

Análise de Dados Longitudinais

M. Salomé Cabral
M. Helena Gonçalves



SOCIEDADE PORTUGUESA
DE ESTATÍSTICA

28 Setembro - 1 Outubro
Nazaré 2011



XIX Congresso Anual da Sociedade Portuguesa de Estatística

Nazaré, 28 de Setembro a 1 de Outubro de 2011

Análise de Dados Longitudinais

M. Salomé Cabral M. Helena Gonçalves

Edições SPE

Ficha Técnica:

Título: Análise de Dados Longitudinais

Autores: M. Salomé Cabral e M. Helena Gonçalves

Editora: Sociedade Portuguesa de Estatística

Capa: Instituto Nacional de Estatística

Impressão: Instituto Nacional de Estatística

Tiragem: 500 exemplares

ISBN: 978-972-8890-24-7

Depósito Legal: 332365/11

Prefácio

Os dados longitudinais surgem sempre que observações repetidas da variável resposta são obtidas ao longo do tempo para cada indivíduo, num ou mais grupos em estudo. As investigações envolvendo este tipo de dados são referidas como estudos longitudinais e desempenham um papel fundamental nas mais variadas áreas de conhecimento na medida em que é possível caracterizar as alterações dentro do indivíduo e relacionar essas alterações com um conjunto de factores que não o tempo.

Os estudos longitudinais constituem, assim, uma importante estratégia na investigação, como é o caso da investigação clínica em que fornecem conhecimentos quer sobre o desenvolvimento e a persistência da doença quer sobre os factores que podem alterar o curso do desenvolvimento dessa doença.

Existe uma grande variedade de desafios na análise de dados longitudinais. Por um lado, devido à sua natureza, as medições repetidas provenientes de estudos longitudinais são multivariadas e têm uma estrutura complexa de autocorrelação cuja modelação desempenha um papel fundamental na análise deste tipo de dados. Por outro, a natureza da variável resposta pode ser contínua ou discreta. Além disso, os estudos longitudinais permitem a introdução de covariáveis que variam ao longo do tempo, o que torna mais complexa a sua

análise. Finalmente, em estudos longitudinais a existência de dados omissos é frequente; a modelação correcta quando existem dados omissos continua a ser um dos maiores desafios da análise de dados longitudinais.

A importância deste tema é não só prática mas também teórica, tendo-se assistido nos últimos 20 anos a um grande desenvolvimento das metodologias estatísticas para analisar dados longitudinais.

O objectivo deste texto é o de apresentar os diferentes métodos de análise usados nos diversos tipos de dados numa linguagem mais descritiva do que orientada para a investigação, de forma a constituir não só um texto de apoio para o mini-curso bem como uma base de trabalho para todos aqueles que queiram utilizar estas metodologias.

Os diferentes métodos de análise descritos requerem o uso de programas estatísticos apropriados. Neste livro é usado o programa R que permite não só a utilização dos diferentes modelos apresentados assim como o acesso aos dados dos exemplos analisados.

Ao longo do livro são feitas diversas referências com o objectivo de fornecer ao leitor uma base bibliográfica de trabalho nesta área.

Não queremos terminar sem agradecer ao Presidente da SPE, Professor Doutor Carlos Braumann e à Comissão Organizadora de XIX Congresso Anual da SPE o convite que nos fizeram para realizar o mini-curso em Análise de Dados Longitudinais assim como a possibilidade de publicar este livro. Por último, a primeira autora agradece ao Francisco Saldanha da Gama pelo encorajamento e pelos “`LATEX` technical advices” que muito contribuíram para a realização deste trabalho.

Lisboa, Maio de 2011

M. Salomé Cabral e M. Helena Gonçalves

Índice

1	Conceitos Básicos e Exemplos	1
1.1	Introdução	1
1.2	Objectivos e Características	3
1.3	Exemplos	4
1.4	Perspectiva Histórica	8
1.4.1	Dados Longitudinais Gaussianos	8
1.4.2	Dados Longitudinais não-Gaussianos	12
1.5	Terminologia e Notação	14
1.6	Modelos para a Análise de Dados Longitudinais	16
1.7	Dados Omissos em Estudos Longitudinais	19
1.8	Organização do Livro	22
2	Modelo Linear Misto	23
2.1	Introdução	23
2.2	O Modelo	24
2.3	Estimação dos Parâmetros no Modelo Marginal	28
2.3.1	Método da Máxima Verosimilhança	28

iv Índice

2.3.2	Método da Máxima Verosimilhança Restrita	31
2.3.3	Máxima Verosimilhança <i>versus</i> Máxima Verosimilhança Restrita	34
2.4	Predição dos Efeitos Aleatórios	34
2.4.1	Melhor Preditor Linear Centrado (BLUP)	35
2.4.2	Estimação de Bayes Empírica	36
2.4.3	Shrinkage	37
2.4.4	Equações do Modelo Misto	38
2.5	Algoritmos de Optimização	40
2.6	Inferência Estatística para o Modelo Marginal	43
2.6.1	Distribuições Assintóticas	43
2.6.2	Inferência para os Efeitos Fixos	45
2.6.3	Inferência para as Componentes da Variância	48
2.7	Inferência Estatística para os Efeitos Aleatórios	52
2.8	Análise de Resíduos	53
2.8.1	Verificação das Condições sobre os Erros Aleatórios	54
2.8.2	Verificação das Condições sobre os Efeitos Aleatórios	55
3	Modelação da Matriz de Variância - Covariância dos Erros Aleatórios	57
3.1	Introdução	57
3.2	O Modelo Linear Misto Geral	58
3.3	Decomposição da Matriz de Variância-Covariância	60
3.4	Modelação da Heterocedasticidade	60

3.4.1	Seleccção do Modelo	62
3.5	Modelação da Dependência	63
3.5.1	Estruturas de Correlação Serial	63
3.5.2	Estruturas de Correlação Espacial	68
3.5.3	Seleccção do Modelo	72
3.6	Efeitos Aleatórios <i>versus</i> Estrutura de Correlação dos Erros	73
3.7	Modelo Linear Geral para Dados Longitudinais Gaussianos	74
3.7.1	Modelo Linear Misto <i>versus</i> Modelo Linear Geral	76
4	Equações de Estimação Generalizadas	79
4.1	Introdução	79
4.2	Estrutura do Modelo Marginal	81
4.3	Equações de Estimação Generalizadas	86
4.4	Propriedades dos Estimadores GEE	89
4.5	Inferência Estatística com base nos Estimadores GEE	91
4.6	GEE e Valores Omissos	92
4.7	Análise de Resíduos	93
4.8	Distribuição Conjunta no Modelo Marginal. Máxima Verossimilhança	93
5	Modelo Linear Generalizado Misto	95
5.1	Introdução	95
5.2	Estrutura do Modelo Linear Generalizado Misto	96
5.3	Interpretação dos Coeficientes do Modelo	98

vi Índice

5.4	Estimação	100
5.4.1	Método de Aproximação dos Dados	101
5.4.2	Método de Aproximação da Função a Integrar	102
5.4.3	Método de Aproximação do Integral	103
5.5	Inferência	107
5.6	Análise de Resíduos	108
5.7	Modelo Marginal <i>versus</i> Modelo com Efeitos Aleatórios	108
6	Exemplos de Aplicação	113
6.1	Introdução	113
6.2	BodyWeight	114
6.2.1	Modelo Linear Misto	114
6.2.2	Modelo Linear Geral para Dados Longitudinais	129
6.3	Seizure	130
6.3.1	Modelo Marginal. Método GEE	132
6.3.2	Modelo Linear Generalizado Misto	136
6.4	Locust	139
6.4.1	Modelo Marginal	140
6.4.2	Modelo Linear Generalizado Misto	144
	Bibliografia	146

Capítulo 1

Conceitos Básicos e Exemplos

1.1 Introdução

Quando numa experiência os indivíduos são medidos repetidas vezes ao longo do tempo em relação a uma mesma característica, num ou mais grupos de tratamento, sendo o próprio tempo um factor de interesse diz-se que se têm dados longitudinais ¹ (Molenberghs & Verbeke, 2005).

Os dados longitudinais podem ser obtidos de uma forma prospectiva ou retrospectiva; na primeira situação os indivíduos são seguidos ao longo do tempo, na segunda situação múltiplas medições em cada indivíduo são extraídas do seu historial (Diggle et al., 2002).

Estudos Longitudinais *versus* Estudos Transversais

Com a existência de medições repetidas ao longo do tempo sobre o mesmo indivíduo os estudos longitudinais permitem, ao contrário dos estudos transversais, separar o que no contexto de um estudo populacional se chama o efeito da coorte do efeito da idade (Diggle et al., 2002). Por exemplo, sabe-se que a percentagem de gordura no corpo das raparigas aumenta após a menarca. Se os investigadores

¹Os sociólogos e os economistas designam os estudos longitudinais por estudos painel.

2 Conceitos Básicos e Exemplos

estão interessados em determinar esse aumento da percentagem de gordura após a menarca e fazem um estudo transversal, eles podem medir essa percentagem em dois grupos: um grupo com 10 anos de idade (uma coorte pré-menarca) e um grupo com 15 anos de idade (uma coorte pós-menarca). Neste estudo transversal a comparação da percentagem de gordura entre os dois grupos de raparigas pode ser feita usando um teste t-Student (não emparelhado). Esta comparação não fornece uma estimativa correcta da alteração da gordura que ocorre no corpo das raparigas desde os 10 aos 15 anos. O efeito do crescimento ou da idade, que é um efeito inerente ao próprio indivíduo, não pode ser estimado num estudo transversal, pois com este estudo não são feitas medições de como os indivíduos mudam ao longo do tempo. Num estudo transversal o efeito da idade está confundido com o efeito da coorte. Dito de outra maneira, existem muitas características que diferenciam as raparigas destes dois grupos que podem distorcer a relação entre idade e gordura. Por outro lado, um estudo longitudinal que mede uma única coorte de raparigas aos 10 anos e aos 15 anos fornece uma estimativa correcta da alteração da percentagem de gordura com a idade. Neste estudo longitudinal a comparação é feita com um teste t-Student emparelhado usando a diferença entre a percentagem de gordura de cada rapariga como variável resposta. Esta comparação dentro de cada indivíduo fornece uma estimativa correcta da alteração da percentagem de gordura dos 10 para os 15 anos. Além disso, cada rapariga funciona como o seu próprio controlo, sendo as alterações na percentagem de gordura durante a duração do estudo estimadas independentemente de qualquer variação de gordura entre indivíduos.

Dados Agrupados e Correlacionados

Um dos aspectos que caracteriza os dados longitudinais é o deles se-

rem agrupados. Nos dados longitudinais os grupos são constituídos pelas medições repetidas sobre o mesmo indivíduo em diferentes ocasiões. Por outro lado, nos estudos longitudinais existe uma ordem implícita nos tempos das observações repetidas em cada indivíduo que leva a uma correlação “serial” entre as observações. As medições repetidas de cada vector resposta tendem a estar correlacionadas e, como tal, a estrutura de autocorrelação desempenha um papel importante na estimação dos parâmetros do modelo ajustado aos dados (Diggle et al., 2002).

1.2 Objectivos e Características

O objectivo principal de um estudo longitudinal é o de caracterizar as alterações da variável resposta com o tempo. Ainda que as medições das alterações dentro do indivíduo sejam um objectivo fundamental de um estudo longitudinal é igualmente objecto de interesse determinar se essas alterações se relacionam com um conjunto de covariáveis, isto é, com o conjunto de factores previamente escolhidos, que não o tempo.

Vão-se seguidamente indicar algumas das características dos estudos longitudinais. A primeira, já referida anteriormente, é o facto, de para cada indivíduo, se ter um vector resposta constituído pelas medições repetidas que estão provavelmente correlacionadas e, como tal, a estrutura de autocorrelação desempenha um papel importante na estimação dos parâmetros do modelo a ajustar aos dados. A segunda é o de existir uma elevada variabilidade entre os diferentes indivíduos. A terceira é a dos indivíduos terem diferente número de observações e de estas terem sido feitas em ocasiões distintas. Este tipo de situação é a regra e não a excepção. A quarta é o de, em alguns estudos longitudinais, poder acontecer que não só a variável resposta se al-

4 Conceitos Básicos e Exemplos

tera ao longo do tempo mas também o valor das covariáveis. Por exemplo, no estudo da relação entre o nível do plasma e o estado de saúde, quer o nível do plasma (a covariável) quer o estado de saúde (a variável resposta) alteram-se ao longo do tempo. O objectivo é estimar a relação dinâmica entre estas duas variáveis com o tempo. Note-se que é uma relação que ocorre dentro dos indivíduos, e que pode variar de indivíduo para indivíduo.

Todas estas características fazem com que a análise estatística deste tipo de dados seja complexa e em muitos casos seja ainda uma área de investigação.

Uma das vantagens destes estudos é o facto da existência de todos os dados disponíveis para cada indivíduo, isto é, das medições repetidas sobre o mesmo indivíduo, poder ser usada na análise o que leva a: (i) captar as alterações dentro de cada indivíduo; (ii) aumentar a potência estatística, separando os erros nas medições dos erros aleatórios; (iii) reduzir o enviesamento.

1.3 Exemplos

Nesta secção são considerados três exemplos de estudos longitudinais retirados das áreas da biologia e da medicina. A sua apresentação tem como objectivo ilustrar as circunstâncias em que estes estudos são apropriados e servirão igualmente para mostrar como as diferentes metodologias de análise são utilizadas. Esta última parte será o objecto do Capítulo 6 onde a análise detalhada dos mesmos será apresentada. Todos estes exemplos encontram-se disponíveis em bibliotecas do R (R Development Core Team, 2011), indicando-se entre parêntesis o nome pelo qual são designados nas respectivas bibliotecas.

Exemplo 1.1 (BodyWeight) - Os dados deste exemplo fazem parte da biblioteca `nlme` (Pinheiro et al., 2011) do R e encontram-se descritos em Crowder & Hand (1990) e em Hand & Crowder (1996). Estes dados dizem respeito ao peso de ratos medidos ao longo de 64 dias. O peso dos ratos (em gramas) foi medido de sete em sete dias a partir do dia 1 até ao dia 64, com uma medição extra no dia 44. A experiência começou várias semanas antes do “dia 1” (início do período de pesagem). Existem 3 grupos de ratos cada grupo correspondendo a uma dieta. O objectivo desta experiência é o de se saber se as taxas de crescimento do peso dos ratos em cada uma das 3 dietas diferem ao longo do tempo.

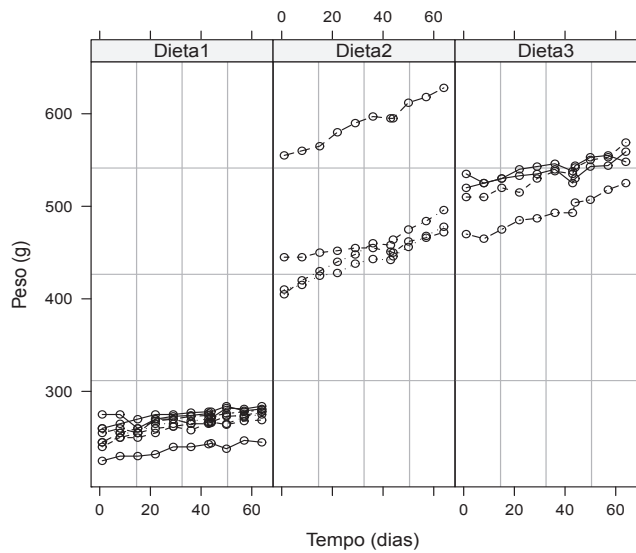


Figura 1.1 Gráfico dos perfis individuais, por dieta, do Exemplo 1.1.

Exemplo 1.2 (locust)- Os dados deste exemplo referem-se ao resultado da análise do efeito da fome no comportamento locomotor

6 Conceitos Básicos e Exemplos

de 24 gafanhotos (*Locusta migratoria*) observado em 161 instantes de tempo. Os indivíduos foram divididos em dois grupos de tratamento (“alimentado” e “não alimentado”) e, dentro de cada grupo, os indivíduos são alternadamente “macho” e “fêmea”. Nesta análise as categorias da variável resposta foram “move-se” e “não se move”. Durante o período de observação, o comportamento de cada indivíduo foi registado a cada trinta segundos. O objectivo desta experiência é o de se saber se a probabilidade de se mover é a mesma nos dois grupos. Este exemplo foi, primeiramente, descrito e analisado por MacDonald & Raubenheimer (1995) e posteriormente por Azzalini & Chiogna (1997) e por Gonçalves & Azzalini (2008) encontrando-se na biblioteca `bird` (Gonçalves et al., 2010) do R.

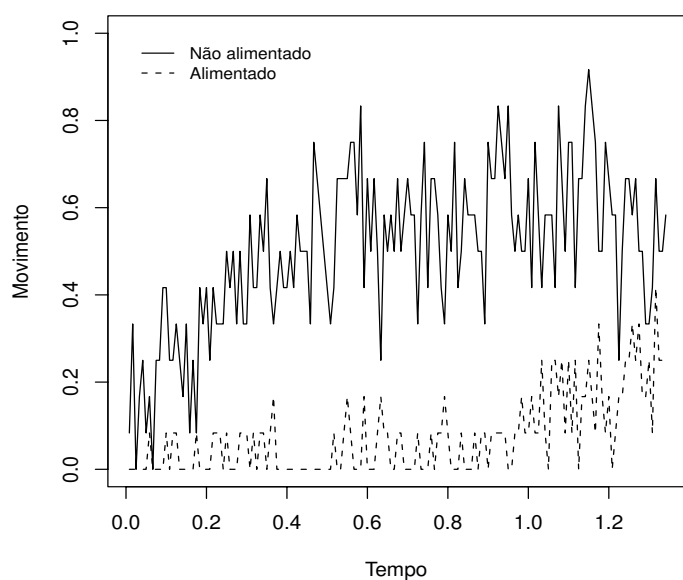


Figura 1.2 Gráfico dos perfis médios, por grupo de tratamento, do Exemplo 1.2.

Exemplo 1.3 (seizure) - Este exemplo encontra-se na biblioteca `geepack` (Højsgaard et al., 2011) do R e refere-se a um ensaio clínico com 59 doentes epiléticos analisado por Thall & Vail (1990), Breslow & Clayton (1993), Diggle et al. (2002), Fitzmaurice et al. (1994) e Gonçalves (2002). Para cada doente foi registado o número de episódios de epilepsia durante um período de 8 semanas (*baseline period*). Foram depois atribuídos aleatoriamente a um de dois grupos de tratamento: o grupo 1 ao qual foi administrado o fármaco anti-epiléptico *prograbide* e o grupo 2 ao qual foi administrado um *placebo* juntamente com a terapia usual. O número de episódios de epilepsia foi registado em 4 intervalos consecutivos de duas semanas (*follow-up period*). O objectivo do ensaio clínico é o de se saber se o *prograbide* reduz a taxa de episódios de epilepsia.

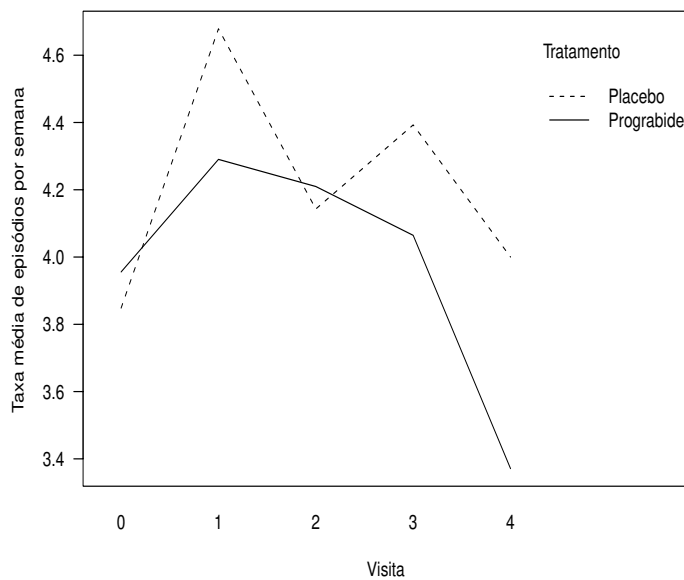


Figura 1.3 Gráfico da taxa média de episódios (por semana) por tratamento para o Exemplo 1.3.

1.4 Perspectiva Histórica

Nesta secção é feito um breve resumo histórico sobre as diferentes metodologias utilizadas na análise de dados longitudinais contínuos (Gaussianos) e discretos (não-Gaussianos) com o objectivo de dar a conhecer o estado da arte.

1.4.1 Dados Longitudinais Gaussianos

Os métodos estatísticos desenvolvidos para a análise de dados longitudinais com respostas contínuas e que podem ser consideradas como seguindo uma distribuição aproximadamente Gaussiana tiveram, tal como os métodos clássicos de regressão linear e o método dos mínimos quadrados, a sua origem na astronomia.

Um dos primeiros métodos propostos para estudar dados longitudinais foi baseado na utilização do modelo de análise de variância (ANOVA) que considera um único efeito aleatório: o indivíduo. Este modelo, designado por modelo misto de análise de variância foi apresentado por Fisher (1918, 1925). É no entanto de referir que quem lançou os fundamentos do modelo linear misto foi o astrónomo britânico George Biddel Airy (Airy, 1981) no seu trabalho sobre um modelo para erros em observações em astronomia. O trabalho apresentado por Fisher é um estudo mais sistemático da sua utilização na ANOVA para a modelação da correlação intraclassas (Fitzmaurice et al., 2008).

O modelo misto de análise de variância tem uma longa história na análise de dados longitudinais e é referido muitas vezes como a “ANOVA univariada” de medições repetidas. Neste modelo assume-se que a correlação existente entre medições repetidas surge porque cada indivíduo tem subjacente (ou latente) um nível de resposta que persiste

ao longo do tempo em todas as medições feitas sobre ele. O efeito individual específico é considerado como uma variável aleatória.

No modelo de análise de medições repetidas *via* ANOVA, a variabilidade de cada indivíduo em torno valor médio da população é explicada através do efeito aleatório. Tem-se então neste modelo duas fontes de variação: a variação entre-indivíduos, e a variação intra-indivíduos o que implica o pressuposto de simetria composta para a estrutura da matriz de variância-covariância das medições repetidas no mesmo indivíduo. Este pressuposto é muitas vezes inadequado em estudos longitudinais na medida em que se espera que a correlação vá diminuindo com o aumento da distância entre os instantes de tempo. Por outro lado este modelo foi concebido para desenhos em que o conjunto de ocasiões em que os indivíduos são medidos é o mesmo para todos os indivíduos, as covariáveis são discretas e todos os indivíduos têm o mesmo número de medições. Assim este modelo não pode ser aplicado a dados longitudinais com espaçamentos irregulares e conseqüentemente em situações em que os indivíduos não têm o mesmo número de observações (dados incompletos) ou que incluem covariáveis quantitativas. Citando Fitzmaurice (Fitzmaurice et al., 2008)

“historicamente a análise de medições repetidas *via* ANOVA foi provavelmente um dos poucos métodos que podia ser aplicado de forma realística ao estudo de dados longitudinais quando os métodos computacionais se encontravam na infância”.

Um outro método para a análise de dados longitudinais com uma longa história mas que já necessita de meios computacionais mais sofisticados consiste na utilização da análise de variância multivariada (MANOVA).

A MANOVA foi desenvolvida para analisar dados em que a cada indivíduo está associado um vector resposta contendo resultados para

10 Conceitos Básicos e Exemplos

mais do que uma variável (pressão sanguínea, nível de glucose no sangue e níveis LDL de colesterol) e não implica restrições na matriz de variância-covariância. No caso dos dados longitudinais está associado a cada indivíduo um vector contendo os resultados de uma única variável resposta medida ao longo do tempo; a situação no entanto é idêntica pelo que a MANOVA passou a ser utilizada na análise destes dados tendo-se tornado numa metodologia apelativa para a análise de dados longitudinais pois não necessita que seja assumida uma estrutura específica para a covariância entre as medições repetidas.

No entanto, a análise de medições repetidas *via* MANOVA tem um conjunto de características que restringe o seu uso na análise de dados longitudinais. Em particular a sua formulação obriga as covariáveis intra-indivíduo a serem iguais para todos os indivíduos. Este facto tem pelo menos duas implicações práticas. A primeira é a de inviabilizar a sua utilização quando o desenho não é equilibrado ao longo do tempo (vectores das medidas repetidas de diferentes comprimentos e/ou obtidos em diferentes sequências de tempo) (Fitzmaurice et al., 2008). A segunda é a de não permitir dados omissos.

Outra abordagem utilizada na análise de dados longitudinais baseia-se em métodos que resumem a sequência de medidas repetidas para o mesmo indivíduo num único valor sumário. A esta abordagem dá-se o nome de análise de medidas sumário e a maior motivação subjacente é o facto de uma sequência de medições repetidas poder ser reduzida a um único número sumário e como tal poder-se aplicar os métodos usuais de análise para respostas univariadas.

