilitar o leitor com uma sensibilidade fundamental à análise dos seus próprios problemas.

O livro compreende sete capítulos. No Capítulo 1 é feita uma definição de MLM e uma resenha histórica dos desenvolvimentos teóricos e de programação informática no século passado.

No Capítulo 2 é apresentado um resumo alargado relativo aos modelos mistos lineares em que são considerados os tipos de modelos e dados fundamentais: agrupados a dois níveis, agrupados a três níveis, medidas repetidas, longitudinais e agrupados longitudinais. Esta tipologia é claramente apresentada de modo a que possa ser relacionada com outras formas alternativas de delineamento experimental baseadas p.ex. em blocos aleatórios. As análises são realizadas em conjuntos de dados balanceados ou não balanceados. Na Secção 2.1 deste capítulo são definidos e clarificados os conceitos fundamentais de efeitos fixos, aleatórios e encaixados. As Secções 2.2 a 2.4 abordam a especificação matricial dos MLM, relativa a um único sujeito compreendendo as componentes fixa e aleatória do modelo, integrando as matrizes de observação e de covariáveis, bem como os vectores das respostas contínuas, dos efeitos fixos, dos efeitos aleatórios e dos resíduos. Na Secção 2.2 são também indicadas estruturas possíveis para as matrizes D e R_i , p. ex. não estruturada, de componentes de variância, estrutura assumida por defeito, de simetria composta ou auto-regressiva, de modo a ser possível a acomodação de várias formas de

correlação entre os resíduos ou as variáveis aleatórias. Na Secção 2.2 são ainda apresentadas formulações alternativas relativas ou à totalidade dos sujeitos em observação, baseadas na sobreposição dos modelos matriciais dos sujeitos individuais em matrizes diagonais por blocos, ou hierárquica, baseada na construção de variáveis específicas a níveis de dados.

Na Secção 2.3 são abordados os conceitos de modelos marginais, ou não condicionais, sem componente aleatória, relativos às médias de populações de níveis de efeitos fixos, não esquecendo os chamados modelos marginais implicados por um MLM, úteis ao estudo de correlação de sujeitos num determinado agrupamento (cluster).

Nas sub-secções 2.4.1 e 2.4.2 são desenvolvidos conceitos fundamentais relativos aos métodos de estimação mais relevantes em MLM a saber, os métodos de máxima verosimilhança (ML) e de máxima verosimilhança restrita (REML).

Na Secção 2.5 são sintetizadas as principais técnicas numéricas para optimização das funções de verosimilhança, de que são exemplos os algoritmos EM (Expectation-Maximization) e de Newton-Raphson.

Nas Secções 2.6 e 2.7 são especificados detalhes relativos aos testes de hipóteses e estatísticas para selecção e construção de modelos, por exclusão ou acréscimo de variáveis associadas a efeitos fixos ou aleatórios a vários níveis de análise. São abordadas as diversas técnicas de construção de modelos segundo as estratégias de *top-down* ou *step-up*, correspondentes a situações em que, respectivamente, a construção se inicie com um número máximo ou mínimo de efeitos fixos.

A Secção 2.8 é dedicada às técnicas de

diagnóstico dos modelos para verificar se os pressupostos de distribuição dos resíduos condicionais, escalados p. ex. por studentização, se verificam ou se o ajustamento do modelo é sensível a observações fora do usual.

Na Secção 2.9 são abordados tópicos relacionados com temas como o do cálculo dos coeficientes de correlação intra-classe, destinados a avaliar a homogeneidade das respostas da variável dependente integrada num dado agrupamento, o da problemática da modelação quando existam dados em falta ou o cálculo dos BLUPS.

Os Capítulos 3 a 7 são destinados à aplicação dos cinco programas informáticos a situações, respectivamente, de dados agrupados a dois níveis, dados agrupados a três níveis, dados de medidas repetidas, dados longitudinais e dados longitudinais agrupados. Os dados são relativos à medicina e ciências sociais. Estes capítulos são elaborados numa tentativa de abordagem integrada de: i) estratégias de *top-down* ou *step-up* para a construção dos modelos ii) de análise de diversas estruturas das matrizes D de variância-covariância dos efeitos aleatórios e R_i dos resíduos, homogéneas ou heterogéneas, segundo um grau de complexidade crescente, iii) de interpretação das estimativas dos parâmetros, iv) de estabelecimento de modelos de resposta condicionais, específicos dos agrupamentos de dados e não condicionais, relativos aos conjuntos das populações e v) de diagnósticos para os modelos finais.

Em cada capítulo é feito um estudo preliminar dos dados de modo a captar fontes de variabilidade que serão consideradas nos diversos modelos. São também apresentadas, para cada problema, tabelas com os conjuntos dos

modelos e respectivos parâmetros fixos ou aleatórios a estudar em cada caso, com os testes de hipóteses realizados e com os resultados da estimação para os diversos programas informáticos.

O Capítulo 3 aborda um problema de análise de peso de crias de ratos em ninhadas, sujeitas a três doses de um composto experimental (alta, baixa e de controle). As variáveis do problema situam-se a dois níveis hierárquicos relativas ao indivíduo e à ninhada, sendo que cada ninhada foi aleatoriamente atribuída a um dado nível de tratamento e as crias de rato foram encaixadas nas diversas ninhadas.

O Capítulo 4 trata um problema sociológico relativo ao aproveitamento em Matemática de um conjunto de alunos em função da escola, origem social dos estudantes, sexo ou a formação dos professores. As variáveis são especificadas a quatro níveis hierárquicos a saber: escola, turma e estudante, sendo que o modelo é elaborado segundo uma estratégia *step-up*, contrariamente a todos os outros exemplos, que seguem o modelo *top-bottom*.

O Capítulo 5 aborda a análise de medidas repetidas, no qual múltiplas medidas são realizadas no mesmo sujeito, sob condições distintas ou ao longo do tempo. Os dados analisados situam-se a dois níveis, o primeiro representando as medidas efectuadas nos sujeitos e o segundo representando os sujeitos. As covariáveis medidas ao nível 2 dos dados representam a variação entre sujeitos e as variáveis medidas ao nível 1 descrevem a variação nos sujeitos.

O Capítulo 6 aborda um problema relativo a modelos longitudinais utilizando um conjunto de dados

relativos a um estudo de capacidades de sociabilidade em crianças com problemas de desenvolvimento e autismo.

Finalmente, o Capítulo 7 propõe o estudo da classe de modelos para dados longitudinais agrupados, segundo os quais as unidades de análise são encaixadas em agrupamentos e as várias medidas da variável dependente são realizadas ao longo do tempo nas várias unidades de análise. Cada agrupamento pode integrar um número distinto de unidades de análise e os instantes temporais em que as medidas se realizam podem ser igualmente distintos. O exemplo apresentado inclui três níveis de análise, sendo que o nível 1 representa o conjunto de medidas longitudinais repetidas efectuadas ao longo do tempo, o nível 2 as unidades de análise e o nível 3 os agrupamentos de unidades.

Como comentário final, considerando a dificuldade de num livro não muito longo expor todas as matérias associadas à complexa problemática dos modelos mistos lineares, julgamos que a obra se pode considerar um bom manual sobre aplicações de modelos mistos lineares, contributivo para futuro desenvolvimento de trabalho nesta ou noutras áreas relacionadas.

Abel Martins Rodrigues
Investigador Auxiliar
INIA/INRB, IP