Por exemplo, a “área sob a curva” (ou AUC - *area under the curve*) é uma das medidas sumário muito utilizadas com especial destaque em estudos de farmacologia. Uma outra medida sumário vulgarmente utilizada é o declive ou a taxa constante de alteração da resposta ao longo do tempo. A relação entre as medidas sumário resultantes e as

covariáveis é feita usando-se métodos *standard* de análise univariada (paramétricos ou não-paramétricos).

A análise de medidas sumário embora apelativa pela simplicidade, apresenta vários problemas. Um desses problemas é o facto de focar a análise num único aspecto das medições repetidas ao longo do tempo. Ora resumir várias medidas ao longo do tempo num único valor implica necessariamente perda de informação, além de que indivíduos com perfis de resposta manifestamente distintos podem produzir igual valor sumário. Outro problema prende-se com o facto de o método não poder aplicar-se quando as covariáveis variam ao longo do tempo. Por outro lado algumas das medidas sumário propostas não estão bem definidas quando existem dados omissos ou espaçamento irregular entre medições. Mesmo quando estão bem definidas estes métodos perdem eficiência e a análise torna-se mais complicada já que as medidas sumárias resultantes deixam de ter a mesma variância, violando o pressuposto da homocedasticidade, fundamental em muitos dos métodos de análise univariada utilizados na análise de medidas sumário.

Embora possam constituir uma base razoável para a análise de dados longitudinais em casos em que o desenho experimental é simples, os métodos baseados na análise de variância, concebidos para planeamentos equilibrados, exibem constrangimentos que limitam o seu espectro de aplicabilidade (Zuur et al., 2009).

No início dos anos 80, Laird e Ware (Laird & Ware, 1982), a partir de uma classe geral de modelos mistos introduzida por Harville (1977), propuseram uma classe flexível de modelos para dados longitudinais, os modelos lineares mistos ². Estes modelos permitem lidar com os problemas da existência de dados não equilibrados e incompletos,

²Também conhecidos por modelos lineares com efeitos mistos (*linear mixed-effects models*).

12 Conceitos Básicos e Exemplos

medições realizadas em diferentes alturas, covariáveis que variam ao longo do tempo e fornecem, ao mesmo tempo, um modelo para a covariância flexível e, no entanto, parcimonioso. Estes modelos incluem, como casos particulares, a análise univariada de medições repetidas *via* ANOVA e as curvas de crescimento apresentando duas vantagens importantes: a primeira é a da existência de poucas restrições sobre as matrizes de desenho quer dos efeitos fixos quer dos efeitos aleatórios; a segunda é a dos parâmetros do modelo poderem ser estimados através do método da máxima verossimilhança.

Apesar do que atrás foi dito, muitas das anteriores abordagens mais elementares e simplistas mantiveram-se em uso, tornando-se parte dos procedimentos de rotina (Fitzmaurice et al., 2008; Zuur et al., 2009).

1.4.2 Dados Longitudinais não-Gaussianos

Em contraste com os métodos para a análise de dados longitudinais com variável resposta contínua, os métodos para analisar dados longitudinais com variável resposta discreta têm-se desenvolvido apenas nos últimos trinta anos, em parte devido aos avanços na área de métodos computacionais e à existência de computadores com uma velocidade de cálculo cada vez maior.

Neste tipo de dados a metodologia de análise mais comum baseia-se nas equações de estimação generalizadas (GEE - *generalized estimated equations*) proposto por Liang & Zeger (1986). Este método, que se designa por método GEE, tornou-se bastante popular uma vez que combina, de forma simples e flexível, a modelação do valor médio da variável resposta em função das covariáveis com a acomodação da dependência entre as observações sobre o mesmo indivíduo sem necessitar de especificar a distribuição conjunta do vector associado a esse indivíduo.

A par do método GEE foram igualmente realizados importantes trabalhos tendo em consideração a verosimilhança de dados longitudinais discretos, com especial relevo para os dados cuja variável resposta é binária, Fitzmaurice & Laird (1993), Fitzmaurice et al. (1993) e Fitzmaurice et al. (1994). As suas teorias assentam numa “parametrização mista” da família exponencial, a qual expressa a probabilidade de sucesso em sucessivos instantes de tempo através de um modelo logístico que incorpora o efeito da covariável.

Uma alternativa à abordagem anterior é dada por Azzalini (1994a,b), novamente para o caso de dados binários. Nesta metodologia, em cada instante de tempo, a probabilidade de sucesso é dada novamente por um modelo de regressão logístico, mas a estrutura de dependência é tratada de uma forma bastante diferente, onde apenas um parâmetro regula a dependência serial entre observações sucessivas. Relacionado com o trabalho de Azzalini tem-se o trabalho de Heagerty (2002) e de Gonçalves & Azzalini (2008).

Ainda com base na verosimilhança foram desenvolvidos modelos que consideram efeitos aleatórios no preditor linear, os chamados modelos lineares generalizados mistos (Breslow & Clayton, 1993; Wolfinger & O’Connell, 1993). Os modelos lineares generalizados mistos (GLMM’s - *generalized linear mixed models*) são um prolongamento dos modelos lineares generalizados, que têm em conta as componentes adicionais da variabilidade devido a efeitos não observáveis, permitindo a acomodação de respostas de uma distribuição não Gaussiana e a especificação de uma ligação possivelmente não linear entre a resposta e os preditores. Por outro lado a incorporação de efeitos aleatórios permite a modelação da sobredispersão e da correlação. Os GLMM’s são bastante úteis na modelação de dados longitudinais, mas a estimação dos parâmetros é complexa, no sentido em que a verosimilhança poderá envolver integrais de elevada dimensão, cuja

14 Conceitos Básicos e Exemplos

solução analítica não é possível.

Vários métodos têm sido propostos no sentido de ultrapassar esse problema tais como o método de quase-verossimilhança marginal (MQL - *marginal quasi-likelihood*) apresentado por Goldenstein (1991) e o método de quase-verossimilhança penalizada (PQL - *penalized quasi-likelihood*), proposto por Breslow & Clayton (1993). Alguns melhoramentos ao método PQL foram apresentados por Breslow & Lin (1995) e Goldenstein & Rasbash (1996). O método adaptativo da quadratura de Gauss-Hermite (Pinheiro & Bates, 1995, 2000) é a solução indicada quando os efeitos aleatórios seguem uma distribuição Gaussiana e o número de efeitos aleatórios é reduzido. Os métodos de Monte Carlo têm igualmente sido propostos como uma abordagem alternativa à integração numérica.

No contexto em que os dados são contagens, Zeger et al. (1988) propôs um modelo de regressão para séries temporais de contagens análogo ao usado por Liang & Zeger (1986). Mais recentemente, Jowaheer & Sutradhar (2002) usaram as equações de estimação generalizadas para modelar dados de contagens ao longo do tempo com sobredispersão e Gonçalves et al. (2007) apresentaram uma abordagem baseada na máxima verossimilhança em que a dependência é do tipo de Markoviano.

1.5 Terminologia e Notação

Antes de prosseguir vão-se estabelecer alguns termos básicos e a notação a ser usada. As unidades estudadas, cada uma medida em várias ocasiões, serão designadas por indivíduos. As medições são feitas em várias ocasiões ou tempos no mesmo indivíduo.

Designa-se por y_{it} ($t = 1, \dots, T_i$) o valor da resposta na ocasião t para o indivíduo i ($i = 1, \dots, n$), realização da variável aleatória (v.a.) Y_{it}

a que se dá o nome de variável resposta e por $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iT_i})^\top$ o vector das variáveis resposta para o indivíduo i . À sequência das medições efectuadas nesse indivíduo $\mathbf{y}_i = (y_{i1}, \dots, y_{iT_i})^\top$ dá-se o nome de *perfil do indivíduo*.

Associado a cada v.a. Y_{it} , no ponto (i, t) do desenho, está o vector $p \times 1$ das covariáveis $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})^\top$, sendo

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{x}_{i1}^\top \\ \mathbf{x}_{i2}^\top \\ \dots \\ \mathbf{x}_{iT_i}^\top \end{pmatrix}$$

a matriz de ordem $T_i \times p$, também dita matriz $T_i \times p$, associada a \mathbf{Y}_i e o vector $p \times 1$ dos parâmetros desconhecidos, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$. Note-se que \mathbf{x}_{it} pode incluir covariáveis cujos valores não se alteram ao longo do tempo e covariáveis cujos valores se alteram ao longo do tempo. O valor esperado e a variância de \mathbf{Y}_i são representados, respectivamente, por:

$$E(\mathbf{Y}_i) = \boldsymbol{\mu}_i \quad e \quad var(\mathbf{Y}_i) = \mathbf{v}_i.$$

Para cada componente de \mathbf{Y}_i , Y_{it} , tem-se $E(Y_{it}) = \mu_{it}$ e $var(Y_{it}) = v_{it}$. O conjunto de todas as variáveis resposta é dado por um vector $N \times 1$,

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1^\top \\ \mathbf{Y}_2^\top \\ \dots \\ \mathbf{Y}_n^\top \end{pmatrix}$$

onde $N = \sum_{i=1}^n T_i$.

Um último aspecto a considerar é a natureza da variável resposta. No caso em que a variável resposta é contínua, como é o caso do peso

ou da pressão sanguínea, têm-se os dados longitudinais contínuos e os modelos paramétricos utilizados na análise destes dados estão, em geral, baseados no pressuposto da distribuição Gaussiana, pelo que também se designam por dados longitudinais Gaussianos. Quando a variável resposta é discreta, por exemplo ausência ou presença de determinado sintoma ou, número de vezes que esse sintoma ocorre, têm-se os dados longitudinais discretos. O pressuposto da distribuição Gaussiana já não é considerado nos modelos paramétricos utilizados na análise dos dados longitudinais discretos pelo que também se designam por dados longitudinais não-Gaussianos (Molenberghs & Verbeke, 2005).

1.6 Modelos para a Análise de Dados Longitudinais

Diggle et al. (2002) consideram três tipos de modelos para a análise de dados longitudinais: o modelo marginal, o modelo com efeitos aleatórios e o modelo de transição ³.

Modelo Marginal

O modelo marginal tem como objectivo fazer inferência sobre o valor médio populacional. Tem-se assim que, os modelos marginais concentram-se no valor médio marginal e na sua dependência com as covariáveis, não necessitando de assumir pressupostos para a distribuição da variável resposta, apenas de um modelo de regressão. Num modelo marginal o valor esperado marginal, $E(Y_{it})$, é modelado como função das covariáveis. O termo valor esperado marginal é usado para realçar o facto, de que não se está a condicionar a outras variáveis res-

³Neste livro só são abordados os dois primeiros modelos.

posta ou a efeitos aleatórios não observáveis (Heagerty, 1999). Como as medições repetidas, em cada indivíduo, não têm tendência a ser independentes, a análise marginal tem que incluir pressupostos em relação à correlação. O modelo marginal tem a vantagem de o valor esperado da variável resposta e a covariância serem modelados separadamente (Diggle et al., 2002). Os modelos marginais não requerem hipóteses sobre a distribuição subjacente às observações, apenas de um modelo de regressão para o valor médio da variável resposta. Isto é, o modelo marginal fornece um método unificado para analisar diversos tipos de variáveis resposta em dados longitudinais, que não necessita de fazer suposições acerca da distribuição do vector das respostas; este método assenta somente nas hipóteses acerca do valor médio da resposta. Este modelo corresponde ao modelo *Population-average* (PA) definido por Zeger et al. (1988).

Modelo com Efeitos Aleatórios

Os modelos com efeitos aleatórios são mais úteis quando o objectivo é descrever as alterações da resposta média de cada indivíduo e a relação destas alterações com as covariáveis. Assim sendo, estes modelos têm como finalidade realizar inferências sobre o indivíduo e não sobre a população. Estes modelos permitem a acomodação de respostas quer sigam uma distribuição Gaussiana ou não. Para além disso, podem modelar a sobredispersão e a correlação através da incorporação de efeitos aleatórios. Neste modelo, além das covariáveis \mathbf{x}_{it} existe um vector \mathbf{b}_i de efeitos aleatórios que acomoda a variabilidade do indivíduo i em relação ao valor médio da população. Condicional a \mathbf{b}_i , os Y_{iT_i} são independentes. Zeger et al. (1988) designam este modelo por *Subject-specific* (SS).

Modelo Condicionado e Modelo de Transição

Nestes modelos o valor esperado e a dependência temporal são simultaneamente modelados, condicionando uma resposta a outras respostas ou a um subconjunto particular de outras respostas. Um caso particular do modelo condicionado é o chamado modelo de transição ou modelo de Markov. Neste modelo a correlação entre Y_{i1}, \dots, Y_{iT_i} existe porque os valores passados de $Y_{i1}, \dots, Y_{iT_i-1}$ influenciam o valor observado de Y_{iT_i} . As variáveis resposta anteriores a Y_{iT_i} são tratadas como covariáveis adicionais (Diggle et al., 2002). Exemplos específicos de modelos de transição para dados binários e para contagens são apresentados por Korn & Whittemore (1979), Ware et al. (1988) e por Zeger & Qaqish (1988). Os modelos de transição ou modelos de Markov discretos são úteis para modelar o valor esperado da resposta condicionada quer às covariáveis, quer às observações passadas.

Modelo Marginal *versus* Modelo com Efeitos Aleatórios

Na análise de dados longitudinais Gaussianos a utilização de modelos com ou sem efeitos aleatórios não implica qualquer alteração na interpretação dos coeficientes de regressão do modelo. Como se verá no Capítulo 2 o modelo com efeitos aleatórios implica um modelo marginal no caso em que se tem um modelo linear misto. Já quando se têm dados longitudinais não-Gaussianos a relação entre os parâmetros fixos do modelo marginal e do modelo com efeitos aleatórios, em geral, não existe. Este facto levou: (i) à necessidade de distinguir os modelos de acordo com a interpretação dos seus coeficientes de regressão e consequentemente ao uso de termos como “modelo marginal” e “modelo com efeitos aleatórios” (modelo misto ou modelo com efeitos mistos) (ii) a que a escolha entre o modelo marginal e o modelo com efeitos aleatórios para dados longitudinais não-Gaussianos deva

ser feita consoante as questões a que o estudo pretende dar resposta. Neste sentido, qualquer que seja o estudo longitudinal, a diferentes questões científicas correspondem diferentes modelos. Por exemplo, considere-se um estudo longitudinal, em que um médico pretende estudar o benefício de um novo tratamento para um dos seus doentes. Neste caso o interesse está em se saber o comportamento de cada um dos doentes relativamente ao tratamento. Assim, a escolha do modelo com efeitos aleatórios torna-se evidente. Por outro lado, considere-se um estudo longitudinal em que investigadores na área da saúde pública pretendem estudar uma potencial redução na mortalidade na população, sendo esta constituída por doentes que receberam o novo tratamento e os que não receberam. Neste exemplo, a escolha recai, no modelo marginal, dado que o objectivo é realizar inferências sobre a população.

1.7 Dados Omissos em Estudos Longitudinais

Tal como foi referido na Secção 1.2 uma das características dos dados longitudinais é a dos indivíduos terem diferente número de observações e de estas terem sido feitas em ocasiões distintas. Esta característica leva à existência de valores omissos e, como tal, os métodos de análise devem ser capazes de trabalhar com dados não equilibrados sem necessitar de retirar da análise os indivíduos com dados omissos.

Por outro lado, quando existem dados omissos há perda de informação e redução na precisão com a qual as alterações na resposta média ao longo do tempo podem ser estimadas. Esta redução na precisão está directamente relacionada com a quantidade de valores omissos presentes nos dados e pode igualmente influenciar, até certo ponto, a forma como a análise lida com os dados omissos. Por último, quando

20 Conceitos Básicos e Exemplos

há dados omissos a validade de qualquer método de análise requer certas hipóteses sobre as razões das omissões, geralmente referidas como mecanismo de omissão de dados (*missing data mechanism*). Como consequência, quando há dados omissos deve-se ter em consideração as razões que levaram à omissão.

Entende-se por mecanismo de omissão de dados um modelo que descreve a probabilidade com que a resposta é ou não observada em determinada ocasião. Esse modelo de acordo com Rubin (1976) e Little & Rubin (1987, Cap. 6) é classificado em: (i) omissão completamente aleatória (MCAR - *missing completely at random*), (ii) omissão aleatória (MAR - *missing at random*) e (iii) omissão não aleatória (NMAR - *not missing at random*). Neste livro vão-se referir somente os dois primeiros modelos.

Diz-se que se tem MCAR quando a probabilidade da resposta estar omissa não está relacionada com o valor específico da mesma ou com o conjunto de respostas observadas. Por exemplo, durante uma experiência laboratorial envolvendo ratos decidiu-se aleatoriamente, no início da experiência, que devido a problemas orçamentais se omitia a observação referente ao instante de tempo 5 nos ratos com número par e a observação referente ao instante de tempo 7 nos ratos com número ímpar. O mecanismo de omissão verificado nada tem a ver com os valores observados pelo que é MCAR.

O aspecto principal de MCAR é o dos dados observados poderem ser considerados uma amostra aleatória dos dados completos. Como resultado, os momentos (por exemplo, valor médio, variância e covariância) e as distribuições para os dados observados não diferem dos correspondentes momentos e nem das distribuições para os dados completos.

Um mecanismo MCAR tem importantes consequências na análise de dados longitudinais. Em geral, todos os métodos para análise

de dados longitudinais que conduzem a inferências válidas quando não existem dados omissos continuam a produzir inferências válidas quando são aplicados a dados onde há valores omissos de acordo com um mecanismo MCAR. Em particular, todos os métodos que serão apresentados nos diferentes capítulos são válidos quando o mecanismo de omissão pode ser considerado MCAR.

Diz-se que se tem MAR quando a probabilidade das respostas serem omissas depender do conjunto de respostas observadas, mas não se encontra relacionado com o valor omissos que deveria ser obtido. Por exemplo, num ensaio clínico sobre perda de peso, as pessoas que ganham peso têm tendência a abandonar o ensaio. As pessoas que estão no ensaio e que ao irem ao centro de saúde descobrem, ao pesarem-se, que ganharam peso decidem abandonar o ensaio. Este abandono está relacionado com os dados observados (isto é com o valor observado da resposta) e como tal o mecanismo de omissão é MAR.

Quando o mecanismo de omissão é MAR certos métodos de análise de dados longitudinais deixam de produzir estimativas válidas se a distribuição conjunta da resposta \mathbf{Y}_i não for correctamente especificada, ou correctamente modelado o mecanismo de omissão. Por exemplo, a análise baseada no método GEE (*standard* tal como é descrito no Capítulo 4), que precisa de um modelo para o valor médio da variável resposta em função das covariáveis sem necessitar de especificar a distribuição conjunta do vector associado a esse indivíduo está, neste caso, comprometida. Já os métodos baseados na máxima verosimilhança, que especificam correctamente a distribuição de \mathbf{Y}_i , conduzem a estimadores e inferências válidas quando o mecanismo de omissão é MAR.

1.8 Organização do Livro

O livro encontra-se dividido em seis capítulos.

Os Capítulos 2 e 3 são dedicados à análise de dados longitudinais contínuos. No Capítulo 2 é apresentado o modelo linear misto básico e no Capítulo 3 é feita a extensão deste modelo através da modelação da heterocedasticidade e da correlação. O modelo linear geral para dados longitudinais finaliza este capítulo.

No Capítulo 4 é feita a extensão do modelo linear generalizado aos dados longitudinais quer sejam contínuos quer sejam discretos e é apresentado o método baseado nas equações de estimação generalizadas proposto por Liang & Zeger (1986) para a análise de modelos marginais. A análise de dados longitudinais não-Gaussianos usando a máxima verosimilhança é igualmente abordada no contexto do modelo marginal.

No Capítulo 5 aborda-se o modelo linear generalizado misto, proposto por Breslow & Clayton (1993). A introdução de efeitos aleatórios levanta dificuldades ao nível do cálculo da função verosimilhança pelo que neste capítulo são apresentadas algumas das metodologias de cálculo numérico utilizadas. Por último, é feita uma discussão entre o modelo marginal e o modelo com efeitos aleatórios.

No Capítulo 6 é apresentada a análise dos exemplos indicados na Secção 1.3 a fim de ilustrar as diferentes metodologias descritas. O programa estatístico utilizado foi o R e diferentes bibliotecas foram seleccionadas. A escolha das bibliotecas foi feita de acordo com o tipo de dados a analisar, a metodologia a utilizar e os respectivos objectivos para cada um dos exemplos.

Capítulo 2

Modelo Linear Misto

2.1 Introdução

Os modelos lineares mistos para dados longitudinais Gaussianos são uma extensão dos modelos lineares e a sua característica básica é a inclusão de efeitos aleatórios no modelo de regressão. Estes efeitos aleatórios permitem descrever, ao longo do tempo, as alterações dentro de cada indivíduo e flexibilizar a representação da estrutura de variância-covariância induzida pelo agrupamento dos dados.

Os modelos lineares mistos devem o seu nome ao facto de incorporarem quer efeitos fixos, isto é, parâmetros associados a toda a população quer efeitos aleatórios, associados aos indivíduos seleccionados aleatoriamente da população.

Em muitas situações o perfil de cada indivíduo pode ser modelado através de uma relação linear o que confere a estes modelos grande aplicação prática. No entanto o aspecto que torna estes modelos mais apelativos é o facto de poderem ser usados quando os dados não são equilibrados no sentido em que os indivíduos têm diferente número de observações e estas terem sido feitas em ocasiões distintas

Neste capítulo vão-se introduzir os modelos lineares mistos, com base em Harville (1977) e usando a forma paramétrica proposta por Laird & Ware (1982) para o caso em que se tem um único nível de agrupa-

24 Modelo Linear Misto

mento, remetendo-se o leitor para Pinheiro & Bates (2000) para os modelos com dois ou mais níveis de agrupamento. São apresentadas as metodologias utilizadas na estimação e inferência, quer para os efeitos fixos, quer para os efeitos aleatórios, assim com os aspectos relacionados com a validação do modelo.

Em termos de dados omissos, estes modelos assumem os mecanismos de omissão de dados MCAR e MAR.

2.2 O Modelo

O modelo linear misto para um único nível de agrupamento, descrito por Laird & Ware (1982), é dado por:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad (2.1)$$

com $i = 1, \dots, n$, onde $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iT_i})^\top$ é o vector $T_i \times 1$ das variáveis resposta para o indivíduo i , $\boldsymbol{\beta}$ é o vector $p \times 1$ dos efeitos fixos (parâmetros desconhecidos); \mathbf{X}_i é a matriz $T_i \times p$ de covariáveis dos efeitos fixos; \mathbf{b}_i é o vector $q \times 1$ dos efeitos aleatórios; \mathbf{Z}_i é a matriz $T_i \times q$ de covariáveis dos efeitos aleatórios e $\boldsymbol{\epsilon}_i$ é o vector $T_i \times 1$ dos erros aleatórios dentro do grupo. As condições subjacentes ao modelo são:

$$\begin{aligned} \mathbf{b}_i &\sim N(\mathbf{0}, \mathbf{D}) \\ \boldsymbol{\epsilon}_i &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_i) \\ \mathbf{b}_i \text{ e } \boldsymbol{\epsilon}_i &\text{ independentes para diferentes grupos } i \text{ e entre si} \end{aligned} \quad (2.2)$$

com $i = 1, \dots, n$, onde \mathbf{D} é uma matriz $q \times q$, $\boldsymbol{\Sigma}_i$ é uma matriz $T_i \times T_i$, sendo ambas definidas positivas.

Com base no modelo (2.1) e nas condições subjacentes (2.2) conclui-se que a distribuição de \mathbf{Y}_i , condicional ao efeito aleatório \mathbf{b}_i , é Gaussi-

ana multivariada com valor médio $\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$ e matriz de variância-covariância $\boldsymbol{\Sigma}_i$, e escreve-se

$$\mathbf{Y}_i|\mathbf{b}_i \sim N(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \boldsymbol{\Sigma}_i),$$

onde a matriz $\boldsymbol{\Sigma}_i$ representa a variação intra-grupo (intra-indivíduos). A função densidade de probabilidade (f.d.p.) é dada por

$$\begin{aligned} f(\mathbf{y}_i|\mathbf{b}_i) &= (2\pi)^{-T_i/2}|\boldsymbol{\Sigma}_i|^{-1/2} \\ &\times \exp\left(-\frac{(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)}{2}\right). \end{aligned} \quad (2.3)$$

Por outro lado, atendendo à distribuição de \mathbf{b}_i a f.d.p. deste é

$$f(\mathbf{b}_i) = (2\pi)^{-q/2}|\mathbf{D}|^{-1/2} \exp\left(-\frac{\mathbf{b}_i^\top \mathbf{D}^{-1}\mathbf{b}_i}{2}\right). \quad (2.4)$$

Como consequência a f.d.p. marginal de \mathbf{Y}_i é dada por

$$\begin{aligned} f(\mathbf{y}_i) &= \int f(\mathbf{y}_i|\mathbf{b}_i)f(\mathbf{b}_i)d\mathbf{b}_i \\ &= (2\pi)^{-T_i/2}|\mathbf{V}_i|^{-1/2} \exp\left(-\frac{(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})}{2}\right) \end{aligned} \quad (2.5)$$

ou seja, $f(\mathbf{y}_i)$ é a função densidade da v.a. de Gauss T_i -dimensional com valor médio $\mathbf{X}_i\boldsymbol{\beta}$ e matriz de variância-covariância $\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top + \boldsymbol{\Sigma}_i$,

$$\mathbf{Y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i). \quad (2.6)$$

A matriz \mathbf{V}_i , $T_i \times T_i$ e definida-positiva, representa a variação entre-grupos (inter-indivíduos).

Combinando todos os modelos especificados por (2.1) para cada grupo obtém-se o modelo

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \quad (2.7)$$

26 Modelo Linear Misto

onde \mathbf{Y} é um vector $N \times 1$, \mathbf{X} é uma matriz $N \times p$, \mathbf{b} e $\boldsymbol{\epsilon}$, são vectores $qn \times 1$ e $N \times 1$, respectivamente. As matrizes \mathbf{Y} , \mathbf{X} e os vectores \mathbf{b} e $\boldsymbol{\epsilon}$ obtêm-se “empilhando” \mathbf{Y}_i , \mathbf{X}_i , \mathbf{b}_i e $\boldsymbol{\epsilon}_i$, respectivamente.

Para o modelo dado por (2.7) tem-se

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$$

e

$$Var(\mathbf{Y}) = \mathbf{Z}\tilde{\mathbf{D}}\mathbf{Z}^\top + \boldsymbol{\Sigma} = \mathbf{V}$$

onde $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_n)$, $\tilde{\mathbf{D}} = \text{diag}(\mathbf{D}, \dots, \mathbf{D})$, $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_n)$ e $\mathbf{V} = \text{diag}(\mathbf{V}_1, \dots, \mathbf{V}_n)$ são matrizes diagonais de blocos $N \times qn$, $qn \times qn$, $N \times N$ e $N \times N$, respectivamente. A distribuição da v.a. \mathbf{Y} é Gaussiana multivariada dada por

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}).$$

O modelo (2.1) pode então ser definido através das f.d.p $f(\mathbf{y}_i|\mathbf{b}_i)$ e $f(\mathbf{b}_i)$. A esta definição dá-se o nome de **formulação hierárquica do modelo linear misto**. À correspondente distribuição de Gauss marginal com f.d.p. dada por (2.5), dá-se o nome de **formulação marginal do modelo** (Verbeke & Molenberghs, 2000).

Modelo marginal *versus* modelo hierárquico

Embora o modelo marginal decorra do modelo hierárquico, os dois modelos não são equivalentes porque o modelo marginal não define explicitamente a estrutura de efeitos aleatórios. Com efeito, o modelo hierárquico implica o modelo marginal mas o inverso não se verifica. Para ilustrar este facto (Molenberghs & Verbeke, 2005) considere-se que se têm dados equilibrados com duas observações por indivíduo correspondentes aos instantes de observação $x_{i1} = 0$ e $x_{i2} = 1$ e os seguintes modelos:

- (i) **modelo 1** com efeito aleatório na intersecção e erros heterogéneos e independentes.

$$Y_{it} = \beta_0 + b_{0i} + \beta_1 x_{it} + \epsilon_{it}, \quad t = 1, 2$$

A matriz \mathbf{V}_i do modelo marginal é dada por

$$\begin{aligned} \mathbf{V}_i &= \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \boldsymbol{\Sigma}_i \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \mathbf{d} \begin{bmatrix} 1 & 1 \end{bmatrix} + \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \\ &= \begin{bmatrix} d_{11} + \sigma_1^2 & d_{11} \\ d_{11} & d_{11} + \sigma_2^2 \end{bmatrix}; \end{aligned}$$

com $\mathbf{d} = d_{11}$.

- (ii) **modelo 2** com efeitos aleatórios na intersecção e declive, não correlacionados, e $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_2$ (onde \mathbf{I}_k designa uma matriz identidade de ordem k)

$$Y_{it} = \beta_0 + b_{0i} + (\beta_1 + b_{1i})x_{it} + \epsilon_{it}, \quad t = 1, 2$$

A matriz \mathbf{V}_i do modelo marginal é dada por

$$\begin{aligned} \mathbf{V}_i &= \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \boldsymbol{\Sigma}_i \\ &= \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} d_{11} & 0 \\ 0 & d_{22} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \\ &= \begin{bmatrix} d_{11} + \sigma^2 & d_{11} \\ d_{11} & d_{11} + d_{22} + \sigma^2 \end{bmatrix}. \end{aligned}$$

Fazendo $\sigma^2 = \sigma_1^2$ e $d_{22} + \sigma^2 = \sigma_2^2$ a matriz de variância-covariância, \mathbf{V}_i , obtida com base no **modelo 2** tem a estrutura da obtida com base no **modelo 1**. Assim (pelo menos) dois modelos hierárquicos distintos implicam o mesmo modelo marginal.

2.3 Estimação dos Parâmetros no Modelo Marginal

A estimação dos parâmetros do modelo linear misto baseia-se, em geral, no modelo marginal sendo o método da máxima verossimilhança (método ML) e o método da máxima verossimilhança restrita (método REML) os dois métodos usados neste livro. Os parâmetros a estimar são os coeficientes de regressão $\boldsymbol{\beta}$ (o vector dos efeitos fixos) e as componentes da variância, isto é, o conjunto de parâmetros dos quais as matrizes de variância-covariância \mathbf{D} e $\boldsymbol{\Sigma}_i$ dependem.

No que se segue assume-se que $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}_{T_i}$ e $\mathbf{D} = \sigma^2 \mathbf{G}$, sendo a análise de outras estruturas para a matriz $\boldsymbol{\Sigma}_i$ feita no Capítulo 3. Ao modelo dado por (2.1) com $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}_{T_i}$, dá-se o nome de **modelo linear misto básico** (Pinheiro & Bates, 2000). A matriz de variância-covariância de \mathbf{Y}_i assume a forma $\mathbf{V}_i = \sigma^2 (\mathbf{Z}_i \mathbf{G} \mathbf{Z}_i^\top + \mathbf{I}_{T_i}) = \sigma^2 \mathbf{M}_i$, o mesmo se aplica a \mathbf{Y} com $\mathbf{V} = \sigma^2 (\mathbf{Z} \tilde{\mathbf{G}} \mathbf{Z}^\top + \mathbf{I}) = \sigma^2 \mathbf{M}$. Sempre que não se revele necessário explicitar σ^2 será usada a matriz \mathbf{V}_i em vez da matriz $\sigma^2 \mathbf{M}_i$ (\mathbf{V} em vez de $\sigma^2 \mathbf{M}$).

2.3.1 Método da Máxima Verossimilhança

Seja $\boldsymbol{\beta}$ o vector dos efeitos fixos, $\boldsymbol{\theta}$ o vector com todas as componentes da variância em \mathbf{G} e $\boldsymbol{\alpha}^\top = (\boldsymbol{\theta}^\top, \sigma^2)$. Assumindo a independência entre os indivíduos, a função verossimilhança é

$$L(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\alpha}) = \prod_{i=1}^n \frac{1}{(2\pi)^{T_i/2}} |\mathbf{V}_i(\boldsymbol{\alpha})|^{-1/2} \times \exp \left(- \frac{(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^\top \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})}{2} \right) \quad (2.8)$$

ou, alternativamente,

$$L(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) = \prod_{i=1}^n \frac{1}{(2\pi\sigma^2)^{T_i/2}} |\mathbf{M}_i(\boldsymbol{\theta})|^{-1/2} \times \exp\left(-\frac{(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \mathbf{M}_i^{-1}(\boldsymbol{\theta})(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})}{2\sigma^2}\right) \quad (2.9)$$

sendo os logaritmos de (2.8) e (2.9), dados respectivamente por:

$$\begin{aligned} \ell(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\alpha}) &= -\frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \log |\mathbf{V}_i(\boldsymbol{\alpha})| \\ &\quad - \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \mathbf{V}_i^{-1}(\boldsymbol{\alpha})(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \end{aligned} \quad (2.10)$$

e por:

$$\begin{aligned} \ell(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) &= -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2} \sum_{i=1}^n \log |\mathbf{M}_i(\boldsymbol{\theta})| \\ &\quad - \sum_{i=1}^n \frac{(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \mathbf{M}_i^{-1}(\boldsymbol{\theta})(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})}{2\sigma^2}. \end{aligned} \quad (2.11)$$

onde \log designa o logaritmo neperiano.

Os estimadores ML para os parâmetros do modelo são obtidos do seguinte modo:

- (i) Para um dado $\boldsymbol{\alpha}$, igualando a zero a derivada de (2.10) em ordem a $\boldsymbol{\beta}$ e resolvendo-a em ordem ao referido parâmetro, obtém-se o estimador ML para $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}_{ML}(\boldsymbol{\alpha})$ dado por:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{ML}(\boldsymbol{\alpha}) &= \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1} \mathbf{Y}_i \right) \\ &= \left(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{Y} \end{aligned} \quad (2.12)$$

que mais não é do que o estimador dos mínimos quadrados generalizados (GLS) de $\boldsymbol{\beta}$, assumindo $\boldsymbol{\alpha}$ conhecido.

30 Modelo Linear Misto

- (ii) Para um dado θ , igualando a zero as derivadas parciais de (2.11) em ordem a β e a σ^2 e resolvendo o sistema em ordem aos referidos parâmetros, obtêm-se os estimadores ML de β e de σ^2 dados por:

$$\hat{\beta}_{ML}(\theta) = \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{M}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{M}_i^{-1} \mathbf{Y}_i \right) \quad (2.13)$$

e

$$\hat{\sigma}_{ML}^2(\theta) = \frac{\sum_{i=1}^n \left(\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}(\theta) \right)^\top \mathbf{M}_i^{-1} \left(\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}(\theta) \right)}{N}. \quad (2.14)$$

- (iii) Quando α é desconhecido, o que acontece na maioria dos casos, o estimador ML de α é obtido maximizando-se (2.10) com respeito a α , após β ter sido substituído por (2.12)¹. O mesmo resultado é obtido maximizando-se (2.10) simultaneamente com respeito a α e β .

Os estimadores ML de α e β são designados por $\hat{\alpha}_{ML}$ e $\hat{\beta}_{ML}$, respectivamente, sendo o estimador de $\hat{\beta}_{ML}$ dado por

$$\begin{aligned} \hat{\beta}_{ML} &= \left(\sum_{i=1}^n \mathbf{X}_i^\top \hat{\mathbf{V}}_i^{-1}(\hat{\alpha}_{ML}) \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i^\top \hat{\mathbf{V}}_i^{-1}(\hat{\alpha}_{ML}) \mathbf{Y}_i \right) \\ &= (\mathbf{X}^\top \hat{\mathbf{V}}_{ML}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{V}}_{ML}^{-1} \mathbf{Y} \end{aligned} \quad (2.15)$$

onde $\hat{\mathbf{V}}_{ML} = \hat{\mathbf{V}}(\hat{\alpha}_{ML})$.

O processo de maximização de (2.8) e (2.9) (ou (2.10) e (2.11)), em qualquer um dos casos atrás referidos, requer métodos numéricos de optimização.

¹Ao logaritmo da função verosimilhança assim obtido dá-se o nome de logaritmo da função verosimilhança perfilada ou função log-verosimilhança perfilada

Um dos problemas do método ML quando se trata de estimar as componentes da variância é o dos estimadores obtidos não serem centrados. Por outro lado, ao usar-se a máxima verosimilhança para se estimar simultaneamente β , σ^2 e θ a forma da matriz do desenho, \mathbf{X} , é envolvida explicitamente na estimação de σ^2 e θ . Como consequência, uma escolha não apropriada daquela matriz pode levar a que os estimadores ML das componentes da variância não sejam consistentes. Uma estratégia para combater a não consistência é a de aumentar o número p de colunas (isto é, de variáveis explicativas) em \mathbf{X} por “razões de segurança”, mas isto torna o problema de enviesamento ainda pior. Para resolver este conflito usa-se o método REML, também conhecido por máxima verosimilhança residual, para estimar as componentes da variância (Diggle et al., 2002).

2.3.2 Método da Máxima Verosimilhança Restrita

No método ML, a perda de graus de liberdade resultante da estimação dos efeitos fixos não é levada em consideração, originando estimadores enviesados das componentes da variância. Este problema é resolvido pelo método REML primeiro introduzido por Patterson & Thompson (1971) e mais tarde desenvolvido por Harville (1974, 1977).

O método REML é baseado na transformação linear dos dados, $\mathbf{U} = \mathbf{A}^\top \mathbf{Y}$, onde \mathbf{A} é uma matriz $N \times (N - p)$ de característica completa, ortogonal às colunas da matriz \mathbf{X} , tal que distribuição de \mathbf{U} não depende de β .

A distribuição de \mathbf{U} é $N(\mathbf{0}, \mathbf{A}^\top \mathbf{V}(\alpha) \mathbf{A})$ qualquer que seja β e aos elementos de \mathbf{U} dá-se o nome de contrastes de erros (Verbeke & Molenberghs, 2000).

Harville (1974) mostrou que a função verosimilhança para os contrastes dos erros (função verosimilhança restrita) pode ser escrita na

32 Modelo Linear Misto

forma

$$\begin{aligned}
 L_{REML}(\mathbf{y}; \boldsymbol{\alpha}) &= (2\pi)^{-(N-p)/2} \left| \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{X}_i \right|^{1/2} \\
 &\times \left| \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) \mathbf{X}_i \right|^{1/2} \prod_{i=1}^n |\mathbf{V}_i(\boldsymbol{\alpha})|^{-1/2} \\
 &\times \exp \left(-\frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^\top \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) \right)
 \end{aligned} \tag{2.16}$$

onde $\hat{\boldsymbol{\beta}}$ é dado por (2.12).

O logaritmo da função de verosimilhança restrita é então dado por:

$$\begin{aligned}
 \ell_R = \log L_{REML}(\mathbf{y}; \boldsymbol{\alpha}) &= \text{const} - \frac{1}{2} \log \left| \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) \mathbf{X}_i \right| \\
 &- \frac{1}{2} \sum_{i=1}^n \log |\mathbf{V}_i(\boldsymbol{\alpha})| \\
 &\times -\frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^\top \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}).
 \end{aligned} \tag{2.17}$$

Maximizando (2.17) em ordem a $\boldsymbol{\alpha}$ obtêm-se os estimadores REML de $\boldsymbol{\alpha}$, $\hat{\boldsymbol{\alpha}}_{REML}$.

Comparando a expressão do logaritmo dada por (2.10) com a dada por (2.17) verifica-se que a diferença entre elas reside no termo

$$-\frac{1}{2} \log \left| \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) \mathbf{X}_i \right|$$

existente em (2.17).

O logaritmo da máxima verosimilhança restrita pode então ser ex-

presso na forma (Verbeke & Molenberghs, 2000):

$$\begin{aligned} \ell_R = \log L_{REML}(\mathbf{y}; \boldsymbol{\alpha}) &= \text{const} - \frac{1}{2} \log \left| \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) \mathbf{X}_i \right| \\ &+ \log L_{ML}(\mathbf{y}; \hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}), \boldsymbol{\alpha}) \end{aligned} \quad (2.18)$$

Como $\left| \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) \mathbf{X}_i \right|$ em (2.18), não depende de $\boldsymbol{\beta}$, os estimadores de REML de $\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$ podem igualmente ser obtidos maximizando-se a chamada função verossimilhança restrita

$$L_{REML}(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\alpha}) = L_{ML}(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\alpha}) \left| \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) \mathbf{X}_i \right|^{-1/2} \quad (2.19)$$

com respeito a todos os parâmetros ($\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$), simultaneamente.

Quer se utilize (2.17) ou (2.19) para se obter o estimador REML para $\boldsymbol{\alpha}$, o estimador REML para $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}_{REML}$, é obtido substituindo em (2.15) \mathbf{V}_i por $\hat{\mathbf{V}}_i(\hat{\boldsymbol{\alpha}}_{REML})$ (ou \mathbf{V} por $\hat{\mathbf{V}}_{REML} = \hat{\mathbf{V}}(\hat{\boldsymbol{\alpha}}_{REML})$).

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{REML} &= \left(\sum_{i=1}^n \mathbf{X}_i^\top \hat{\mathbf{V}}_i^{-1}(\hat{\boldsymbol{\alpha}}_{REML}) \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i^\top \hat{\mathbf{V}}_i^{-1}(\hat{\boldsymbol{\alpha}}_{REML}) \mathbf{Y}_i \right) \\ &= (\mathbf{X}^\top \hat{\mathbf{V}}_{REML}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{V}}_{REML}^{-1} \mathbf{Y} \end{aligned} \quad (2.20)$$

É importante notar que, qualquer que seja \mathbf{A} , se obtêm os mesmos estimadores das componentes da variância, isto é, não dependem de \mathbf{A} e que o estimador de $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}_{REML}$, não é igual ao estimador $\hat{\boldsymbol{\beta}}_{ML}$.

A principal justificação do método REML foi dado por Patterson & Thompson (1971) ao provarem que, na ausência de informação sobre $\boldsymbol{\beta}$, nenhuma informação sobre $\boldsymbol{\alpha}$ é perdida quando a inferência é baseada em \mathbf{U} em vez de \mathbf{Y} . Harville (1974) mostrou, de um ponto de vista Bayesiano, que usando somente contrastes de erros para fazer inferências sobre $\boldsymbol{\alpha}$ é equivalente a usar todos os dados para fazer essas inferências, mas ignorando qualquer distribuição *a priori* sobre $\boldsymbol{\beta}$.

2.3.3 Máxima Verosimilhança *versus* Máxima Verosimilhança Restrita

Em relação aos dois métodos apresentados convém notar que o método ML fornece estimadores para os efeitos fixos, enquanto que o método REML, por si só, não.

Por outro lado, tendo em atenção o termo $\frac{1}{2} \log \left| \sum_{i=1}^n \mathbf{X}_i^T \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) \mathbf{X}_i \right|$ em (2.18) conclui-se que quaisquer alterações na matriz \mathbf{X} dos efeitos fixos alteram também $\log L_{REML}(\mathbf{y}; \boldsymbol{\alpha})$ enquanto que os estimadores ML não variam face a reparametrizações dos efeitos fixos (Pinheiro & Bates, 2000). Este aspecto inviabiliza a comparação, com base na função verosimilhança restrita, de modelos lineares mistos com diferentes estruturas de efeitos fixos.

Os dois métodos (ML e REML) produzem habitualmente estimativas semelhantes, embora a diferença aumente com o aumento do número de termos fixos no modelo (Zuur et al., 2009). Em particular, as estimativas ML das componentes da variância são menores das que as obtidas por REML (Pinheiro & Bates, 2000).

2.4 Predição dos Efeitos Aleatórios

Em muitas situações o interesse principal reside na estimação dos parâmetros do modelo marginal (os efeitos fixos $\boldsymbol{\beta}$ e as componentes da variância) no entanto a obtenção dos preditores dos efeitos aleatórios é igualmente importante. Essa importância tem a ver com o facto dos efeitos aleatórios \mathbf{b}_i serem variáveis aleatórias e reflectirem o “desvio” na evolução do i -ésimo indivíduo em relação ao valor esperado da população $\mathbf{X}_i\boldsymbol{\beta}$. Assim, os seus preditores podem ser interpretados como resíduos que podem ser úteis na detecção de perfis

especiais (Molenberghs & Verbeke, 2005).

Nesta secção vai-se assumir que o modelo (2.1) com as condições (2.2) é apropriado.

2.4.1 Melhor Preditor Linear Centrado (BLUP)

Atendendo a que os \mathbf{b}_i são variáveis aleatórias, é natural usarem-se os métodos Bayesianos para obter os seus preditores.

Na Secção 2.2 viu-se que a distribuição de \mathbf{Y}_i condicional a \mathbf{b}_i e a distribuição de \mathbf{b}_i são, respectivamente,

$$\mathbf{Y}_i | \mathbf{b}_i \sim N(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \boldsymbol{\Sigma}_i)$$

e

$$\mathbf{b}_i \sim N(0, \mathbf{D})$$

então, de acordo com a abordagem Bayesiana, a distribuição *a priori* para os \mathbf{b}_i é $N(0, \mathbf{D})$.

A distribuição *a posteriori* de \mathbf{b}_i , definida como a distribuição de \mathbf{b}_i condicional a $\mathbf{Y}_i = \mathbf{y}_i$ ($\mathbf{b}_i | \mathbf{y}_i$), pode ser calculada. Designando por $f(\mathbf{b}_i | \mathbf{y}_i)$ a densidade *a posteriori* de \mathbf{b}_i dado $\mathbf{Y}_i = \mathbf{y}_i$ tem-se

$$f(\mathbf{b}_i | \mathbf{y}_i) \equiv f(\mathbf{b}_i | \mathbf{Y}_i = \mathbf{y}_i) = \frac{f(\mathbf{y}_i | \mathbf{b}_i) f(\mathbf{b}_i)}{\int f(\mathbf{y}_i | \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i}.$$

Com base nas propriedades do modelo Gaussiano (Azzalini, 1996) a distribuição conjunta de \mathbf{b}_i e de \mathbf{Y}_i é Gaussiana multivariada

$$\begin{bmatrix} \mathbf{b}_i \\ \mathbf{Y}_i \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ \mathbf{X}_i \boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} \mathbf{D} & \mathbf{DZ}_i^\top \\ \mathbf{Z}_i \mathbf{D} & \mathbf{Z}_i \mathbf{DZ}_i^\top + \boldsymbol{\Sigma}_i \end{bmatrix} \right)$$

de onde se conclui que

$$\mathbf{b}_i | \mathbf{y}_i \sim N(\mathbf{DZ}_i^\top (\mathbf{Z}_i \mathbf{DZ}_i^\top + \boldsymbol{\Sigma}_i)^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}), \mathbf{V}_{11,2})$$

36 Modelo Linear Misto

com $\mathbf{V}_{11,2} = \mathbf{D} - \mathbf{D}\mathbf{Z}_i^\top (\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top + \boldsymbol{\Sigma}_i)^{-1}\mathbf{Z}_i\mathbf{D}$.

O valor esperado da distribuição *a posteriori* é então dado por

$$\begin{aligned} E(\mathbf{b}_i | \mathbf{Y}_i = \mathbf{y}_i) &= \int \mathbf{b}_i f(\mathbf{b}_i | \mathbf{y}_i) d\mathbf{b}_i \\ &= \mathbf{D}\mathbf{Z}_i^\top (\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top + \boldsymbol{\Sigma}_i)^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \\ &= \mathbf{D}\mathbf{Z}_i^\top \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}). \end{aligned}$$

Assumindo que $\boldsymbol{\alpha}$ é conhecido, o melhor preditor linear centrado (BLUP) de \mathbf{b}_i , é obtido substituindo, na expressão anterior, $\boldsymbol{\beta}$ por $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y}$ (McCulloch & Searle, 2001) e tem-se

$$\tilde{\mathbf{b}}_i(\boldsymbol{\alpha}) = \mathbf{D}\mathbf{Z}_i^\top \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}))$$

ou, alternativamente

$$\tilde{\mathbf{b}}(\boldsymbol{\alpha}) = \tilde{\mathbf{D}}\mathbf{Z}^\top \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}))$$

Condicional a $\boldsymbol{\alpha}$, o preditor da combinação linear $\mathbf{u} = \mathbf{a}_\beta^\top \boldsymbol{\beta} + \mathbf{a}_b^\top \mathbf{b}_i$ do vector dos efeitos fixos $\boldsymbol{\beta}$ e do vector \mathbf{b}_i dos efeitos aleatórios para os vectores \mathbf{a}_β e \mathbf{a}_b conhecidos de dimensão $p \times 1$ e $q \times 1$ respectivamente, é dado por

$$\tilde{\mathbf{u}}(\boldsymbol{\alpha}) = \mathbf{a}_\beta^\top \hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) + \mathbf{a}_b^\top \tilde{\mathbf{b}}_i(\boldsymbol{\alpha})$$

e prova-se que $\tilde{\mathbf{u}}(\boldsymbol{\alpha})$ é o BLUP de \mathbf{u} (Harville, 1976; Searle et al., 1992).

2.4.2 Estimação de Bayes Empírica

Na secção anterior foi assumido que as componentes da variância $\boldsymbol{\alpha}$ eram conhecidas. No caso em que não são conhecidas, a utilização de uma abordagem Bayesiana completa requer que sejam especificadas as distribuições *a priori* de $\tilde{\mathbf{D}}(\mathbf{D})$ e $\boldsymbol{\Sigma}(\boldsymbol{\Sigma}_i)$ e, posteriormente, proceder à integração em relação aos parâmetros (Davidian & Giltinan, 1995).

Esta tarefa é complicada pelo que na prática se opta por substituir $\tilde{\mathbf{D}}(\mathbf{D})$ e $\Sigma(\Sigma_i)$ pelas suas estimativas ML ou REML.

A esta abordagem de substituir certos parâmetros pelas suas estimativas é designada por estimação de Bayes empírica. Ao valor daí resultante para os efeitos aleatórios dá-se o nome de **estimativa de Bayes empírica (empirical Bayes- EB)**. Os estimadores **EB** são designados por $\hat{\mathbf{b}}(\hat{\mathbf{b}}_i)$ (Molenberghs & Verbeke, 2005).

2.4.3 Shrinkage

Para ilustrar a interpretação dos estimadores **EB**, considera-se o preditor, $\hat{\mathbf{Y}}_i$, do perfil do i -ésimo indivíduo:

$$\begin{aligned}\hat{\mathbf{Y}}_i &= \mathbf{X}_i\hat{\boldsymbol{\beta}} + \mathbf{Z}_i\hat{\mathbf{b}}_i \\ &= \mathbf{X}_i\hat{\boldsymbol{\beta}} + (\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top + \mathbf{V}_i^{-1})(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}) \\ &= (\mathbf{I}_{T_i} - \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top + \mathbf{V}_i^{-1})\mathbf{X}_i\hat{\boldsymbol{\beta}} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top + \mathbf{V}_i^{-1}\mathbf{y}_i \\ &= (\Sigma_i\mathbf{V}_i^{-1})\mathbf{X}_i\hat{\boldsymbol{\beta}} + (\mathbf{I}_{T_i} - \Sigma_i\mathbf{V}_i^{-1})\mathbf{y}_i\end{aligned}\quad (2.21)$$

$\hat{\mathbf{Y}}_i$ é uma média pesada do perfil médio da população $\mathbf{X}_i\hat{\boldsymbol{\beta}}$ e dos dados \mathbf{y}_i , com pesos $\Sigma_i\mathbf{V}_i^{-1}$ e $(\mathbf{I}_{T_i} - \Sigma_i\mathbf{V}_i^{-1})$, respectivamente. O “numerador” de $\Sigma_i\mathbf{V}_i^{-1}$ é a matriz de variância-covariância dos erros e o “denominador” é a matriz de variância-covariância global. Assim, se a variabilidade dos erros é grande em comparação com a variabilidade entre os indivíduos (modelada através dos efeitos aleatórios) é dado mais peso a $\mathbf{X}_i\hat{\boldsymbol{\beta}}$ ao passo que será dado mais peso aos dados na situação contrária. Esta propriedade é conhecida na literatura Bayesiana por *shrinkage*. A expressão (2.21) mostra como o preditor do perfil do i -ésimo indivíduo é contraído (*shrinks*) para o perfil médio da população.

Esta propriedade é válida para qualquer combinação linear $\boldsymbol{\lambda}$ dos

38 Modelo Linear Misto

efeitos aleatórios

$$\text{var}(\boldsymbol{\lambda}^\top \hat{\mathbf{b}}_i) \leq \text{var}(\boldsymbol{\lambda}^\top \mathbf{b}_i)$$

basta ter em atenção que

$$\begin{aligned} \text{var}(\mathbf{b}_i) &= \text{var}[E(\mathbf{b}_i|\mathbf{y}_i)] + E[\text{var}(\mathbf{b}_i|\mathbf{y}_i)] \\ &= \text{var}(\tilde{\mathbf{b}}_i) + \text{valor positivo} \end{aligned}$$

onde $\tilde{\mathbf{b}}_i = E(\mathbf{b}_i|\mathbf{y}_i)$ é o preditor e portanto

$$\text{var}(\tilde{\mathbf{b}}_i) \leq \text{var}(\mathbf{b}_i).$$

2.4.4 Equações do Modelo Misto

Henderson (Henderson et al., 1959) mostrou que se pode obter o estimador dos efeitos fixos, $\boldsymbol{\beta}$, e os preditores dos efeitos aleatórios, \mathbf{b} , através da resolução de um sistema de equações lineares às quais se dá o nome de equações do modelo misto (Davidian & Giltinian, 1995; Verbeke & Molenberghs, 2000).

Para obter essas equações necessita-se da função densidade de probabilidade conjunta de \mathbf{Y} e \mathbf{b} . Tendo em atenção que $\mathbf{Y}|\mathbf{b} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \boldsymbol{\Sigma})$ e que $\mathbf{b} \sim N(\mathbf{0}, \tilde{\mathbf{D}})$ a densidade conjunta de \mathbf{Y} e \mathbf{b} é

$$\begin{aligned} f(\mathbf{y}, \mathbf{b}) &= f(\mathbf{y}|\mathbf{b})f(\mathbf{b}) \\ &= (2\pi)^{-N/2}|\boldsymbol{\Sigma}|^{-1/2} \\ &\times \exp\left(-\frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})}{2}\right) \\ &\times (2\pi)^{-qn/2}|\tilde{\mathbf{D}}|^{-1/2} \exp\left(-\frac{\mathbf{b}^\top \tilde{\mathbf{D}}^{-1}\mathbf{b}}{2}\right) \\ &= (2\pi)^{-(N+qn)/2}|\boldsymbol{\Sigma}|^{-1/2}|\tilde{\mathbf{D}}|^{-1/2} \\ &\times \exp\left(-\frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) + \mathbf{b}^\top \tilde{\mathbf{D}}^{-1}\mathbf{b}}{2}\right). \end{aligned}$$

Sendo o logaritmo dado por

$$\log f(\mathbf{y}, \mathbf{b}) = -\frac{N+qn}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \log |\tilde{\mathbf{D}}| + \left(-\frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) + \mathbf{b}^\top \tilde{\mathbf{D}}^{-1} \mathbf{b}}{2} \right)$$

Maximizando aquele logaritmo em relação a $\boldsymbol{\beta}$ e a \mathbf{b} simultaneamente obtém-se o sistema de equações

$$\begin{cases} \frac{\partial \log f(\mathbf{y}, \mathbf{b})}{\partial \boldsymbol{\beta}} = 0 \Leftrightarrow \mathbf{X}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) = 0 \\ \frac{\partial \log f(\mathbf{y}, \mathbf{b})}{\partial \mathbf{b}} = 0 \Leftrightarrow \mathbf{Z}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) - \tilde{\mathbf{D}}^{-1} \mathbf{b} = 0 \end{cases} \quad (2.22)$$

que sob a forma matricial é dado por:

$$\begin{bmatrix} \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X} & \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X} & \mathbf{Z}^\top \boldsymbol{\Sigma}^{-1} \mathbf{Z} + \tilde{\mathbf{D}}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} \\ \mathbf{Z}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} \end{bmatrix} \quad (2.23)$$

ao qual se dá o nome de **equações do modelo misto de Henderson**.

Para $\tilde{\mathbf{D}}$ e $\boldsymbol{\Sigma}$ conhecidas, os estimadores dos efeitos fixos, $\hat{\boldsymbol{\beta}}$, e os preditores dos efeitos aleatórios, $\tilde{\mathbf{b}}$, são dados, respectivamente, por

$$\begin{aligned} \hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) &= (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y} \\ &= \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1} \mathbf{y}_i \end{aligned} \quad (2.24)$$

e

$$\tilde{\mathbf{b}}(\boldsymbol{\alpha}) = \tilde{\mathbf{D}} \mathbf{Z}^\top \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})) \quad (2.25)$$

ou

$$\tilde{\mathbf{b}}_i(\boldsymbol{\alpha}) = \mathbf{D} \mathbf{Z}_i^\top \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})). \quad (2.26)$$

O estimador de $\boldsymbol{\beta}$ dado por (2.24) é também o estimador GLS de $\boldsymbol{\beta}$ (que é igualmente o estimador ML dado por (2.12)). O preditor de

\mathbf{b} dado por (2.25) é o melhor preditor linear centrado (BLUP) para \mathbf{b} (Searle et al., 1992).

Um dos méritos das equações do modelo misto é a sua maior facilidade de cálculo na medida em que não é necessário inverter a matriz \mathbf{V} de ordem $N \times N$. As inversas das matrizes $\tilde{\mathbf{D}}$ e $\mathbf{\Sigma}$ de ordem $qn \times qn$ e $N \times N$, respectivamente, são mais fáceis pois qn é geralmente menor do que N e $\mathbf{\Sigma}$ é, em geral, uma matriz diagonal. No entanto, para grandes conjuntos de dados, o cálculo computacional pode tornar-se intenso.

2.5 Algoritmos de Optimizaçãõ

A maximização do logaritmo da função verosimilhança assim como do logaritmo da função verosimilhança restrita, note-se que eles só diferem porque este último tem o termo extra $\frac{1}{2} \log \left| \sum_{i=1}^n \mathbf{X}_i^T \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) \mathbf{X}_i \right|$, requer métodos numéricos de optimização em virtude das equações envolvidas não serem lineares em $\boldsymbol{\alpha}$.

O algoritmo EM (*Expected-Maximization*) (Dempster et al., 1977; Laird & Ware, 1982; Laird et al., 1987; Lindstrom & Bates, 1988; Pinheiro & Bates, 2000) e o método Newton-Raphson (Thisted, 1988; Pinheiro & Bates, 2000) são os métodos numéricos de optimização usualmente utilizados e são os que se encontram implementados nos procedimentos de cálculo para os modelos lineares mistos em programas estatísticos como o R, o S-PLUS, o SAS e o SPSS.

Em Pinheiro & Bates (2000); Monte-Pegado (2003); Longford (1993, 2005) e Wolfinger et al. (1994) é descrita a utilização destes métodos numéricos para o modelo linear misto.

Algoritmo EM

O algoritmo EM é um método iterativo muito utilizado na estimação por máxima verossimilhança e por máxima verossimilhança restrita no caso em que se tem dados incompletos. No contexto dos modelos lineares mistos, a amostra completa será constituída por \mathbf{y}_i juntamente com \mathbf{b}_i e ϵ_i . Como estes últimos não são observados, as observações \mathbf{y}_i da variável resposta são consideradas uma amostra incompleta.

O algoritmo EM é um procedimento iterativo em que cada iteração envolve dois passos:

Passo E: cálculo de $E\left[\ell(\boldsymbol{\theta})|_{b|y}\right]$,

Passo M: maximização de $E\left[\ell(\boldsymbol{\theta})|_{b|y}\right]$.

Pode dizer-se que o passo M simula a estimação ML que seria possível se tivéssemos dados completos. Como se está perante dados incompletos, a maximização do passo M é realizada num substituto da verossimilhança produzido durante o passo E (Laird et al., 1987).

A escolha dos valores iniciais para as iterações é importante, uma vez que o algoritmo pode não atingir a convergência no caso de terem sido dados valores iniciais arbitrários. A velocidade de convergência deste método é geralmente linear, dependendo dos dados e da estrutura assumida para a matriz de variância-covariância.

É possível demonstrar (Dempster et al., 1977) que a verossimilhança aumenta em cada iteração. Por outro lado, importa salientar que as restrições de não negatividade para as componentes da variância são automaticamente satisfeitas em cada iteração, desde que os valores iniciais satisfaçam essa restrição (Davidian & Giltinian, 1993).

Algoritmo de Newton-Raphson

O algoritmo de Newton-Raphson é um dos métodos de maximização mais utilizados. Este algoritmo utiliza uma expansão de primeira ordem da função *score* (derivada do logaritmo da função verosimilhança).

A função *score* é um vector e portanto ao igualar este vector a $\mathbf{0}$ fica-se com um sistema de equações, geralmente, não lineares, que é necessário resolver para encontrar as estimativas ML ou REML para θ .

Cada iteração requer o cálculo da função *score* e da sua derivada (a matriz Hessiana do logaritmo da verosimilhança). Sob condições gerais, geralmente satisfeitas na prática, o algoritmo de Newton-Raphson tem convergência quadrática. O cálculo da matriz Hessiana em cada iteração pode ser uma tarefa complicada. No sentido de reduzir estes cálculos Thisted (1988) apresentou um método alternativo designado por Quasi-Newton.

Qualquer processo iterativo requer valores iniciais para os parâmetros. Estes valores podem obter-se com base em estudos anteriores ou a partir dos próprios dados.

EM *versus* Newton-Raphson

O algoritmo EM tem a vantagem de as iterações individuais serem relativamente fáceis de calcular e de as iterações iniciais se aproximarem rapidamente do óptimo. No entanto próximo do óptimo a convergência é lenta. Desta forma é por vezes difícil decidir se a iteração terminou ou não.

Para o método de Newton-Raphson as iterações individuais são mais complexas podendo ser muito instáveis quando muito longe do óptimo. No entanto, convergem rapidamente quando próximo dele.

Desta forma, é útil começar por produzir estimativas utilizando o algoritmo EM e quando perto do óptimo passar para o método de Newton-Raphson. Este esquema de optimização híbrido encontra-se implementado na função `lme` do R.

2.6 Inferência Estatística para o Modelo Marginal

O ajustamento de um modelo aos dados não é, em geral, o objetivo final de uma análise estatística. Na prática o interesse principal reside na inferência para os parâmetros do modelo de modo a que os resultados obtidos possam ser generalizados para a população, a partir da qual foi obtida a amostra.

O primeiro passo de uma inferência com base no estimador de um parâmetro é o conhecimento da distribuição desse estimador pelo que se vai começar pelas distribuições dos estimadores ML e REML dos vários parâmetros envolvidos no modelo.

2.6.1 Distribuições Assintóticas

O estimador de β

$$\hat{\beta}(\alpha) = \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1} \mathbf{Y}_i$$

sob a hipótese do modelo marginal dado por (2.6) e condicional a α , tem distribuição Gaussiana multivariada com valor esperado

$$E(\hat{\beta}) = \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1} E[\mathbf{Y}_i] \right) = \beta$$

44 Modelo Linear Misto

e matriz de variância-covariância

$$\begin{aligned} \text{var}(\hat{\boldsymbol{\beta}}) &= \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1} \text{var}(\mathbf{Y}_i) \mathbf{V}_i^{-1} \mathbf{X}_i \right) \\ &\times \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \\ &= \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \end{aligned}$$

tendo-se designado $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ por $\hat{\boldsymbol{\beta}}$ para simplificar a notação.

No entanto, na prática $\boldsymbol{\alpha}$ é desconhecido pelo se tem de recorrer a resultados assintóticos.

Pinheiro (1994, *in* Pinheiro & Bates (2000)) provou que, sob certas condições de regularidade, geralmente satisfeitas na prática, os estimadores ML são consistentes e a distribuição assintótica é Gaussiana multivariada. A matriz de variância-covariância aproximada dos estimadores é dada pela inversa da matriz de informação de Fisher correspondente ao logaritmo da função de verosimilhança (Cox and Hinkley (1974) *in* Pinheiro & Bates (2000)).

Dado que

$$E \left[\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\theta}^\top} \right] = 0 \quad e \quad E \left[\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \sigma^2} \right] = 0$$

os estimadores ML dos efeitos fixos $\boldsymbol{\beta}$ são assintoticamente não correlacionados com os estimadores ML de $\boldsymbol{\theta}$ e σ^2 .

A distribuição assintótica dos estimadores ML é:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &\sim N(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^\top \mathbf{M}^{-1}(\boldsymbol{\theta}) \mathbf{X})^{-1}) \\ \begin{bmatrix} \hat{\boldsymbol{\theta}} \\ \log \hat{\sigma} \end{bmatrix} &\sim N \left(\begin{bmatrix} \boldsymbol{\theta} \\ \log \sigma \end{bmatrix}, \mathbf{I}^{-1}(\boldsymbol{\theta}, \sigma) \right) \end{aligned} \quad (2.27)$$

com

$$\mathbf{I}(\boldsymbol{\theta}, \sigma) = \begin{bmatrix} \frac{\partial^2 \ell(\boldsymbol{\theta}, \sigma^2)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} & \frac{\partial^2 \ell(\boldsymbol{\theta}, \sigma^2)}{\partial \log \sigma \partial \boldsymbol{\theta}^\top} \\ \frac{\partial^2 \ell(\boldsymbol{\theta}, \sigma^2)}{\partial \boldsymbol{\theta} \log \sigma} & \frac{\partial^2 \ell(\boldsymbol{\theta}, \sigma^2)}{\partial^2 \log \sigma} \end{bmatrix}$$

onde $\ell(\boldsymbol{\theta}, \sigma^2)$ designa o logaritmo da função verosimilhança perfilada nos efeitos fixos e $\mathbf{I}(\boldsymbol{\theta}, \sigma)$ a matriz empírica de informação de Fisher. Foi utilizado $\log(\sigma)$, em vez de σ , de modo a fornecer uma parametrização sem restrições, para a qual a aproximação à distribuição Gaussiana tende a ser mais expedita.

Pinheiro (1994) mostrou igualmente que os estimadores REML são consistentes e a distribuição assintótica é Gaussiana multivariada. As expressões são idênticas às anteriores bastando substituir $\ell(\boldsymbol{\theta}, \sigma^2)$ por $\ell_R(\boldsymbol{\theta}, \sigma^2)$, logaritmo da função verosimilhança restrita perfilada nos efeitos fixos.

Na prática, os parâmetros desconhecidos $\boldsymbol{\theta}$ e σ^2 são substituídos pelos seus respectivos estimadores ML ou REML nas expressões que dão as matrizes de variância-covariância aproximadas em (2.27) (Pinheiro & Bates, 2000; Verbeke & Molenberghs, 2000).

A construção de testes e intervalos de confiança para os parâmetros fixos do modelo linear misto baseiam-se nas distribuições assintóticas dos estimadores.

2.6.2 Inferência para os Efeitos Fixos

Teste de Razão de Verosimilhanças

O teste de razão de verosimilhanças é utilizado quando se pretende comparar vários modelos encaixados, diferindo somente na estrutura de efeitos fixos.

46 Modelo Linear Misto

A estatística de teste é dada por:

$$2 \log \left(\frac{L_1}{L_0} \right) = 2(\log L_1 - \log L_0) \quad (2.28)$$

onde L_1 é a verosimilhança do modelo mais geral, isto é, com mais parâmetros, L_0 a do modelo encaixado e $k_1 - k_0$ é a diferença entre o número de parâmetros dos dois modelos. Sob a hipótese nula, o modelo mais restrito é adequado, a distribuição assintótica da estatística de teste é um qui-quadrado com $k_1 - k_0$ graus de liberdade ($\chi^2_{(k_1 - k_0)}$).

Note-se que este teste só é válido se os estimadores dos parâmetros fixos nos dois modelos forem estimados pelo método ML. Este aspecto deve-se ao facto do logaritmo da função verosimilhança restrita se alterar se as especificações dos efeitos fixos forem igualmente alteradas (Secção 2.3.3).

O teste de razão de verosimilhanças realizado nestas circunstâncias tende a ser anti-conservativo, isto é, o valor-p resultante é geralmente inferior ao verdadeiro valor-p do teste. Por outro lado, à medida que, no modelo mais restrito, o número de parâmetros removido dos efeitos fixos aumenta em comparação com o número total de observações, aumenta também a imprecisão dos valores-p resultantes (Pinheiro & Bates, 2000). Por este motivo, alguns autores recomendam a utilização dos testes-t e testes-F aproximados, para avaliar a significância dos efeitos fixos (Pinheiro & Bates, 2000; Bolker et al., 2008).

Teste - t e Teste - F Aproximados

Outra maneira de testar hipóteses envolvendo os termos fixos do modelo é usar o teste-t e o teste-F aproximados. Assim, para testar

$$H_0 : \beta_j = 0 \quad vs \quad H_1 : \beta_j \neq 0, \quad j = 1, \dots, p$$

utiliza-se a estatística do teste-t dada por:

$$\frac{\hat{\beta}_j}{\hat{\sigma}_{REML} \sqrt{\left[\left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{M}_i^{-1}(\hat{\boldsymbol{\theta}}) \mathbf{X}_i \right)^{-1} \right]_{jj}}} \quad (2.29)$$

que, sob a hipótese nula, tem distribuição assintótica t de Student com gl_j graus de liberdade.

Pinheiro & Bates (2000) designam este teste por teste-t condicional na medida em que a estatística de teste está condicionada ao estimador $\hat{\boldsymbol{\theta}}$ do vector $\boldsymbol{\theta}$ (o vector com todas as componentes da variância em \mathbf{G}), sendo σ^2 substituído pelo seu estimador REML. O teste-t aproximado avalia a significância marginal de cada coeficiente dos efeitos fixos quando todos os outros efeitos fixos estão presentes no modelo (Pinheiro & Bates, 2000).

No caso de se pretender testar

$$H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0} \quad vs \quad H_1 : \mathbf{L}\boldsymbol{\beta} \neq \mathbf{0}, \quad j = 1, \dots, p$$

em que \mathbf{L} é uma matriz conhecida, é usada a estatística de teste do teste F dada por:

$$F = \frac{\hat{\boldsymbol{\beta}}^\top \mathbf{L}^\top \left[\mathbf{L} \hat{\sigma}_{REML} \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{M}_i^{-1}(\hat{\boldsymbol{\theta}}) \mathbf{X}_i \right)^{-1} \mathbf{L}^\top \right] \mathbf{L} \hat{\boldsymbol{\beta}}}{r(\mathbf{L})} \quad (2.30)$$

que, sob a hipótese nula, tem distribuição assintótica F de Snedecor com (l, v) graus de liberdade. O número de graus de liberdade, l , do numerador do teste F é dado pela característica da matriz \mathbf{L} , $r(\mathbf{L})$. O teste-F aproximado testa a significância de termos nos efeitos fixos do modelo, que podem incluir vários coeficientes. Pinheiro & Bates (2000) designam este teste por teste-F condicional com base no argumento utilizado para o teste-t.

Quer para o teste-t quer para o teste-F aproximados o número de graus de liberdade do denominador, gl_j e v , respectivamente, necessitam de serem estimados a partir dos dados e existem diversos métodos (Verbeke & Molenberghs, 2000). Os diferentes métodos conduzem a diferentes resultados. No entanto, na análise de dados longitudinais, os diferentes indivíduos contribuem com informação independente, o que se traduz num número de graus de liberdade suficientemente grande, qualquer que seja o método utilizado para o estimar, e consequentemente leva a valores-p muito semelhantes (Verbeke & Molenberghs, 2000).

Intervalos de Confiança

Os intervalos de confiança aproximados para os efeitos fixos são construídos com base nas estatísticas dos testes-t aproximados.

Seja gl_j o número de graus de liberdade do denominador no teste-t aproximado correspondente ao j -ésimo efeito fixo. O intervalo de confiança aproximado para β_j ao nível de confiança $(1 - \alpha)$ é:

$$\hat{\beta}_j \pm t_{(gl_j, 1-\alpha/2)} \hat{\sigma}_{REML} \sqrt{\left[\left(\sum_{i=1}^n \mathbf{X}_i^T \mathbf{M}_i^{-1} (\hat{\boldsymbol{\theta}}) \mathbf{X}_i \right)^{-1} \right]_{jj}} \quad (2.31)$$

em que $t_{(gl_j, 1-\alpha/2)}$ representa o quantil $(1 - \alpha/2)$ da distribuição t de Student com gl_j graus de liberdade (Pinheiro & Bates, 2000).

2.6.3 Inferência para as Componentes da Variância

Na maior parte das vezes está-se interessado em estudar o comportamento médio da população e não o número de efeitos aleatórios existentes no modelo ou a estrutura de correlação existente entre eles. No entanto, a modelação desta matriz quer quanto ao número de efeitos

aleatórios quer quanto à estrutura de correlação existente entre eles é útil na interpretação da variação (aleatória) dos dados, sendo mesmo essencial para obter inferências válidas para os parâmetros do valor médio do modelo.

Com efeito, a sobreparametrização de um modelo conduz a uma ineficiente estimação e a uma avaliação deficiente do desvio padrão dos efeitos fixos enquanto que a subparametrização invalida as inferências feitas sobre esses parâmetros quando a estrutura de covariância não é correcta.

Assim sendo, há todo o interesse em se testar a necessidade de quais os efeitos aleatórios a incluir no modelo e qual a estrutura da matriz \mathbf{D} . A não ser que optemos por inferência bayesiana esta “necessidade” só pode ser testada indirectamente através do modelo marginal (Verbeke & Molenberghs, 2000).

Teste de Razão de Verosimilhanças

Para um dado β , a comparação de modelos encaixados com diferentes estruturas de variância-covariância dos efeitos aleatórios, corresponde a testar a hipótese nula de haver q efeitos aleatórios *versus* a hipótese alternativa de haver $q + k$ efeitos aleatórios, ou testar, por exemplo, a hipótese nula de os q efeitos aleatórios serem independentes *versus* a hipótese alternativa de não o serem.

O teste razão de verosimilhanças (Secção 2.6.2) é utilizado e, ao contrário do que acontece com os efeitos fixos, as funções de verosimilhança são comparáveis quando os parâmetros do modelo são estimados através do método REML, já que a estrutura fixa é a mesma nos dois modelos a serem comparados (Secção 2.3.3).

Uma das condições exigidas para que a estatística de teste tenha, assintoticamente, uma distribuição qui-quadrado com o número de

graus de liberdade igual à diferença entre as dimensões dos espaços parâmetro especificados em H_1 e H_0 , é a de que a hipótese nula não esteja na fronteira do espaço parâmetro. De acordo com Stram & Lee (1994), quando tal acontece, a distribuição assintótica da estatística de teste do teste de razão de verossimilhanças é uma mistura de qui-quadrados. No caso particular em que se testa a hipótese de existência de q efeitos aleatórios *versus* a hipótese alternativa de existirem $q + 1$ efeitos aleatórios, essa distribuição assintótica é uma mistura 50:50 de um χ_q^2 com um χ_{q+1}^2 e escreve-se $0.5\chi_q^2 + 0.5\chi_{q+1}^2$ (para mais detalhes ver Verbeke & Molenberghs (2000)). No entanto, segundo Pinheiro & Bates (2000), este ajustamento nem sempre é bem sucedido, pelo estes autores aconselham a utilização *naive* da distribuição assintótica de um qui-quadrado, com os graus de liberdade dados pela diferença entre os parâmetros estimados no modelo especificado nas hipóteses alternativa e nula, respectivamente. Esta solução é a que está implementada na biblioteca `nlme` do R e quando da sua utilização deve ter-se em conta que o valor-p calculado desta forma é conservativo, isto é, é maior do que o verdadeiro valor-p (Pinheiro & Bates, 2000).

Intervalos de Confiança para as Componentes da Variância

À semelhança do que acontece para os efeitos fixos, os intervalos de confiança aproximados para as componentes da matriz de variância-covariância são construídos à custa da distribuição assintótica dos estimadores ML e REML e dos testes t aproximados descritos na Secção 2.6.2 (Pinheiro & Bates, 2000).

O intervalo de confiança aproximado para o desvio padrão σ é obtido com base na distribuição (2.27). Designando por $[\mathbf{I}^{-1}]_{\sigma\sigma}$ o último elemento da diagonal da inversa da matriz de informação de Fisher, um intervalo de confiança aproximado ao nível de confiança $(1 - \alpha)$

para σ é:

$$\left[\hat{\sigma} \exp \left(-z_{(1-\alpha/2)} \sqrt{[\mathbf{I}^{-1}]_{\sigma\sigma}} \right), \hat{\sigma} \exp \left(z_{(1-\alpha/2)} \sqrt{[\mathbf{I}^{-1}]_{\sigma\sigma}} \right) \right] \quad (2.32)$$

em que $z_{(1-\alpha/2)}$ representa o quantil $(1 - \alpha/2)$ da Gaussiana padrão. Este intervalo de confiança pode ser utilizado, tanto para o estimador ML como para o estimador REML, com as necessárias alterações.

Para as componentes da matriz de variância-covariância dos efeitos aleatórios os intervalos de confiança aproximados são um pouco mais difíceis de construir e são estimados com menor precisão do que os dos efeitos fixos e do que o do desvio padrão dentro dos grupos. O aumento da precisão na estimação dos primeiros só é possível com o aumento do número de grupos estudados (Pinheiro & Bates, 2000).

Critérios de Informação

Até agora tem-se estado a explicitar o modelo da hipótese nula como um caso particular do modelo da hipótese alternativa, mas nem sempre é esse o caso. Quando tal acontece o teste de razão de verossimilhanças não é apropriado quer se esteja a testar a significância dos termos na estrutura dos efeitos fixos quer dos termos na estrutura dos efeitos aleatórios. A comparação entre modelos não encaixados é então feita com base nos critérios de informação.

A ideia base dos critérios de informação consiste em comparar modelos construídos com base na maximização do logaritmo da verossimilhança, penalizando os modelos com mais parâmetros. Desta forma, quanto menor for o valor do critério de informação de um modelo melhor é o modelo.

Os critérios de informação mais comuns são o Critério de Informação Akaike (AIC), proposto por Akaike (1974), e dado por:

$$AIC = -2\ell(\hat{\beta}, \hat{\alpha}) + 2n_{par} \quad (2.33)$$

52 Modelo Linear Misto

onde n_{par} é o número de parâmetros no modelo e o Critério de Informação Bayesiana (BIC), proposto por Schwarz (1978), e dado por:

$$BIC = -2\ell(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) + 2n_{par} \log(N) \quad (2.34)$$

onde n_{par} é o número de parâmetros no modelo e N é o número total de observação usadas no modelo ajustado.

Os dois critérios são muito semelhantes, sendo o critério BIC mais sensível ao número de parâmetros incluídos no modelo, penalizando o modelo que tem mais parâmetros.

No caso de se estar a trabalhar com a função verosimilhança restrita, ℓ é substituído por ℓ_R . Recorde-se que a comparação com base na máxima verosimilhança restrita só é possível para modelos com a mesma estrutura média ($\boldsymbol{\beta}$).

2.7 Inferência Estatística para os Efeitos Aleatórios

No que se refere aos efeitos aleatórios $\tilde{\mathbf{b}}_i$ tem-se que o valor esperado e a variância do BLUP de \mathbf{b}_i , $\tilde{\mathbf{b}}_i$, são dados respectivamente por:

$$\begin{aligned} E[\tilde{\mathbf{b}}_i(\boldsymbol{\alpha})] &= E[\mathbf{D}\mathbf{Z}_i^\top \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})] \\ &= \mathbf{D}\mathbf{Z}_i^\top \mathbf{V}_i^{-1} (\mathbf{X}_i \boldsymbol{\beta} - \mathbf{X}_i \boldsymbol{\beta}) = 0 = E[\mathbf{b}_i] \end{aligned} \quad (2.35)$$

e

$$\text{var}[\tilde{\mathbf{b}}_i(\boldsymbol{\alpha})] = \mathbf{D}\mathbf{Z}_i^\top \left[\mathbf{V}_i^{-1} - \mathbf{V}_i^{-1} \mathbf{X}_i \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \mathbf{X}_i^\top \mathbf{V}_i^{-1} \right] \mathbf{Z}_i \mathbf{D} \quad (2.36)$$

Note-se que (2.36) subestima a variabilidade em $\tilde{\mathbf{b}}_i(\boldsymbol{\alpha}) - \mathbf{b}_i$ uma vez que ignora a variação de \mathbf{b}_i . Como consequência as inferência para

\mathbf{b}_i são geralmente baseadas em

$$\text{var}[\tilde{\mathbf{b}}_i(\boldsymbol{\alpha}) - \mathbf{b}_i] = \mathbf{D} - \text{var}(\tilde{\mathbf{b}}_i(\boldsymbol{\alpha})) \quad (2.37)$$

como estimador da variação de $\tilde{\mathbf{b}}_i(\boldsymbol{\alpha}) - \mathbf{b}_i$ (Laird & Ware, 1982).

Na prática $\boldsymbol{\alpha}$ é substituído pelo seu estimador ML ou REML e, neste caso, quer (2.36) quer (2.37) subestimam a verdadeira variabilidade do estimador EB de \mathbf{b}_i , $\hat{\mathbf{b}}_i$, uma vez que não têm em consideração a variabilidade introduzida ao substituir-se $\boldsymbol{\alpha}$ pelo seu estimador.

A inferência sobre os \mathbf{b}_i é baseada, à semelhança dos efeitos fixos, em testes-t e testes-F aproximados, com processos idênticos de estimação do número de graus de liberdade (Verbeke & Molenberghs, 2000).

2.8 Análise de Resíduos

A análise de resíduos é o meio de diagnóstico utilizado para verificar se os pressupostos subjacentes ao modelo ajustado aos dados são válidos e conseqüentemente as inferências também o são. No caso do modelo linear misto as condições subjacentes a verificar são:

- os **erros aleatórios** dentro do grupo são independentes e identicamente distribuídos, com distribuição Gaussiana de valor médio nulo e variância constante σ^2 e são independentes dos efeitos aleatórios.
- os **efeitos aleatórios** têm distribuição Gaussiana com valor médio nulo e matriz de variância-covariância \mathbf{D} (não dependente do grupo) e são independentes para diferentes grupos.

A utilização dos gráficos de diagnóstico é a forma mais utilizada para a verificação destes pressupostos e é, em geral, a única disponibilizada nos programas estatísticos para estes modelos. A justificação deste facto é dada por Pinheiro & Bates (2000) ao referirem que a

validade das condições subjacentes ao modelo podem, igualmente, ser verificadas de uma maneira formal através de um teste de hipóteses. No entanto, raramente as conclusões baseadas num teste de hipóteses contradizem a informação contida num gráfico de diagnóstico.

2.8.1 Verificação das Condições sobre os Erros Aleatórios

A verificação dos pressupostos para os erros aleatórios baseia-se na análise dos resíduos dentro do grupo. Os resíduos consistem, na prática, na diferença entre a resposta observada e o respectivo valor ajustado pelo modelo dentro de cada grupo. Os resíduos dentro do grupo, condicionados às componentes da matriz de variância-covariância dos efeitos aleatórios, são as estimativas EB dos erros dentro do grupo, ϵ_i , já que as componentes da matriz de variância-covariância têm de ser substituídas pelas respectivas estimativas. Não obstante, fornecem geralmente uma boa estimativa dos erros dentro do grupo, pelo que são usados para validar os pressupostos impostos aos erros aleatórios (Pinheiro & Bates, 2000).

Os gráficos mais utilizados para validar as condições impostas aos erros aleatórios incluem as caixas de bigodes dos resíduos por grupo e os gráficos dos resíduos padronizados *versus* os valores ajustados ou *versus* as covariáveis de interesse. Aos gráficos anteriores acrescenta-se ainda o gráfico dos valores observados *versus* os valores estimados, o gráfico da função de autocorrelação empírica e/ou gráfico do semivariograma amostral assim como o papel de probabilidades da Gaussiana. Os tipos de resíduos utilizados incluem resíduos correntes (*raw residuals*), resíduos padronizados (ou de Pearson) e ainda resíduos normalizados. No caso do pressuposto da homocedasticidade ou da independência dos resíduos dos erros aleatórios ser violada procede-se à modelação da matriz de variância-covariância dos mesmos tal como

é descrito no Capítulo 3.

2.8.2 Verificação das Condições sobre os Efeitos Aleatórios

A verificação das hipóteses subjacentes aos efeitos aleatórios é feita com base nas suas estimativas EB. Não obstante, mesmo que o modelo linear misto esteja correctamente especificado, os estimadores EB, $\hat{\mathbf{b}}_i$, têm todas distribuições diferentes, a menos que as matrizes \mathbf{X}_i e \mathbf{Z}_i sejam as mesmas. Este facto levanta a questão da interpretabilidade dos gráficos habitualmente utilizados, nomeadamente, o papel de probabilidades e os histogramas das estimativas EB não padronizadas. Por este motivo, DeGruttola et al. (1991) propõem a padronização das estimativas EB antes da verificação do pressuposto da distribuição Gaussiana. Por outro lado, devido à propriedade de “contração” (Secção 2.4.2) os estimadores EB não reflectem na totalidade a heterogeneidade dos dados, pelo que o histograma das estimativas EB pode não ser adequado para detectar desvios à distribuição assumida.

Verbeke & Molenberghs (2000) fazem ainda notar que, se o interesse for somente a inferência para o modelo marginal, especialmente se o interesse recair sobre os efeitos fixos, obtêm-se inferências válidas mesmo que tenha sido assumida incorrectamente a hipótese da distribuição dos efeitos aleatórios ser Gaussiana.

Capítulo 3

Modelação da Matriz de Variância - Covariância dos Erros Aleatórios

3.1 Introdução

O modelo linear misto, tal como foi descrito no Capítulo 2, permite uma certa flexibilidade em relação à estrutura dos efeitos aleatórios mas impõe a condição $\Sigma_i = \sigma^2 \mathbf{I}_{T_i}$ para a estrutura de erros aleatórios. Esta condição, embora permita a aplicação do modelo linear misto a muitas situações práticas, quando considerada no contexto de dados longitudinais é pouco realista na medida em que neste tipo de dados as medições ou observações realizadas sobre o mesmo indivíduo estão geralmente correlacionadas pelo que aquela estrutura não é a adequada.

Neste capítulo serão apresentadas outras estruturas para a matriz Σ_i que permitem modelar não só a correlação mas também a heterocedasticidade dos erros aleatórios no modelo linear misto. O modelo com estas estruturas será designado por **modelo linear misto geral**.

A possibilidade de relaxar as condições impostas aos erros aleatórios permitindo a heterocedasticidade e/ou correlação dá a estes modelos uma grande flexibilidade na análise de dados reais.

Por último é apresentado o modelo sem efeitos aleatórios que mais não é do que o modelo linear geral para dados longitudinais apresentado

58 Modelação da Matriz de Variância - Covariância dos Erros Aleatórios
em Diggle et al. (2002).

3.2 O Modelo Linear Misto Geral

O modelo continua a ser dado por

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad (3.1)$$

mas com a generalização

$$\boldsymbol{\epsilon}_i \sim N(0, \sigma^2 \boldsymbol{\Lambda}_i), \quad i = 1, \dots, n.$$

onde $\boldsymbol{\Lambda}_i$ uma matriz $T_i \times T_i$ definida positiva parametrizada por um número, geralmente pequeno, de parâmetros que se designa por $\boldsymbol{\lambda}$ (Pinheiro & Bates, 2000).

Dado que $\boldsymbol{\Lambda}_i$ é uma matriz definida positiva, significa que admite raiz quadrada invertível $\boldsymbol{\Lambda}_i^{1/2}$ (Thisted, 1988), com inversa $\boldsymbol{\Lambda}_i^{-1/2}$, de modo que:

$$\boldsymbol{\Lambda}_i = (\boldsymbol{\Lambda}_i^{1/2})^\top \boldsymbol{\Lambda}_i^{1/2}$$

e

$$\boldsymbol{\Lambda}_i^{-1} = \boldsymbol{\Lambda}_i^{-1/2} (\boldsymbol{\Lambda}_i^{-1/2})^\top.$$

Considere-se a seguinte reparametrização do modelo:

$$\begin{aligned} \mathbf{Y}_i^* &= (\boldsymbol{\Lambda}_i^{-1/2})^\top \mathbf{Y}_i \\ \mathbf{X}_i^* &= (\boldsymbol{\Lambda}_i^{-1/2})^\top \mathbf{X}_i \\ \mathbf{Z}_i^* &= (\boldsymbol{\Lambda}_i^{-1/2})^\top \mathbf{Z}_i \\ \boldsymbol{\epsilon}_i^* &= (\boldsymbol{\Lambda}_i^{-1/2})^\top \boldsymbol{\epsilon}_i \end{aligned} \quad (3.2)$$

tendo em atenção que:

$$\begin{aligned} E[\boldsymbol{\epsilon}_i^*] &= (\boldsymbol{\Lambda}_i^{-1/2})^\top E[\boldsymbol{\epsilon}_i] = 0 \\ \text{var}[\boldsymbol{\epsilon}_i^*] &= (\boldsymbol{\Lambda}_i^{-1/2})^\top \text{var}[\boldsymbol{\epsilon}_i] (\boldsymbol{\Lambda}_i^{-1/2}) = \sigma^2 \mathbf{I} \end{aligned} \quad (3.3)$$

então:

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

Desta forma, pode-se reescrever o modelo (3.1) como

$$\mathbf{Y}_i^* = \mathbf{X}_i^* \boldsymbol{\beta} + \mathbf{Z}_i^* \mathbf{b}_i + \boldsymbol{\epsilon}_i^*, \quad (3.4)$$

$$\mathbf{b}_i \sim N(0, \mathbf{D}), \quad \boldsymbol{\epsilon}_i^* \sim N(0, \sigma^2 \mathbf{I}), \quad i = 1, \dots, n;$$

isto é, \mathbf{Y}_i^* é descrito através de um modelo linear misto básico (Pinheiro & Bates, 2000).

Atendendo a que $d\mathbf{y}_i^* = |\boldsymbol{\Lambda}_i^{-1/2}| dy_i$ a função verosimilhança para o modelo (3.1) com base numa amostra \mathbf{y} é dada por:

$$\begin{aligned} L(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\lambda}) &= \prod_{i=1}^n f(\mathbf{y}_i; \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\lambda}) \\ &= \prod_{i=1}^n f(\mathbf{y}_i^*; \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\lambda}) |\boldsymbol{\Lambda}_i^{-1/2}| \\ &= L(\mathbf{y}^*; \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\lambda}) \prod_{i=1}^n |\boldsymbol{\Lambda}_i^{-1/2}|. \end{aligned} \quad (3.5)$$

Tendo em atenção que a função verosimilhança para o modelo (3.4) corresponde à função verosimilhança do modelo básico os resultados obtidos no capítulo anterior são válidos. O mesmo se pode dizer em relação à função verosimilhança restrita do modelo (3.4) que é dada por:

$$L_{REML}(\mathbf{y}; \boldsymbol{\theta}, \sigma^2, \boldsymbol{\lambda}) = L_{REML}(\mathbf{y}^*; \boldsymbol{\theta}, \sigma^2, \boldsymbol{\lambda}) \prod_{i=1}^n |\boldsymbol{\Lambda}_i^{-1/2}|. \quad (3.6)$$

3.3 Decomposição da Matriz de Variância-Covariância

As matrizes $\mathbf{\Lambda}_i$ podem ser decompostas num produto de matrizes mais simples

$$\mathbf{\Lambda}_i = \mathbf{W}_i \mathbf{C}_i \mathbf{W}_i \quad (3.7)$$

onde \mathbf{W}_i é uma matriz diagonal e \mathbf{C}_i é uma matriz de correlação, ou seja, uma matriz definida positiva com todos os elementos da diagonal iguais a 1. A matriz \mathbf{W}_i pode não ser única, na medida em que se pode multiplicar cada uma das suas linhas por -1 e obter a mesma decomposição. Para garantir a sua unicidade vai-se impor que \mathbf{W}_i tenha todos os elementos da diagonal principal positivos. Por outro lado

$$var(\epsilon_{it}) = \sigma^2 [\mathbf{W}_i]_{tt}^2$$

e

$$corr(\epsilon_{it}, \epsilon_{it'}) = [\mathbf{C}_i]_{tt'},$$

pelo que \mathbf{W}_i descreve a variância e \mathbf{C}_i descreve a correlação dos erros ϵ_i dentro do grupo (medições repetidas sobre o mesmo indivíduo, tal como definido no Capítulo 1). Esta decomposição de $\mathbf{\Lambda}_i$ em duas componentes, uma de estrutura de variância e outra de estrutura de correlação, permite a modelação destas estruturas separadamente dando ao modelo linear misto geral grande flexibilidade.

3.4 Modelação da Heterocedasticidade

De acordo com a parametrização proposta por Davidian & Giltinian (1995) a função variância dos erros dentro do grupo associada ao modelo (3.1) pode ser definida da seguinte forma:

$$var(\epsilon_{it} | \mathbf{b}_i) = \sigma^2 g^2(\mu_{it}, \mathbf{v}_{it}, \boldsymbol{\delta}), \quad i = 1, \dots, n; t = 1, \dots, T_i, \quad (3.8)$$

onde $\mu_{it} = E[y_{it}|\mathbf{b}_i]$; \mathbf{v}_{it} é o vector de covariáveis da variância; $\boldsymbol{\delta}$ é o vector dos parâmetros da variância; $g(\cdot)$ é a função variância, contínua em $\boldsymbol{\delta}$. Esta função é escolhida de modo a reflectir a variabilidade intra-indivíduos, podendo ser, por exemplo, a função exponencial, logarítmica, potência ou uma combinação destas funções.

Esta formulação da função de variância proposta em (3.8) é bastante flexível e intuitiva na medida em que permite que a variância dentro do indivíduo dependa dos efeitos fixos $\boldsymbol{\beta}$ e dos efeitos aleatórios \mathbf{b}_i , através dos valores esperados μ_{it} . No entanto, como discutido por Davidian & Giltinian (1995), coloca alguns problemas teóricos e computacionais, na medida em que os erros dentro do grupo e os efeitos aleatórios deixam de ser independentes. Partindo do pressuposto que $E[\epsilon_{it}|\mathbf{b}_i] = 0$ então $var(\epsilon_{it}) = E[var(\epsilon_{it}|\mathbf{b}_i)]$ pelo que a dependência dos erros dentro do indivíduo, em relação aos efeitos aleatórios, pode ser evitada integrando-se em relação aos efeitos aleatórios.

Devido ao facto de a função variância $g(\cdot)$ não ser, em geral, linear em \mathbf{b}_i a integração em relação aos efeitos aleatórios, tendo como função de variância o modelo indicado em (3.8), é geralmente complicada do ponto de vista computacional. Davidian & Giltinian (1995) sugerem que se use um modelo aproximado em que os valores esperados μ_{it} são substituídos pelos seus BLUP $\hat{\mu}_{it} = x_{it}^\top \boldsymbol{\beta} + z_{it}^\top \hat{\mathbf{b}}_i$ em que x_{it} e z_{it} representam, respectivamente as t -ésimas linhas de \mathbf{X}_i e \mathbf{Z}_i , deste modo:

$$var(\epsilon_{it}) \approx \sigma^2 g^2(\hat{\mu}_{it}, \mathbf{v}_{it}, \boldsymbol{\delta}), \quad i = 1, \dots, n; t = 1, \dots, T_i. \quad (3.9)$$

Com esta aproximação os erros aleatórios e os efeitos aleatórios deixam de estar correlacionados, tal como acontece no modelo (3.1), pelo que os resultados obtidos na Secção 3.2 continuam válidos (Pinheiro & Bates, 2000).

É de salientar que se o modelo de variância condicional apresentado

62 Modelação da Matriz de Variância - Covariância dos Erros Aleatórios

em (3.8) não depender de μ_{it} , então o modelo (3.9) fornece a variância marginal exacta, não sendo por isso necessário recorrer a uma aproximação.

Na Tabela 3.1 são dadas algumas funções de variância que se encontram na biblioteca `nlme` do R, podendo ainda optar-se por uma combinação daquelas funções de variância.

Tabela 3.1 *Funções de variância.*

Descrição da classe	Variância ($\text{var}(\epsilon_{it})$)
Variância fixa	$\sigma^2 v_{it}$
Variâncias diferentes por estrato	$\sigma^2 \delta_{s_{it}}^2$
Potência de uma covariável	$\sigma^2 v_{it} ^{2\delta}$
Exponencial de uma covariável	$\sigma^2 \exp(2\delta v_{it})$
Constante + Potência de uma covariável	$\sigma^2 (\delta_1 + v_{it} ^{\delta_2})^2$

v_{it} -covariável; s_{it} -variável de estratificação; $\delta_1 > 0$

3.4.1 Selecção do Modelo

Na selecção do modelo que melhor se ajusta aos dados, o teste de razão de verosimilhanças (Secção 2.6.2) é o usado para comparar o modelo que assume a homocedasticidade com qualquer um dos modelos que incorporam a heterocedasticidade, já que estão encaixados. O mesmo acontece se duas estruturas de heterocedasticidade encaixadas são comparadas. Quando a comparação é feita entre duas estruturas não encaixadas são utilizados os critérios de informação AIC ou BIC (Secção 2.6.3). Note-se que os modelos têm a mesma estrutura fixa.

3.5 Modelação da Dependência

Historicamente as estruturas de correlação foram desenvolvidas para dois conjuntos de dados particulares: as séries temporais e os dados espaciais. Neste tipo de dados há, respectivamente, uma indexação das observações a uma variável tempo (unidimensional) ou a um vector espaço (bidimensional). No contexto do modelo linear misto, são usadas para modelar a dependência entre os erros aleatórios ϵ_{it} dentro do grupo e no que se segue é assumido que as estruturas de correlação são *isotrópicas* (Cressie, 1993), isto é, que a correlação entre dois erros ϵ_{it} e $\epsilon_{it'}$ depende dos vectores de posição \mathbf{p}_{it} e $\mathbf{p}_{it'}$, respectivamente a eles associados, apenas através da distância entre eles $d(\mathbf{p}_{it}, \mathbf{p}_{it'})$, e não dos valores particulares que assumem. A expressão geral para a estrutura de correlação dentro do grupo pode ser expressa como (Pinheiro & Bates, 2000):

$$\text{corr}(\epsilon_{it}, \epsilon_{it'}) = h[d(\mathbf{p}_{it}, \mathbf{p}_{it'}), \boldsymbol{\rho}], \quad i = 1, \dots, n; t = 1, \dots, T_i, \quad (3.10)$$

onde $\boldsymbol{\rho}$ é um vector de *parâmetros de correlação* e $h(\cdot)$ é uma função de correlação que assume valores entre -1 e 1, contínua em $\boldsymbol{\rho}$ e tal que $h(0, \boldsymbol{\rho}) = 1$ (Pinheiro & Bates, 2000). Desta forma, quanto mais próximos (no espaço ou no tempo) estiverem dois erros aleatórios, maior a sua dependência.

3.5.1 Estruturas de Correlação Serial

As estruturas de correlação serial são utilizadas para modelar a dependência em dados de séries temporais, isto é, com observações feitas sequencialmente ao longo do tempo. Simplificando o pressuposto de isotropia, o modelo geral de correlação serial é dado por:

$$\text{corr}(\epsilon_{it}, \epsilon_{it'}) = h[|\mathbf{p}_{it}, \mathbf{p}_{it'}|, \boldsymbol{\rho}]$$

onde a função de correlação $h(\cdot)$ é designada por função de autocorrelação.

Função de Autocorrelação Empírica

A função de autocorrelação empírica, que consiste numa estimativa não paramétrica da função de autocorrelação (Box et al., 1994), é bastante útil para verificar a correlação serial em dados de séries temporais. Sejam $r_{it} = (y_{it} - \hat{y}_{it})/\hat{\sigma}_{it}$ onde $\hat{\sigma}_{it}$ é o estimador de $var(\epsilon_{it})$, os resíduos padronizados do modelo linear misto ajustado. A função de autocorrelação empírica no espaçamento (*lag*) l é definida por:

$$\hat{\rho}(l) = \frac{\sum_{i=1}^n \sum_{t=1}^{T_i-l} r_{it}r_{i(t+l)}/N(l)}{\sum_{i=1}^n \sum_{t=1}^{T_i} r_{it}^2/N(0)} \quad (3.11)$$

onde $N(l)$ representa o número de pares de resíduos utilizados no somatório do numerador da função. Quando as observações estão igualmente espaçadas a função de autocorrelação empírica é o meio gráfico usado para identificar o processo. Analisando esta função pode-se concluir que:

- Se os valores se aproximam de zero gradualmente então o processo pode ser identificado como auto-regressivo.
- Se a função de auto correlação é consistente dentro de

$$\pm z_{(1-\alpha/2)}/\sqrt{N(l)}z_{(1-\alpha/2)}$$

após o *lag* 2 ou 3 então o modelo pode ser identificado como processo de médias móveis de *ordem* 1 ou 2.

As estruturas de correlação serial mais frequentes são:

(i) Geral

Cada correlação entre as observações é dada por um parâmetro diferente, correspondendo à função de correlação:

$$h(k, \rho) = \rho_k, \quad k = 1, 2, \dots$$

Como o número de parâmetros aumenta quadraticamente com o número máximo de observações dentro do grupo, esta estrutura conduz frequentemente a modelos sobreparametrizados, sendo útil apenas quando existem poucas observações por grupo.

(ii) Simetria Composta

Nesta estrutura assume-se uma correlação igual entre todos os erros aleatórios pertencentes ao mesmo grupo. O modelo de correlação correspondente é dado por:

$$\text{corr}(\epsilon_{it}, \epsilon_{it'}) = \rho, \forall t \neq t', \quad h(k, \rho) = \rho, \quad k = 1, 2, \dots$$

onde o único parâmetro de correlação ρ é designado por coeficiente de correlação intraclasse.

Auto-regressivo - Médias Móveis Esta família de estruturas de correlação inclui diferentes classes de modelos lineares estacionários (Box et al., 1994). Nestes modelos, assume-se que as observações são feitas em intervalos de tempo inteiros e, para simplificar a notação vai-se omitir o índice referente ao indivíduo, pelo que, ϵ_t designa a observação que ocorreu no instante de tempo t . A distância (*lag*), ou período de tempo, entre duas observações ϵ_t e ϵ_s é dada por $|t - s|$, pelo que o *lag1* se refere a observações feitas com uma unidade de distância, e assim sucessivamente.

(iii) Modelos Auto-regressivos

Estes modelos expressam uma determinada observação como uma função linear das observações anteriores acrescentada de um ruído homocedástico, a_t , que se assume ter valor médio nulo e ser independente das observações anteriores

$$\epsilon_t = \phi_1 \epsilon_{t-1} + \dots + \phi_p \epsilon_{t-p} + a_t.$$

O número, p , de observações anteriores incluídas no modelo linear é designado por ordem do modelo auto-regressivo e escreve-se $AR(p)$. Deste modo, existem p parâmetros de auto-correlação num modelo $AR(p)$ dados por $\Phi = (\phi_1, \dots, \phi_p)$.

- **Auto-regressivo de ordem 1 - $AR(1)$**

É o mais simples e um dos mais utilizados (Pinheiro & Bates, 2000). Neste modelo, os erros no tempo t são modelados como uma função dos erros no tempo $t - 1$, juntamente com o ruído, ou seja:

$$\epsilon_t = \phi \epsilon_{t-1} + a_t, \quad |\phi| < 1.$$

A respectiva função de correlação decresce exponencialmente, em valor absoluto, com o espaçamento (*lag*) e é dada por:

$$h(k, \phi) = \phi^k, \quad k = 0, 1, \dots$$

O modelo $AR(1)$ é uma das poucas estruturas de correlação serial que pode ser generalizada para tempo contínuo e designa-se por $CAR(1)$.

- **Auto-regressivo para tempo contínuo - $CAR(1)$**

A função de correlação é dada por:

$$h(s, \phi) = \phi^s, \quad s \geq 0, \phi \geq 0.$$

(iv) Modelos de Médias Móveis

Nestes modelos é assumida que determinada observação é uma função linear de termos de ruído, isto, é,

$$\epsilon_t = \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} + a_t.$$

O número, q , de termos de ruído incluídos no modelo linear é a ordem do modelo de médias móveis e designa-se por $MA(q)$. Neste modelo existem q parâmetros de correlação dados por $\theta = (\theta_1, \dots, \theta_q)$. A função de correlação para um modelo $MA(q)$ assume a forma:

$$h(k, \theta) = \begin{cases} \frac{\theta_k + \theta_1 \theta_{k-1} + \dots + \theta_{k-q} \theta_q}{1 + \theta_1^2 + \dots + \theta_q^2}, & k=1, \dots, q; \\ 0, & k=q+1, q+2, \dots \end{cases}$$

As observações separadas por mais do que q unidades de tempo não estão correlacionadas, uma vez que não partilham qualquer termo de ruído a_t .

- **Médias móveis de ordem 1- $MA(1)$**

É o mais simples dos modelos com

$$\epsilon_t = \theta_1 a_{t-1} + a_t,$$

e tem-se

$$h(1, \theta) = \rho_1 = \frac{\theta_1}{1 + \theta_1^2}, \quad |\rho_1| < 0.5.$$

(v) Modelos Auto-regressivo de Médias Móveis

Estes modelos, tal como o nome indica, são obtidos pela combinação dos modelos auto-regressivos com os modelos de médias móveis, e são designados abreviadamente por $ARMA$. Um modelo $ARMA(p, q)$ é dado por

$$\epsilon_t = \sum_{i=1}^p \phi_i \epsilon_{t-i} + \sum_{j=1}^q \theta_j a_{t-j} + a_t.$$

68 Modelação da Matriz de Variância - Covariância dos Erros Aleatórios

Neste modelo tem-se $p + q$ parâmetros de correlação, que correspondem à combinação dos p parâmetros auto-regressivos $\Phi = (\phi_1, \dots, \phi_p)$ e dos q parâmetros de médias móveis $\theta = (\theta_1, \dots, \theta_q)$.

Por convenção $ARMA(p, 0) = AR(p)$ e $ARMA(0, q) = MA(q)$; desta forma os modelos auto-regressivos e os modelos de médias móveis constituem um caso particular do modelo auto-regressivo de médias móveis geral.

A função de correlação de um modelo $ARMA(p, q)$ é dada por:

$$h(k, \rho) = \begin{cases} \phi_1 h(|k-1|, \rho) + \dots + \phi_p h(|k-p|, \rho) + \\ \theta_1 \psi(k-1, \rho) + \dots + \theta_q \psi(k-q, \rho), & k=1, \dots, q; \\ \phi_1 h(|k-1|, \rho) + \dots + \phi_p h(|k-p|, \rho), & k=q+1, \dots, \end{cases}$$

onde $\psi(k, \phi, \theta) = \frac{E[\epsilon_{t-k} a_t]}{var(\epsilon_t)}$. É de salientar que $\psi(k, \phi, \theta) = 0, k = 1, 2, \dots$ dado que, neste caso ϵ_{t-k} e a_t são independentes e $E[a_t] = 0$.

- $ARMA(1, 1)$

É um modelo “intermédio” entre os modelos $AR(1)$ e $MA(2)$ que apresenta um decaimento exponencial da função de auto-correlação para $lags \geq 2$ mas que permite maior flexibilidade na primeira auto-correlação (Pinheiro & Bates, 2000).

$$\begin{aligned} \epsilon_t &= \phi \epsilon_{t-1} + \theta a_{t-j} + a_t \\ h(1, \rho) &= \rho_1 = \frac{(1 + \phi_1 \theta_1)(\phi_1 + \theta_1)}{(1 + \theta_1^2 + 2\phi_1 \theta_1)} \\ h(k, \rho) &= \rho_k = \phi_1^{k-1} \rho_1, \quad k \geq 2. \end{aligned}$$

3.5.2 Estruturas de Correlação Espacial

As estruturas de correlação espacial foram originalmente desenvolvidas para modelar a dependência em dados associados a vectores de posição bidimensionais, tais como dados geoestatísticos, dados em

treliça ou padrões pontuais. No entanto, a propriedade de isotropia permite a sua generalização para outros casos, podendo ser utilizados em séries temporais (Diggle et al., 2002; Pinheiro & Bates, 2000).

Seja ϵ_x a observação correspondente à posição $\mathbf{x} = (x_1, \dots, x_r)^\top$. As estruturas de correlação espaciais isotrópicas podem apoiar-se em qualquer distância sendo a mais comum a distância euclidiana, ou L_2

definida por $d_E(\epsilon_x, \epsilon_y) = \sqrt{\sum_{i=1}^r (x_i - y_i)^2}$. Outras distâncias igualmente utilizadas são a distância Manhattan, ou L_1 , $d_{Man}(\epsilon_x, \epsilon_y) =$

$\sum_{i=1}^r |x_i - y_i|$ e a distância máxima $d_{Max}(\epsilon_x, \epsilon_y) = \max_{i=1, \dots, r} |x_i - y_i|$.

As estruturas de correlação espacial são as indicadas para dados onde as observações têm espaçamento irregular entre si (no tempo ou no espaço).

Semivariograma

As estruturas de correlação espacial são geralmente representadas através do seu semivariograma em vez da sua estrutura de correlação (Cressie, 1993). O semivariograma de uma estrutura espacial isotrópica com uma função distancia $d(\cdot)$ é definido por:

$$\gamma[d(\epsilon_x, \epsilon_y), \boldsymbol{\lambda}] = \frac{1}{2} \text{var}(\epsilon_x - \epsilon_y) = \frac{1}{2} E[\epsilon_x - \epsilon_y]^2.$$

Os erros dentro do grupo podem ser padronizados de modo a terem variância unitária, sem que a sua estrutura de correlação seja alterada. Assim, sem perda de generalidade, assume-se que $\text{var}(\epsilon_x) = 1, \forall x$. Neste caso, $\gamma(\cdot)$ vai depender exclusivamente dos parâmetros de correlação $\boldsymbol{\rho}$ e como

$$\gamma(s, \boldsymbol{\rho}) = \frac{1}{2} \text{var}(\epsilon_x - \epsilon_y) = 1 - \text{corr}(\epsilon_x, \epsilon_y),$$

tem-se

$$\gamma(s, \boldsymbol{\rho}) = 1 - h(s, \boldsymbol{\rho}),$$

sendo $\gamma(\mathbf{0}, \boldsymbol{\rho}) = 0$ pois $h(\mathbf{0}, \boldsymbol{\rho}) = 1$.

Um valor reduzido para $\gamma(s, \boldsymbol{\rho})$ revela dependência entre observações enquanto que valores superiores indicam independência. No semivariograma as distâncias entre observações são representadas no eixo das abcissas e, no eixo das ordenadas, figuram os valores estimados do semivariograma. Na Figura 3.1 está representado o semivariograma referente ao modelo exponencial com a indicação dos elementos componentes do mesmo. O **patamar** (*sill*) é o valor do semivariograma correspondente ao seu **alcance** (*range*), isto é, o valor da distância máxima ente observações a partir da qual já não existe dependência espacial entre observações.

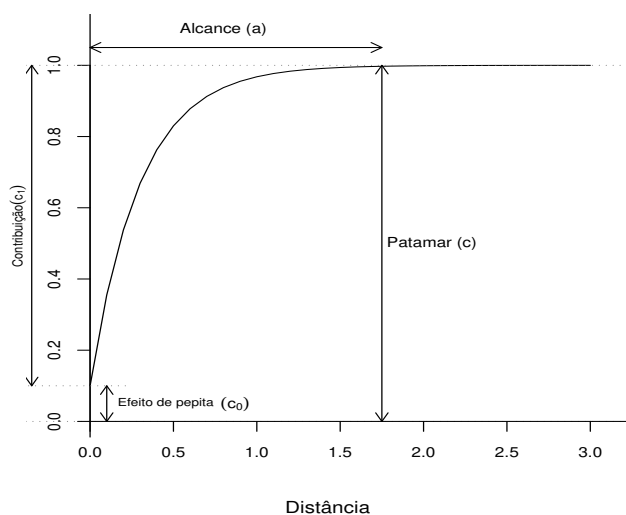


Figura 3.1 Semivariograma correspondente ao modelo exponencial.

No caso em que junto à origem há pouca variabilidade isso vai provocar alguma descontinuidade em $\gamma(\cdot)$, ou seja, nos valores assumidos por y quando a distância tende para zero. Para contornar esta questão é por vezes considerado o **efeito de pepita** c_0 , impondo-se que $\gamma(s, \boldsymbol{\rho}) \rightarrow c_0$ quando $s \rightarrow 0$, com $0 < c_0 < 1$. Em termos de função de correlação, o efeito de pepita traduz-se em $h(s, \boldsymbol{\rho}) \rightarrow 1 - c_0$ quando $s \rightarrow 0$. A **contribuição** (c_1) é a diferença entre o patamar (c) e o efeito de pepita (c_0).

O estimador clássico do semivariograma (Matheron, 1997) é:

$$\hat{\gamma}(s) = \frac{1}{2N(s)} \sum_{i=1}^{T_i} \sum_{d(p_{it}, p_{it'})=s} (r_{it} - r_{it'})^2$$

onde $N(s)$ é o número de pares de resíduos com distância s entre si. Como o estimador $\hat{\gamma}(s)$ utiliza os quadrados das diferenças entre pares de resíduos, pode ser bastante sensível a *outliers*, pelo que Cressie & Hawkins (1980) desenvolveram um estimador robusto dado por:

$$\hat{\gamma}(s) = \frac{\left(\frac{1}{2N(s)} \sum_{i=1}^{T_i} \sum_{d(p_{it}, p_{it'})=s} |r_{it} - r_{it'}|^{1/2} \right)^4}{\left(0.457 + \frac{0.494}{N(s)} \right)}$$

onde o termo do denominador é incluído para assegurar o não enviesamento (Jardim, 2001).

Existem diferentes estruturas de correlação espacial, que implicam formulações matemáticas distintas para a função $h(\cdot)$. Na Tabela 3.2 apresentam-se alguns dos modelos de variogramas isotrópicos (Pinheiro & Bates, 2000) e onde $I(s < \rho)$ representa uma variável binária que assume o valor 1 quando $s < \rho$ e 0 caso contrário.

As funções de correlação para estruturas como as da Tabela 3.2 po-

Tabela 3.2 *Alguns modelos de variograma isotrópicos para estruturas de correlação espacial.*

Esférico	$\gamma(s, \rho) = 1 - \left(1 - 1.5 \frac{s}{\rho} + 0.5 \left(\frac{s}{\rho}\right)^3\right) I(s < \rho)$
Exponencial	$\gamma(s, \rho) = 1 - \exp\left(-\frac{s}{\rho}\right)$
Gaussiano	$\gamma(s, \rho) = 1 - \exp\left[-\left(\frac{s}{\rho}\right)^2\right]$
Linear	$\gamma(s, \rho) = 1 - \left(1 - \frac{s}{\rho}\right) I(s < \rho)$
Quadrático Racional	$\gamma(s, \rho) = \frac{\left(\frac{s}{\rho}\right)^2}{1 + \left(\frac{s}{\rho}\right)^2}$
Efeito Pepita	$\gamma(s, c_0, \rho) = \begin{cases} c_0 + (1 - c_0)\gamma(s, \rho), & s > 0; \\ 0, & s = 0. \end{cases}$

dem obter-se usando a relação $h(s, \rho) = 1 - \gamma(s, \rho)$. Em estatística espacial o parâmetro ρ é referido como *range* (Littell et al., 1996).

Na Figura 3.2 estão representados os gráficos dos semivariogramas *versus* a distância para diferentes estruturas de correlação espacial.

Os modelos apresentados na Tabela 3.2 encontram-se na biblioteca `nlme` do R.

3.5.3 Selecção do Modelo

A selecção do modelo com a estrutura de correlação adequada é feita recorrendo-se ao teste de razão de verosimilhanças (Secção 2.6.2) sempre que as estruturas a comparar se encontrem encaixadas. Quando a comparação é feita entre duas estruturas não encaixadas são utilizados os critérios de informação AIC e BIC (Secção 2.6.3).

Uma vez modelada a matriz de variância-covariância dos erros aleatórios, procede-se novamente à verificação dos pressupostos para o

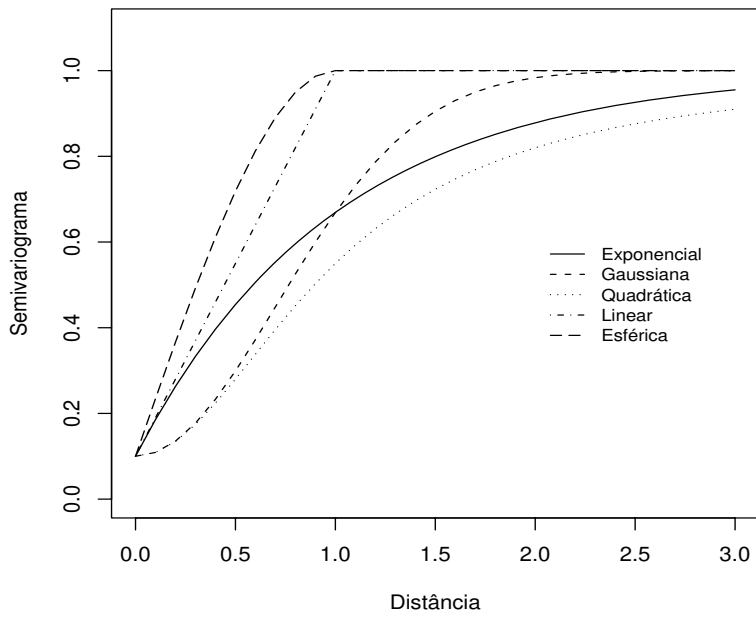


Figura 3.2 Gráficos de semivariogramas versus distância para correlações espaciais isotrópicas com $\rho = 1$ e efeito pepita=0.1.

novo modelo, recorrendo aos métodos descritos na Secção 2.8.

3.6 Efeitos Aleatórios *versus* Estrutura de Correlação dos Erros

Não é invulgar o modelo de efeitos aleatórios e o modelo de correlação dos erros dentro dos grupos competirem um com o outro, no sentido em que o número de efeitos aleatórios necessários é menor quando as estruturas de correlação dos erros dentro dos grupos estão presentes, e vice-versa. Com efeito a matriz de variância-covariância do modelo

linear misto geral

$$\text{var}(\mathbf{Y}_i) = \mathbf{V}_i = \sigma^2(\mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^\top + \mathbf{\Lambda}_i)$$

é constituída por duas componentes que podem ser usadas para modelar a heterocedasticidade e a correlação: uma componente associada aos **efeitos aleatórios**, $\mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^\top$, e uma componente, $\mathbf{\Lambda}_i$, **intra-grupo**. Nas sucessivas fases de construção do modelo que melhor se ajusta aos dados estas matrizes vão sendo modeladas. Pode acontecer que a matriz de variância-covariância, \mathbf{G} , dos efeitos aleatórios necessite de ser mais complexa quando se assume que $\mathbf{\Lambda}_i = \mathbf{I}_{T_i}$ e que após se permitir que $\mathbf{\Lambda}_i$ tenha outras estruturas diferentes de \mathbf{I}_{T_i} se verifique que essa complexidade já não é necessária, podendo em alguns casos não ser mesmo necessária a presença de efeitos aleatórios. O que se pode dizer é que existe um compromisso entre a complexidade das duas componentes de \mathbf{V}_i e que na modelação das mesmas se deve ter o cuidado de prevenir problemas de não-identificabilidade dos parâmetros no modelo.

3.7 Modelo Linear Geral para Dados Longitudinais Gaussianos

Em algumas aplicações práticas pode-se mesmo pretender evitar a existência de efeitos aleatórios para acomodar a dependência entre as observações preferindo-se utilizar a componente **intra-grupo** $\mathbf{\Lambda}_i$ para modelar directamente a estrutura de variância-covariância. O modelo para os dados é então

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2\mathbf{\Lambda}_i) \quad i = 1, \dots, n \quad (3.12)$$

ao qual se dá o nome de **modelo linear geral** para dados longitudinais (Diggle et al., 2002) A estimação dos parâmetros deste modelo,

assumindo que Λ_i é conhecido, é designado por problema dos mínimos quadrados generalizados.

Usando a transformação dada em (3.2), pode-se reescrever (3.12) como um modelo “clássico” de regressão linear (Pinheiro & Bates, 2000)

$$\mathbf{Y}_i^* = \mathbf{X}_i^* \boldsymbol{\beta} + \boldsymbol{\epsilon}_i^*, \quad \boldsymbol{\epsilon}_i^* \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{T_i}) \quad i = 1, \dots, n. \quad (3.13)$$

Assim, fixando $\boldsymbol{\lambda}$, os estimadores ML de $\boldsymbol{\beta}$ e de σ^2 são:

$$\begin{aligned} \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) &= \left(\mathbf{X}^{*\top} \mathbf{X}^* \right)^{-1} \left(\mathbf{X}^{*\top} \mathbf{Y}^* \right) \\ \sigma^2(\boldsymbol{\lambda}) &= \frac{\left(\mathbf{Y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) \right)^\top \left(\mathbf{Y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) \right)}{N}. \end{aligned} \quad (3.14)$$

A função log-verosimilhança perfilada que só é função de $\boldsymbol{\lambda}$, é obtida substituindo $\boldsymbol{\beta}$ e σ^2 na função verosimilhança da amostra pelos seus estimadores ML condicionados (3.14) e é dada por:

$$\begin{aligned} l(\mathbf{y}|\boldsymbol{\lambda}) &= \log L(\mathbf{y}|\boldsymbol{\lambda}) \\ &= \text{const} - \frac{N}{2} \log \left[\left(\mathbf{y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) \right)^\top \left(\mathbf{y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) \right) \right] \\ &\quad - \frac{1}{2} \sum_{i=1}^n \log |\Lambda_i|. \end{aligned} \quad (3.15)$$

Os estimadores de máxima verosimilhança são obtidos, primeiro otimizando (3.15) em relação a $\boldsymbol{\lambda}$ e depois substituindo em (3.14) $\boldsymbol{\lambda}$ pelo seu estimador ML $\hat{\boldsymbol{\lambda}}$ para obter os estimadores ML de $\boldsymbol{\beta}$ e de σ^2 (Pinheiro & Bates, 2000).

A função verosimilhança restrita para o modelo (3.12) é deduzida tal

como foi feito para o modelo linear misto dado por (3.1)

$$\begin{aligned}
 l_R(\mathbf{y}|\boldsymbol{\lambda}) &= \log L_{REML}(\mathbf{y}|\boldsymbol{\lambda}) \\
 &= \text{const} - \frac{N-p}{2} \log \left[\left(\mathbf{y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) \right)^\top \left(\mathbf{y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) \right) \right] \\
 &\quad - \log |\mathbf{X}^{*\top} \mathbf{X}^*| - \frac{1}{2} \sum_{i=1}^n \log |\boldsymbol{\Lambda}_i|. \tag{3.16}
 \end{aligned}$$

O estimador REML de σ^2 é dado por

$$\sigma_{REML}^2(\boldsymbol{\lambda}) = \frac{\left(\mathbf{Y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) \right)^\top \left(\mathbf{Y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) \right)}{N-p}. \tag{3.17}$$

O cálculo da maximização do logaritmo da função verosimilhança assim como do logaritmo da função verosimilhança restrita é feito recorrendo a métodos numéricos de optimização.

A modelação da matriz $\boldsymbol{\Lambda}_i$ é feita tal como foi descrito anteriormente assim como a selecção do modelo. No que se refere à análise de resíduos esta é feita usando-se os gráficos de diagnóstico, tal como no modelo linear misto.

3.7.1 Modelo Linear Misto *versus* Modelo Linear Geral

De acordo com Pinheiro & Bates (2000) a escolha entre o modelo linear misto e o modelo linear geral deve ter em conta os objectivos da análise e não se limitar ao valor do AIC ou ao valor-p dado com base no teste de razão de verosimilhanças. O modelo linear misto tem uma estrutura hierárquica que, em muitas aplicações, fornece uma maneira mais intuitiva de acomodar a dependência dentro do grupo do que a modelação directa da estrutura marginal de variância-covariância das respostas que é feita quando se usa o modelo linear geral. Além

disso, no modelo linear misto, obtêm-se os valores preditos dos efeitos aleatórios que podem, por si só, ter interesse.

O modelo linear geral para dados longitudinais, tal como o modelo linear misto, é válido para os mecanismos de omissão de dados MCAR e MAR.

A biblioteca `nlme` do R possui a função `gls` que permite ajustar o modelo dos mínimos quadrados generalizados usando quer a máxima verossimilhança quer a máxima verossimilhança restrita.

Capítulo 4

Equações de Estimação Generalizadas

4.1 Introdução

Nos dois últimos capítulos foram considerados dois métodos de análise de dados longitudinais quando a resposta é contínua, dando origem a dois modelos: o modelo linear misto e o modelo linear geral. Em qualquer um dos casos o modelo linear foi estendido de modo a acomodar a correlação entre observações de um mesmo indivíduo e a interpretação dos coeficientes de regressão é independente das hipóteses feitas acerca da correlação entre as medições repetidas, pelo que é a mesma nos dois modelos.

Quando a resposta é discreta, como é o caso em que se têm dados categorizados ou contagens, aqueles métodos não são apropriados pois o modelo linear deixa de ser adequado para relacionar as alterações do valor médio da resposta com as covariáveis. Os vários modelos para analisar dados longitudinais discretos são então obtidos através da extensão do modelo linear generalizado (GLM) do caso univariado para o multivariado.

Um outro aspecto relacionado com a análise de dados longitudinais discretos, e que se distingue da análise de dados longitudinais contínuos, é o facto de diferentes hipóteses acerca da fonte de associação entre observações de um mesmo indivíduo poder levar a coeficientes

de regressão com interpretações diferentes.

A necessidade de distinguir os modelos para analisar dados longitudinais discretos de acordo com a interpretação dos coeficientes leva ao uso de termos como modelos marginais e modelos com efeitos aleatórios (Diggle et al., 2002). Os primeiros são muitas vezes designados por *population-average models* (PA) e os segundos por *subject-specific models* (SS) (Zeger et al., 1988). Estes modelos, devido ao facto da interpretação dos coeficientes serem diferentes, têm objectivos de inferência distintos. Para os modelos marginais o objectivo da inferência é a população enquanto que para os modelos com efeitos aleatórios o objectivo da inferência é o indivíduo. A escolha do modelo para analisar os dados longitudinais discretos tem então de ser feita tendo em conta a questão a que se quer dar resposta.

Este capítulo é dedicado aos modelos marginais e a estimação dos parâmetros do modelo é feita usando o método de equações de estimação generalizadas (método GEE) tal como foi proposto por Liang & Zeger (1986). Nos últimos 20 anos esta metodologia tem-se mostrado extremamente útil na análise de dados longitudinais, especialmente quando a resposta é de natureza discreta podendo igualmente ser usada quando a variável resposta é contínua.

Este método baseia-se na estimação por quase-verosimilhança e como tal não necessita que a distribuição conjunta de \mathbf{Y}_i seja especificada. A sua popularidade deve-se ao facto de combinar, de forma simples e flexível, a modelação do valor médio da variável resposta em função das covariáveis, isto é, a componente de maior interesse, com a acomodação da dependência entre as observações sobre o mesmo indivíduo sem necessitar de especificar a estrutura de dependência (Fitzmaurice et al., 2004).

Em termos de dados omissos, a utilização deste método pressupõe o mecanismo de omissão de dados MCAR (em oposição ao mecanismos

de omissão de dados MAR assumido pelo modelo linear misto e pelo modelo linear geral).

No que se segue apresenta-se o modelo marginal e o método GEE e por último é feita uma referência à análise do modelo marginal quando a distribuição conjunta de \mathbf{Y}_i é especificada.

4.2 Estrutura do Modelo Marginal

Os modelos marginais são uma maneira natural de estender os modelos lineares generalizados a dados longitudinais (discretos ou contínuos). Num modelo marginal o valor esperado marginal, $E(Y_{it})$, é modelado como função das covariáveis. O termo valor esperado marginal é usado para realçar o facto de que não se está a condicionar o modelo a outras variáveis resposta ou a efeitos aleatórios não observáveis (Heagerty, 1999). Como as medições repetidas, em cada indivíduo, não têm tendência a ser independentes, a análise marginal tem que incluir pressupostos em relação à correlação. O modelo marginal tem a vantagem de o valor esperado da variável resposta e a covariância serem modelados separadamente (Diggle et al., 2002).

À semelhança dos capítulos anteriores seja y_{it} a resposta observada do i -ésimo indivíduo no tempo t , onde $t = 1, \dots, T_i$ e $i = 1, \dots, n$, realização da variável resposta Y_{it} . O vector das variáveis resposta para o i -ésimo indivíduo é dado por $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iT_i})^\top$ e $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})^\top$ representa o vector $p \times 1$ das covariáveis associado ao ponto (i, t) do desenho. Para cada componente de \mathbf{Y}_i , Y_{it} , tem-se $E(Y_{it}) = \mu_{it}$ e $\text{var}(Y_{it}) = v_{it}$.

O modelo linear pode ser escrito na forma

$$\begin{aligned} Y_{it} &= \beta_1 x_{it1} + \beta_2 x_{it2} + \dots + \beta_p x_{itp} + \epsilon_{it} \\ &= \mathbf{x}_{it}^\top \boldsymbol{\beta} + \epsilon_{it} \end{aligned}$$

82 Equações de Estimação Generalizadas

onde $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$ é o vector $p \times 1$ dos parâmetros de regressão desconhecidos e ϵ_{it} é uma variável aleatória com valor esperado zero.

Neste capítulo vai-se considerar que Y_{it} tem uma distribuição pertencente à família exponencial cuja densidade marginal é definida por:

$$f(y_{it}) = \exp \left[\frac{\omega_{it}}{\phi} (y_{it}\theta_{it} - c(\theta_{it})) + d(y_{it}, \phi) \right].$$

Especificamente, um modelo marginal para dados longitudinais tem as seguintes hipóteses:

1. A esperança marginal para cada variável resposta,

$$E(Y_{it}) = \mu_{it} = c'(\theta_{it}),$$

dependente das covariáveis, x_{it} , através de uma função de ligação,

$$g(\mu_{it}) = \eta_{it} = \mathbf{x}_{it}^\top \boldsymbol{\beta},$$

tal como a *identidade* para respostas contínuas, o *logit* para respostas binárias e o *log* para contagens;

2. A variância marginal depende do valor médio marginal de acordo com

$$\text{var}(Y_{it}) = v_{it} = V(\mu_{it}) \frac{\phi}{\omega_{it}} = c''(\theta_{it}) \frac{\phi}{\omega_{it}},$$

onde V é uma *função de variância* conhecida, ϕ é um parâmetro de dispersão, ou de escala, que pode ser conhecido ou estimado e ω_{it} é uma constante conhecida;

3. A correlação entre Y_{it} and $Y_{it'}$ é uma função do valor médio marginal e possivelmente de parâmetros adicionais, α ; e escreve-se $\text{corr}(Y_{it}, Y_{it'}) = \rho(\mu_{it}, \mu_{it'})$ onde $\rho(\cdot)$ é uma função conhecida.

A especificação do modelo marginal nas três hipóteses acima indicadas, torna evidente a extensão dos modelos lineares generalizados para dados longitudinais.

As duas primeiras hipóteses do modelo marginal descrevem o valor médio e a variância de Y_{it} , e correspondem ao modelo linear generalizado *standard*; a única diferença reside na passagem do caso univariado para o multivariado. É a terceira hipótese, com a incorporação da correlação entre as respostas referentes ao mesmo indivíduo, que representa a principal extensão dos modelos lineares generalizados para dados longitudinais.

Seguidamente vão-se considerar alguns exemplos de modelos marginais para dados longitudinais.

Modelo Marginal para Respostas Contínuas

Seja $\mathbf{x}_{it} = (1, t)^\top$ e $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ e considere-se Y_{it} uma variável resposta contínua sendo a *identidade* a função de ligação. As hipóteses do modelo marginal assumem a forma:

1. $\mu_{it} = \beta_0 + \beta_1 t$;
2. $\text{var}(Y_{it}) = \phi$;
3. $\text{corr}(Y_{it}, Y_{it'}) = \alpha^{|t-t'|}$, $0 \leq \alpha \leq 1$.

A correlação entre medições feitas sobre mesmo indivíduo é modelada através de uma função do parâmetro α via um modelo $AR(1)$.

Modelo Marginal para Contagens

Seja $\mathbf{x}_{it} = (1, t)^\top$ e $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ e considere-se Y_{it} uma variável resposta sob a forma de contagens e o

\log é a escolha natural da função de ligação. As hipóteses do modelo marginal assumem a forma:

1. $\log(\mu_{it}) = \beta_0 + \beta_1 t$;
2. $\text{var}(Y_{it}) = \phi \mu_{it}$;
3. $\text{corr}(Y_{it}, Y_{it'}) = \alpha_{tt'}$.

O modelo marginal descrito é um modelo de regressão *log-linear*, em que se tem como pressuposto uma função de variância de Poisson. A associação dentro de cada indivíduo é dada por uma estrutura de correlação não definida. Outras escolhas para a função de ligação e de variância são possíveis, assim como outras estruturas de correlação propostas por Liang & Zeger (1986).

Em muitas aplicações os dados sob a forma de contagens têm bastante variabilidade, ou seja, existe sobredispersão. A sobredispersão é uma regra, não a exceção, quando estamos na presença de dados longitudinais em que a variável resposta está sob a forma de contagens.

A sobredispersão pode ser considerada, incluindo um *parâmetro de dispersão* ou *escala*, ϕ , na especificação da variância.

Modelo Marginal para Dados Binários

Seja $\mathbf{x}_{it} = (1, t)^\top$ e $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ e considere-se Y_{it} uma resposta binária tomando os valores 0 e 1 e o logit é a escolha natural para a função de ligação. Neste caso as hipóteses do modelo marginal assumem a forma:

1. $\text{logit}(\mu_{it}) = \log\left(\frac{\mu_{it}}{1 - \mu_{it}}\right) = \log\left(\frac{\Pr(Y_{it} = 1)}{\Pr(Y_{it} = 0)}\right) = \beta_0 + \beta_1 t;$
2. $\text{var}(Y_{it}) = \mu_{it}(1 - \mu_{it});$
3. $\text{corr}(Y_{it}, Y_{it'}) = \alpha.$

Ao modelo marginal acima indicado, dá-se o nome de modelo de regressão logístico, o qual assume o pressuposto de variância de Bernoulli, $\text{var}(Y_{it}) = \mu_{it}(1 - \mu_{it})$.

Para descrever a terceira hipótese, tem que se ter atenção ao uso do termo *correlação*. Nas variáveis resposta discretas, as correlações estão condicionadas pelo valor médio da variável resposta, e vice-versa, como é o caso em que a variável resposta é binária.

Com efeito a correlação entre duas respostas binárias é dada por

$$\text{corr}(Y_{it-1}, Y_{it}) = \frac{\Pr(Y_{it-1} = 1, Y_{it} = 1) - \mu_{it-1}\mu_{it}}{[\mu_{it-1}(1 - \mu_{it-1})\mu_{it}(1 - \mu_{it})]^{1/2}},$$

onde $\Pr(Y_{it-1} = 1, Y_{it} = 1)$ satisfaz a desigualdade

$$\max(0, \mu_{it-1} + \mu_{it} - 1) < \Pr(Y_{it-1} = 1, Y_{it} = 1) < \min(\mu_{it-1}, \mu_{it}).$$

Por exemplo, se $\mu_{it-1} = E(Y_{it-1}) = \Pr(Y_{it-1} = 1) = 0.2$ e $\mu_{it} = E(Y_{it}) = \Pr(Y_{it} = 1) = 0.8$, então $\text{corr}(Y_{it-1}, Y_{it}) \leq 0.25$. O que significa que a correlação não pode exceder 0.25, quando as probabilidades de sucesso têm os valores 0.2 e 0.8.

Com este resultado verifica-se que, quando a variável resposta é binária, a correlação não é a medida mais natural de associação entre as respostas para um mesmo indivíduo pelo que, muitos autores utilizam o *odds ratio* ou o *log odds ratio* (Liang et al., 1992; Azzalini, 1994a; Gonçalves & Azzalini, 2008).

Em qualquer um dos três exemplos apresentados o vector β descreve o modo como o valor médio da população se comporta ao longo do tempo e não como a resposta do indivíduo depende da covariável.

4.3 Equações de Estimação Generalizadas

Ao analisarem-se as três hipóteses do modelo marginal, verifica-se que não é necessário qualquer pressuposto em relação à distribuição subjacente às observações, apenas se necessita de um modelo de regressão para o valor médio da variável resposta. Esta omissão de um pressuposto em relação à distribuição é vantajosa sobretudo no caso em que a variável resposta é discreta, onde especificar a distribuição conjunta multivariada de \mathbf{Y}_i não é trivial. Por outro lado, ao omitir-se a distribuição de \mathbf{Y}_i , os métodos de estimação baseados na máxima verosimilhança não podem ser utilizados pelo que se recorre ao método GEE.

O método GEE proposto por Liang & Zeger (1986) é a extensão multivariada do método de estimação por quase-verosimilhança (Wedderburn, 1974). Mais concretamente é a extensão do método de estimação por quase-verosimilhança quando o segundo momento não pode ser completamente especificado em termos de valor esperado e, além disso, é necessário estimar correlação adicional. Este método é também conhecido por GEE1 e em Molenberghs & Verbeke (2005) são descritas várias alternativas a este método.

O método GEE para $\boldsymbol{\beta}$ tem a mesma forma das equações *score* para os modelos univariados o que difere é a forma de escolher a matriz de variância-covariância.

Assim, tendo em atenção que as componentes de $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iT_i})^\top$ para cada indivíduo não são independentes, o método GEE para $\boldsymbol{\beta}$ proposto por Liang & Zeger (1986) é dado por

$$\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^n \mathbf{D}_i^\top \mathbf{V}_i [(\boldsymbol{\alpha})]^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \quad (4.1)$$

onde $\mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}$, é uma matriz $T_i \times p$, $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iT_i})^\top$, $\boldsymbol{\mu}_i$ o vector

do valor esperado de \mathbf{Y}_i e $\mathbf{V}_i(\boldsymbol{\alpha})$ é uma matriz $T_i \times T_i$ que para o i -ésimo indivíduo é dada por

$$\mathbf{V}_i(\boldsymbol{\alpha}) = \phi(\mathbf{A}_i^{1/2} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}), \quad (4.2)$$

à qual se dá o nome de matriz de variância-covariância “corrente”. O termo “corrente” é utilizado porque não se espera que seja, inicialmente, bem especificada.

Em (4.2) \mathbf{A}_i , é uma matriz diagonal $T_i \times T_i$, cujo elemento na diagonal t é $V(\mu_{it})$, função variância ($\text{var}(Y_{it}) = \phi V(\mu_{it})$) e $\mathbf{R}_i(\boldsymbol{\alpha}) = \text{corr}(\mathbf{Y}_i)$ é uma matriz $T_i \times T_i$ à qual se dá o nome de matriz de correlação “corrente” (*working correlation matrix*).

Se $\mathbf{R}_i(\boldsymbol{\alpha}) = \mathbf{I}$, matriz identidade $T_i \times T_i$, então o método GEE reduz-se às equações de quase-verosimilhança para um modelo linear generalizado que assume que as medições repetidas são independentes. Quando $\mathbf{R}_i(\boldsymbol{\alpha})$ está correctamente especificada então $\mathbf{V}_i(\boldsymbol{\alpha}) = \text{var}(\mathbf{Y}_i)$.

O sistema de equações (4.1), concebido para estimar $\boldsymbol{\beta}$, necessita de $\boldsymbol{\alpha}$ e ϕ mas, a estimação destes depende de $\boldsymbol{\beta}$. Este processo de dependência é quebrado através da utilização de um processo iterativo.

O processo iterativo proposto por Liang & Zeger (1986) divide-se em quatro passos:

1. Determinar-se as estimativas iniciais de $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}^{(0)}$, através do ajustamento de um modelo linear generalizado, assumindo que as observações são independentes;
2. (i) Calcular os resíduos de Pearson e_{it} .
 (ii) Calcular $\boldsymbol{\alpha}$.
 (iii) Calcular $\mathbf{R}_i(\boldsymbol{\alpha})$.
 (iv) Calcular ϕ .
 (v) Calcular $\mathbf{V}_i(\boldsymbol{\alpha}) = \phi(\mathbf{A}_i^{1/2} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2})$;

3. Actualizar a estimativa de β :

$$\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} - \left[\sum_{i=1}^n \hat{\mathbf{D}}_i^\top [\mathbf{V}_i(\hat{\alpha})]^{-1} \hat{\mathbf{D}}_i \right]^{-1} \times \left[\sum_{i=1}^n \hat{\mathbf{D}}_i^\top [\mathbf{V}_i(\hat{\alpha})]^{-1} (y_i - \hat{\mu}_i) \right]; \quad (4.3)$$

4. Repetir os passos 2-3 até o processo convergir.

As estimativas de α e ϕ são obtidas a partir dos resíduos de Pearson, e_{it} , definidos por

$$e_{it} = \frac{(y_{it} - \hat{\mu}_{it})}{\sqrt{\text{var}(\hat{\mu}_{it})}}$$

onde o parâmetro de dispersão ou de escala, ϕ , é estimado através de

$$\hat{\phi} = \frac{1}{n} \sum_{i=1}^n \frac{1}{T_i} \sum_{t=1}^{T_i} e_{it}^2$$

e a estimação de α , pode ser obtida de forma semelhante, dependendo da estrutura de correlação definida. Liang & Zeger (1986) sugeriram as estruturas de correlação apresentadas na Tabela 4.1.

Tabela 4.1 Estruturas de correlação propostas por Liang & Zeger (1986).

	$\text{corr}(Y_{it}, Y_{ik})$	Estimativa
Independente	0	—
Dependente	α	$\hat{\alpha} = \frac{1}{n} \sum_{i=1}^n \frac{1}{T_i(T_i - 1)} \sum_{t \neq k} e_{it} e_{ik}$
AR(1)	$\alpha^{ t-k }$	$\hat{\alpha} = \frac{1}{n} \sum_{i=1}^n \frac{1}{(T_i - 1)} \sum_{t < T_i - 1} e_{it} e_{i,t+1}$
Não Definida	α_{tk}	$\hat{\alpha} = \frac{1}{n} \sum_{i=1}^n e_{it} e_{ik}$

Outras estruturas de correlação podem ser consideradas, em particular, para o caso em que a variável resposta é binária em que a

correlação através do *odds-ratio* marginal é uma escolha natural para a estrutura de correlação.

4.4 Propriedades dos Estimadores GEE

Liang & Zeger (1986) mostraram que, sob certas condições de regularidade, o estimador de β , $\hat{\beta}$, solução das equações de estimação generalizadas, possui as seguintes propriedades:

1. É um estimador consistente de β .

Esta propriedade é válida quer se tenha ou não, modelado correctamente a estrutura de correlação. Desta forma, para que $\hat{\beta}$ proporcione estimativas válidas de β necessita-se apenas que o modelo para o valor médio da variável resposta tenha sido correctamente especificado. Esta é uma importante propriedade da robustez de $\hat{\beta}$, que torna a metodologia das equações de estimação generalizadas bastante apelativa nas mais variadas aplicações.

2. Para grandes amostras, a distribuição amostral de $\hat{\beta}$ é Gaussiana multivariada, com valor médio β e matriz de variância-covariância

$$I_0^{-1} I_1 I_0^{-1} \quad (4.4)$$

com

$$I_0 = \sum_{i=1}^n \mathbf{D}_i^\top [\mathbf{V}_i(\boldsymbol{\alpha})]^{-1} \mathbf{D}_i$$

$$I_1 = \sum_{i=1}^n \mathbf{D}_i^\top [\mathbf{V}_i(\boldsymbol{\alpha})]^{-1} \text{var}(Y_i) [\mathbf{V}_i(\boldsymbol{\alpha})]^{-1} \mathbf{D}_i$$

calculada em $(\hat{\beta}, \hat{\boldsymbol{\alpha}})$. Este estimador da matriz de variância-covariância é consistente à medida que o número de indivíduos

que contribuem para cada elemento da matriz tende para infinito (Diggle et al., 2002) e é conhecido como estimador *sandwich* de $var(\hat{\beta})$, I_0 é o “pão” e I_1 é o “fiambre”.

Em síntese, o método GEE possui um conjunto de propriedades interessantes para a estimação dos parâmetros de regressão no modelo marginal. Primeiro, porque, em muitos estudos longitudinais, o estimador $\hat{\beta}$ é quase tão preciso e eficiente como o obtido pelo método de máxima verosimilhança. Segundo, porque o estimador $\hat{\beta}$ para ser consistente só necessita que o modelo para o valor médio da variável resposta tenha sido correctamente especificado. Portanto, independentemente da matriz de correlação “corrente” ter sido ou não bem especificada, obtêm-se estimadores consistentes dos parâmetros de regressão.

Embora com a metodologia GEE se obtenham estimadores consistentes de β , mesmo quando a matriz correlação “corrente” não está correctamente especificada, os erros padrão usuais dos estimadores não são consistentes nesta situação. Felizmente, a utilização do estimador *sandwich* de $var(\hat{\beta})$ permite obter, em muitas situações, erros padrão válidos para $\hat{\beta}$. Baseados no estimador *sandwich* obtêm-se igualmente versões robustas do teste Wald.

Apesar da importante propriedade de somente ser necessário que o modelo para o valor médio da variável resposta seja correctamente especificado para obter estimadores consistentes de β , a modelação da covariância é bastante importante. São duas as razões para esse facto. Primeiro porque quanto mais próxima a matriz de variância-covariância, \mathbf{V}_i , estiver da verdadeira matriz de variância-covariância, maior é a eficiência ou precisão com a qual β pode ser estimado. Segundo, porque a propriedade de robustez é uma propriedade assintótica, que é válida apenas para amostras de grandes dimensões. Além disso, o estimador *sandwich* de $var(\hat{\beta})$ ajusta-se melhor em

estudos longitudinais equilibrados, onde o número de indivíduos é elevado e o número de medições repetidas é relativamente pequeno.

O estimador apresentado em (4.4) não é tão apelativo quando o número de indivíduos é pequeno (relativamente ao número de medições repetidas), quando o delineamento não é equilibrado ou quando os indivíduos não podem ser agrupados sob a mesma matriz de covariáveis. Em particular, quando o número de indivíduos é reduzido ($n \leq 30$), o estimador *sandwich* de $var(\hat{\beta})$ exhibe enviesamento pelo que Paik (1988) recomendou o uso do estimador *jackknife* de $var(\hat{\beta})$ dado por:

$$\frac{n-p}{n} \sum_{i=1}^n (\hat{\beta}_{-i} - \hat{\beta})(\hat{\beta}_{-i} - \hat{\beta})^\top$$

onde $\hat{\beta}_{-i}$ é o estimador de β obtido após o i -ésimo indivíduo ter sido removido.

4.5 Inferência Estatística com base nos Estimadores GEE

Quando se utiliza o método GEE o teste de razão de verosimilhanças não pode ser usado em virtude da estimação se ter baseado na quase-verosimilhança e não na verosimilhança. Este aspecto tem igualmente repercursões para inferências quando existem valores omissos. No entanto, tendo em atenção a distribuição assintótica dos estimadores GEE, o teste de Wald pode ser usado.

Por exemplo, quando o objectivo é testar

$$H_0 : \beta_j = 0 \quad vs \quad H_1 : \beta_j \neq 0,$$

para $j = 1, \dots, p$ a estatística de teste de Wald é dada por

$$W = \frac{\hat{\beta}_j}{\sqrt{var(\hat{\beta}_j)}}$$

92 Equações de Estimação Generalizadas

que, sob H_0 , tem uma distribuição assintótica Gaussiana e onde $var(\hat{\beta}_j)$ é o j -ésimo elemento da diagonal da matriz (4.4).

Para testar a hipótese mais geral

$$H_0 : \mathbf{L}\boldsymbol{\beta} = 0 \quad vs \quad H_1 : \mathbf{L}\boldsymbol{\beta} \neq 0,$$

onde \mathbf{L} é uma matriz $r \times p$ de característica completa r , a estatística de teste de Wald é dada por

$$\mathbf{W} = (\mathbf{L}\hat{\boldsymbol{\beta}})^\top \left[\mathbf{L}var(\hat{\boldsymbol{\beta}})\mathbf{L}^\top \right]^{-1} (\mathbf{L}\hat{\boldsymbol{\beta}})$$

que, sob H_0 , tem uma distribuição assintótica de um qui-quadrado com r graus de liberdade.

Fitzmaurice et al. (2008) consideram que embora a estatística do teste de Wald permita a comparação de modelos encaixados, esta comparação pode ser problemática no caso em que se tem dados longitudinais binários. Os mesmos autores apresentam como alternativa um teste baseado nos *scores* proposto por Rotnizky & Jewell (1990).

4.6 GEE e Valores Omissos

Quando há valores omissos as inferências com base em estimadores obtidos com o método GEE só são válidas se o mecanismo de omissão de dados é MCAR.

Quando o mecanismo de omissão de dados é MAR uma das soluções é o método de equações de estimação generalizadas com pesos ou pesado (WGEE - *weighted generalized estimating equations*) (Robins et al., 1995; Fitzmaurice et al., 1995). Uma das versões mais simples do método WGEE é a da contribuição individual para o método GEE *standard* ser pesada inversamente pela probabilidade de ser observada nos instantes de tempo dados.

Neste livro este tema não é desenvolvido remetendo-se o leitor para Molenberghs & Verbeke (2005); Molenberghs & Kenward (2007), ou Daniels & Hogan (2007).

O método GEE encontra-se implementado num grande número de programas estatísticos no entanto em muitos deles o parâmetro de escala ϕ é invariante com o tempo pelo que Fitzmaurice et al. (2004) não recomendam o seu uso na análise de dados longitudinais quando a resposta é contínua.

4.7 Análise de Resíduos

A análise gráfica dos resíduos continua a ser o método de diagnóstico utilizado. No entanto, quando se analisa dados longitudinais não-Gaussianos nem sempre é fácil essa análise, mesmo no caso simples de regressão logística para dados independente essa análise é difícil devido à extrema discretização dos dados binários.

Lin et al. (2002) desenvolveram técnicas de diagnóstico baseadas em “somas acumuladas” ou “somas móveis” de resíduos com o objectivo de distinguir o “sinal” do “ruído” (Fitzmaurice et al., 2008). Infelizmente esta metodologia não se encontra, até à data, nas bibliotecas do R específicas para o método GEE.

4.8 Distribuição Conjunta no Modelo Marginal. Máxima Verosimilhança

A formulação do modelo marginal nas três hipóteses apresentadas na Secção 4.2 não determina a distribuição conjunta de \mathbf{Y}_i . Para que os métodos de estimação baseados na máxima verosimilhança (ou na máxima verosimilhança restrita) possam ser utilizados é necessário

especificar essa distribuição conjunta.

Quando os dados longitudinais são Gaussianos, \mathbf{Y}_i segue uma distribuição Gaussiana multivariada e a análise do modelo marginal, usando o método ML ou o método REML foi apresentada nos Capítulos 2 e 3.

No caso em que os dados longitudinais são não-Gaussianos a primeira dificuldade surge logo na especificação da distribuição conjunta de \mathbf{Y}_i . É uma tarefa difícil que requer a especificação do vector esperado assim como a associação entre as respostas. A segunda dificuldade está relacionada com a complexidade computacional no cálculo da maximização da função verosimilhança.

Apesar das dificuldades alguns autores têm desenvolvido metodologias nesse sentido, em particular, para dados binários (Fitzmaurice & Laird, 1993; Fitzmaurice et al., 1993, 1994; Gonçalves & Azzalini, 2008).

Na metodologia descrita em Gonçalves & Azzalini (2008) para dados longitudinais binários, o método ML é o usado na estimação dos parâmetros fixos do modelo marginal e a dependência entre sucessivas observações de um mesmo indivíduo é considerada do tipo Markoviano de primeira ordem ψ_1 ou de segunda ordem ψ_2

$$OR(Y_{t-1}, Y_{t-2}) = \psi_1 = OR(Y_{t-1}, Y_t) \quad (4.5)$$

$$OR(Y_{t-2}, Y_t | Y_{t-1} = 0) = \psi_2 = OR(Y_{t-2}, Y_t | Y_{t-1} = 1) \quad (4.6)$$

onde a notação $OR(U, V)$ designa o *odds ratio* da distribuição conjunta de um par de variáveis aleatórias (U, V) e ψ_1 e ψ_2 são dois parâmetros positivos. Alguns padrões de valores omissos são permitidos assim como dados não equilibrados.

Esta metodologia encontra-se implementada na biblioteca `bild` do R.

Capítulo 5

Modelo Linear Generalizado Misto

5.1 Introdução

Os modelos lineares generalizados mistos (GLMM), propostos por Breslow & Clayton (1993), são uma extensão dos modelos lineares generalizados (GLM) que permitem a adição de componentes de variabilidade devida a efeitos não observados. Os GLMM são portanto, modelos lineares generalizados que incluem efeitos aleatórios no preditor linear além dos efeitos fixos.

A introdução destes efeitos aleatórios no preditor linear permite modelar a estrutura de correlação entre observações que pertencem ao mesmo indivíduo. A correlação entre observações de um mesmo indivíduo surge assim do facto destas partilharem os mesmos efeitos aleatórios não observáveis.

Os GLMM têm como objectivo descrever as alterações da resposta média de cada indivíduo e a relação destas alterações com as covariáveis. Assim sendo, estes modelos têm como finalidade realizar inferências para o indivíduo e não para a população.

A estrutura não linear associada à introdução de efeitos aleatórios levanta dificuldades não só ao nível computacional como também ao nível de interpretação de alguns parâmetros estimados.

Ao longo deste capítulo ir-se-á apresentar o modelo assim como algumas das metodologias utilizadas nos cálculos numéricos. A interpretação dos parâmetros do modelo é igualmente abordada e por último é feita a comparação dos GLMM com os modelos marginais na análise de dados longitudinais discretos. Em termos de dados omissos, estes modelos assumem os mecanismos de omissão de dados MCAR e MAR.

5.2 Estrutura do Modelo Linear Generalizado Misto

Tal como no capítulo anterior, seja y_{it} a resposta observada do indivíduo i para o tempo t realização da variável resposta Y_{it} onde $t = 1, \dots, T_i$ e $i = 1, \dots, n$, sendo $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})^\top$ o vector $p \times 1$ das covariáveis associado a cada vector resposta, Y_{it} e $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$ o vector $p \times 1$ dos parâmetros de regressão desconhecidos. Considere-se agora \mathbf{z}_{it} um vector $q \times 1$ de covariáveis (usualmente um subconjunto de \mathbf{x}_{it}) associado a um vector $q \times 1$ de efeitos aleatórios \mathbf{b}_i .

Os GLMM podem ser especificados do seguinte modo:

1. Dado \mathbf{b}_i , as variáveis resposta Y_{i1}, \dots, Y_{iT_i} são mutuamente independentes e seguem um GLM com densidade

$$f(y_{it}|\mathbf{b}_i) = \exp\left[\frac{\omega_{it}}{\phi}(y_{it}\theta_{it} - c(\theta_{it})) + d(y_{it}, \phi)\right].$$

O valor médio e a variância condicional são dados, respectivamente, por:

$$E(Y_{it}|\mathbf{b}_i) = \mu_{it}^b = c'(\theta_{it}),$$

e

$$\text{var}(Y_{it}|\mathbf{b}_i) = v_{it}^b = c''(\theta_{it})\frac{\phi}{\omega_{it}},$$

que se assume que satisfaçam,

$$g(\mu_{it}^b) = \mathbf{x}_{it}^\top \boldsymbol{\beta} + \mathbf{z}_{it}^\top \mathbf{b}_i,$$

e

$$v_{it}^b = V(\mu_{it}^b) \frac{\phi}{\omega_{it}},$$

onde $g(\cdot)$ é uma função de ligação e V uma função de variância, ambas conhecidas, ϕ é um parâmetro de dispersão ou de escala e ω_{it} é uma constante conhecida;

2. os efeitos aleatórios, \mathbf{b}_i , $i = 1, \dots, n$ são independentes entre si com uma distribuição multivariada comum F .

De uma maneira geral a equação do GLMM pode ser escrita na forma

$$g\{E(Y_{it}|\mathbf{b}_i)\} = \eta_{it}^b = \mathbf{x}_{it}^\top \boldsymbol{\beta} + \mathbf{z}_{it}^\top \mathbf{b}_i \quad (5.1)$$

com as especificações dadas em 1 e 2, sendo a distribuição comum aos efeitos aleatórios \mathbf{b}_i a Gaussiana multivariada com valor esperado zero e matriz de variância-covariância \mathbf{D} .

De forma a tornar mais clara a especificação do GLMM a partir do GLM vai-se considerar o modelo linear generalizado com efeitos aleatórios na intersecção e no declive quando a variável resposta assume a forma de contagens e quando é binária.

Seja $\mathbf{z}_{it} = \mathbf{x}_{it} = (1, t)^\top$, $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ e $\mathbf{b}_i = (b_{0i}, b_{1i})^\top$, onde se assume que \mathbf{b}_i é um vector aleatório independente com uma distribuição Gaussiana bivariada, com valor médio zero e matriz de variância-covariância, \mathbf{D} .

Modelo Linear Generalizado Misto para Contagens

Y_{it} é uma variável resposta sob a forma de contagens e o log é a escolha natural da função de ligação.

O GLMM tem as seguintes hipóteses:

1. Condicional ao vector de efeitos aleatórios, \mathbf{b}_i , Y_{it} são independentes e seguem uma distribuição de Poisson, com

$$\text{var}(Y_{it}|\mathbf{b}_i) = E(Y_{it}|\mathbf{b}_i), \quad (\phi = 1);$$

2. $\log(\mu_{it}^b) = \mathbf{x}_{it}^\top \boldsymbol{\beta} + \mathbf{z}_{it}^\top \mathbf{b}_i = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t$.

Modelo Linear Generalizado Misto para Dados Binários

Y_{it} é uma resposta binária que assume os valores 0 e 1 e o *logit* é a escolha natural para a função de ligação. Neste caso as hipóteses do GLMM assumem a forma:

1. Condicional ao vector de efeitos aleatórios, \mathbf{b}_i , Y_{it} são independentes e seguem uma distribuição de Binomial, com

$$\text{var}(Y_{it}|\mathbf{b}_i) = E(Y_{it}|\mathbf{b}_i)\{1 - E(Y_{it}|\mathbf{b}_i)\} = \mu_{it}^b(1 - \mu_{it}^b), \quad (\phi = 1);$$

2. $\text{logit}(\mu_{it}^b) = \mathbf{x}_{it}^\top \boldsymbol{\beta} + \mathbf{z}_{it}^\top \mathbf{b}_i = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t$.

5.3 Interpretação dos Coeficientes do Modelo

A introdução de efeitos aleatórios quer no modelo linear quer no modelo linear generalizado permite modelar a estrutura de correlação entre observações que pertencem ao mesmo indivíduo. No entanto, enquanto que no modelo linear esses efeitos não afectam a interpretação dos parâmetros fixos do modelo, o mesmo não se passa no modelo linear generalizado.

Com efeito $E(Y_{it})$, é dado por

$$E(Y_{it}) = E\{E(Y_{it}|\mathbf{b}_i)\} = E\left[g^{-1}(\mathbf{x}_{it}^\top \boldsymbol{\beta} + \mathbf{z}_{it}^\top \mathbf{b}_i)\right],$$

e devido à não linearidade da função $g^{-1}(\cdot)$ tem-se

$$E[g(Y_{it}|\mathbf{b}_i)] \neq g[E(Y_{it}|\mathbf{b}_i)].$$

Uma relação directa entre os parâmetros do modelo marginal e o do GLMM só existe em alguns casos especiais (Molenberghs & Verbeke, 2005).

Assim no GLMM têm-se as seguintes interpretações para os parâmetros fixos e para a componente aleatória do modelo.

- (i) **Parte fixa** As componentes de β , β_k , têm uma interpretação em termos específicos do indivíduo. Isto é, elas representam a influência das covariáveis na alteração da resposta esperada de um indivíduo específico. Porque a interpretação de β_k depende dos efeitos aleatórios do i -ésimo indivíduo assumirem um valor fixo (não observado), designam-se por coeficientes de regressão específicos do indivíduo (*subject-specific regression coefficients*).
- (ii) **Parte aleatória** Relativamente à componente aleatória uma maneira de interpretar as estimativas das variâncias dos efeitos aleatórios é produzir percentis dos efeitos baseados na hipótese Gaussiana. Por exemplo, para um modelo com efeito aleatório na intersecção, $\hat{\beta}_0 - 1.96\sqrt{\hat{d}_{11}}$ e $\hat{\beta}_0 + 1.96\sqrt{\hat{d}_{11}}$ dão-nos os percentis aproximados 2.5 e 97.5, respectivamente. Por outro lado, com base nos percentis assim obtidos e aplicando a inversa da função de ligação, obtém-se os limites de variação dos valores esperados condicionais μ_{it}^b . Uma outra maneira é a representação gráfica das trajectórias (perfis) condicionadas.

5.4 Estimação

Os GLMM permitem a acomodação de variáveis resposta com distribuição não-Gaussiana, a especificação de funções de ligação não lineares entre o valor esperado da resposta e os preditores, e podem modelar a sobredispersão e a correlação através da introdução de efeitos aleatórios. No sentido de serem feitas inferências correctas esta correlação tem de ser considerada e a distribuição conjunta, quer do vector das respostas, quer do vector dos efeitos aleatórios, tem de ser completamente especificada. Como resultado a função verosimilhança é usada e, conseqüentemente, a estimação e a inferência baseiam-se no método da máxima verosimilhança.

No GLMM a contribuição do i -ésimo indivíduo para a função de verosimilhança é

$$L_i^A(\boldsymbol{\beta}, \phi, \mathbf{D}) = \int \prod_{t=1}^{T_i} f(y_{it} | \mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i | \mathbf{D}) d\mathbf{b}_i. \quad (5.2)$$

O sobrescrito A em $L_i^A(\boldsymbol{\beta}, \phi, \mathbf{D})$ é a notação utilizada para indicar que se trata da função verosimilhança do modelo com efeitos aleatórios. Para os n indivíduos a função verosimilhança é apresentada na forma

$$\begin{aligned} L^A(\boldsymbol{\beta}, \phi, \mathbf{D}) &= \prod_{i=1}^n L_i^A(\boldsymbol{\beta}, \phi, \mathbf{D}) \\ &= \prod_{i=1}^n \int L_i^F(\boldsymbol{\beta}, \phi | \mathbf{D}) f(\mathbf{b}_i | \mathbf{D}) d\mathbf{b}_i, \end{aligned} \quad (5.3)$$

com $L_i^F(\boldsymbol{\beta}, \phi | \mathbf{D}) = \prod_{t=1}^{T_i} f(y_{it} | \mathbf{b}_i, \boldsymbol{\beta}, \phi)$, onde o sobrescrito F refere-se ao modelo com efeitos fixos.

O logaritmo da função verosimilhança é dado por

$$\ell^A(\boldsymbol{\beta}, \phi, \mathbf{D}) = \sum_{i=1}^n \log L_i^A(\boldsymbol{\beta}, \phi, \mathbf{D}). \quad (5.4)$$

que assume a expressão

$$\begin{aligned} \ell^A(\boldsymbol{\beta}, \phi, \mathbf{D}) &= \sum_{i=1}^n \log \int L_i^F(\boldsymbol{\beta}, \phi | \mathbf{D}) |\mathbf{D}|^{-1/2} \\ &\times \frac{1}{(2\pi)^{q/2}} \exp\left(-\frac{1}{2} \mathbf{b}_i^\top \mathbf{D}^{-1} \mathbf{b}_i\right) d\mathbf{b}_i. \end{aligned} \quad (5.5)$$

O problema chave da maximização de (5.3) ou, o que é equivalente, de (5.5) é a existência de n integrais no espaço q -dimensional do vector dos efeitos aleatórios. Em geral, o problema não tem uma solução analítica, pelo que se tem de recorrer ao cálculo numérico para realizar inferência com base na máxima verosimilhança. O cálculo numérico possui igualmente várias limitações, uma vez que envolve integrais complexos cuja dimensão depende da estrutura dos efeitos aleatórios.

Vários métodos têm sido propostos de forma a resolver este problema numérico. De um modo geral podem-se dividir as aproximações numéricas utilizadas em três grupos: as que são baseadas na aproximação dos dados, as que são baseadas na aproximação da função a integrar e as que são baseadas na aproximação do integral (Molenberghs & Verbeke, 2005).

5.4.1 Método de Aproximação dos Dados

A metodologia adoptada na aproximação dos dados consiste na decomposição dos dados em valor médio e uma componente apropriada, com um desenvolvimento em série de Taylor em torno do valor médio que é uma função não linear do preditor.

Considere-se o modelo,

$$Y_{it} = g^{-1}(\eta_{it}^b) = g^{-1}(\mathbf{x}_{it}^\top \boldsymbol{\beta} + \mathbf{z}_{it}^\top \mathbf{b}_i) + \epsilon_{it}, \quad (5.6)$$

onde $g^{-1}(\cdot)$ é a inversa da função de ligação, e $\eta_{it}^b = \mathbf{x}_{it}^\top \boldsymbol{\beta} + \mathbf{z}_{it}^\top \mathbf{b}_i$ é o preditor linear. O preditor linear é composto por uma componente

fixa, $\mathbf{x}_{it}^\top \boldsymbol{\beta}$, e por uma componente aleatória, $\mathbf{z}_{it}^\top \mathbf{b}_i$, em que $E(\mathbf{b}_i) = 0$, $\text{var}(\mathbf{b}_i) = \mathbf{D}$, $\text{var}(\boldsymbol{\epsilon}_i) = \sigma^2 \mathbf{I}$ e $\text{var}(\mathbf{b}_i, \boldsymbol{\epsilon}_i) = 0$. O objectivo é, portanto, linearizar a função de ligação de forma a realizar inferências para os parâmetros de regressão. Inserem-se nesta última metodologia o método de quase-verosimilhança penalizada (PQL), proposto por Breslow & Clayton (1993) e o método de quase-verosimilhança marginal (MQL) apresentado por Goldenstein (1991). A diferença essencial entre estes dois métodos reside no facto do MQL não incorporar o efeito aleatório \mathbf{b}_i no preditor linear (Molenberghs & Verbeke, 2005).

5.4.2 Método de Aproximação da Função a Integrar

A metodologia de aproximação da função a integrar tem como base a aproximação de Laplace. O método de Laplace (Tierny & Kadane, 1986) consiste em aproximar integrais na forma

$$I = \int e^{Q(\mathbf{b})} d\mathbf{b}, \quad (5.7)$$

em que $Q(\mathbf{b})$ é uma função limitada conhecida, de uma variável \mathbf{b} , com dimensão q . Seja $\hat{\mathbf{b}}$ o valor de \mathbf{b} para o qual Q é maximizada.

Considere-se o desenvolvimento de Taylor de segunda ordem, pelo que a função $Q(\mathbf{b})$ assume a forma

$$Q(\mathbf{b}) \approx Q(\hat{\mathbf{b}}) + \frac{1}{2} (\mathbf{b} - \hat{\mathbf{b}})' Q''(\hat{\mathbf{b}}) (\mathbf{b} - \hat{\mathbf{b}}), \quad (5.8)$$

onde $Q''(\hat{\mathbf{b}})$ corresponde à matriz Hessiana de Q em relação a $\hat{\mathbf{b}}$.

Substituindo $Q(\mathbf{b})$ em (5.7) pela sua aproximação em (5.8), obtém-se

$$\mathbf{I} \approx (2\pi)^{q/2} \left| -Q''(\hat{\mathbf{b}}) \right|^{-1/2} e^{Q(\hat{\mathbf{b}})}. \quad (5.9)$$

O integral da contribuição do i -ésimo indivíduo para a função de verosimilhança (5.2) é proporcional ao integral (5.7), para funções $Q(\mathbf{b})$ com a forma

$$Q(\mathbf{b}) = \phi^{-1} \sum_{t=1}^{T_i} \left[y_{it}(\mathbf{x}_{it}^\top \boldsymbol{\beta} + \mathbf{z}_{it}^\top \mathbf{b}_i) - c(\mathbf{x}_{it}^\top \boldsymbol{\beta} + \mathbf{z}_{it}^\top \mathbf{b}_i) \right] - \frac{1}{2} \mathbf{b}^\top \mathbf{D}^{-1} \mathbf{b},$$

onde o método de Laplace pode ser aplicado. O facto de $Q(\hat{\mathbf{b}})$ depender de $\boldsymbol{\beta}$, ϕ e \mathbf{D} , todos parâmetros desconhecidos, faz com que em cada iteração da maximização da verosimilhança, $\hat{\mathbf{b}}$ seja recalculado de forma condicional às estimativas destes parâmetros. A aproximação de Laplace é exacta quando $Q(\mathbf{b})$ é uma função quadrática de \mathbf{b} .

5.4.3 Método de Aproximação do Integral

Dos vários métodos de integração que podem ser usados, muitos dos quais se encontram implementados em vários programas estatísticos, salientam-se o método numérico da quadratura de Gauss-Hermite e o método numérico adaptativo da quadratura de Gauss-Hermite. Estes métodos baseiam-se na aproximação de integrais na forma

$$\int f(z)\phi(z)dz, \quad (5.10)$$

em que $f(z)$ é uma função densidade conhecida e $\phi(z)$ é a função densidade de uma distribuição Gaussiana multivariada. Começa-se por padronizar os efeitos aleatórios de forma a obter-se a matriz de variância-covariância. Seja $\boldsymbol{\delta}_i$, com $\boldsymbol{\delta}_i = \mathbf{D}^{-1/2} \mathbf{b}_i$, em que $\boldsymbol{\delta}_i$ segue uma distribuição de Gauss com valor médio zero e matriz de variância-covariância dada por \mathbf{I} . Desta forma o preditor linear assume a forma

$$\eta_{it} = \mathbf{x}_{it}^\top \boldsymbol{\beta} + \mathbf{z}_{it}^\top \mathbf{D}^{-1/2} \boldsymbol{\delta}_i$$

onde \mathbf{b}_i foi substituído por $\mathbf{D}^{-1/2} \boldsymbol{\delta}_i$.

A contribuição do i -ésimo indivíduo para a função de verosimilhança é, assim, dada por

$$\begin{aligned} L_i^A(\boldsymbol{\beta}, \phi, \mathbf{D}) &= \int \prod_{t=1}^{T_i} f(y_{it}|\mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i|\mathbf{D}) d\mathbf{b}_i \\ &= \int \prod_{t=1}^{T_i} f(y_{it}|\boldsymbol{\delta}_i, \boldsymbol{\beta}, \mathbf{D}, \phi) f(\boldsymbol{\delta}_i) d\boldsymbol{\delta}_i. \end{aligned} \quad (5.11)$$

A expressão (5.11) está portanto na forma da expressão (5.10), a qual é necessária para aplicar o método da quadratura de Gauss-Hermite (ou o método adaptativo da quadratura de Gauss-Hermite).

Método da Quadratura de Gauss-Hermite

No método da quadratura de Gauss-Hermite, $\int f(z)\phi(z)dz$ é aproximado por uma soma pesada,

$$\int f(z)\phi(z)dz \approx \sum_{q=1}^Q w_q f(z_q),$$

em que Q é a ordem da aproximação. Quanto maior for Q mais precisa é a aproximação. Os pontos de quadratura, z_q , são as soluções de um polinómio de Hermite de ordem Q e w_q são os pesos escolhidos. Os pontos de quadratura, z_q , e os pesos w_q encontram-se tabelados.

No caso de integração univariada, a aproximação consiste em subdividir a região de integração em intervalos e aproximar a superfície de integração numa superfície obtida através da aproximação de rectângulos. Dependendo do suporte de $f(z)$, os pontos de quadratura, z_q , corresponderão ou não à região de interesse. Neste sentido, a escolha de z_q é feita para que $f(z)\phi(z)$ se comporte, aproximadamente, como $\phi(z)$, ou seja, como uma função densidade de uma distribuição de Gauss. No caso presente a função $f(z)$ assume uma função densidade da família exponencial.

Método Adaptativo da Quadratura de Gauss-Hermite

No método adaptativo da quadratura de Gauss-Hermite, os pontos de quadratura são centrados e dimensionados para que $f(z)\phi(z)$ assumam uma distribuição Gaussiana. O valor médio desta distribuição será igual à estimativa, \hat{z} , de $\ln [f(z)\phi(z)]$, e a variância igual a

$$\left[-\frac{\partial^2}{\partial z^2} \ln [f(z)\phi(z)] \Big|_{z=\hat{z}} \right]^{-1}.$$

Assim sendo, a quadratura de pontos adaptativa assume a forma,

$$z_q^+ = \hat{z} + \left[-\frac{\partial^2}{\partial z^2} \ln [f(z)\phi(z)] \Big|_{z=\hat{z}} \right]^{-1/2} z_q$$

com os respectivos pesos,

$$w_q^+ = \left[-\frac{\partial^2}{\partial z^2} \ln [f(z)\phi(z)] \Big|_{z=\hat{z}} \right]^{-1/2} \frac{\phi(z_q^+)}{\phi(z_q)} w_q.$$

De forma análoga à quadratura de Gauss-Hermite, o integral é aproximado por uma soma pesada, dada por

$$\int f(z)\phi(z)dz \approx \sum_{q=1}^Q w_q^+ f(z_q^+).$$

De salientar o facto de que quando o método de quadratura de Gauss-Hermite ou o método adaptativo são ajustados a um GLMM é aplicada uma aproximação à contribuição dos n indivíduos na função verosimilhança. De uma forma geral, quanto maior for a ordem de Q melhor é a aproximação dos n integrais da verosimilhança.

Um caso particular é quando $Q = 1$. Seja $Q(z) = \ln [f(z)\phi(z)]$. Como $Q = 1$, $z_1 = 0$ e $w_1 = 1$, obtém-se $z_1^+ = \hat{z}$, o máximo de $Q(z)$. Desta forma, os pesos adaptativos assumem a forma,

$$w_q^+ = |Q''(\hat{z})|^{-1/2} \frac{\phi(\hat{z})}{\phi(0)} = (2\pi)^{q/2} |Q''(\hat{z})|^{-1/2} \frac{e^{Q(\hat{z})}}{f(\hat{z})}.$$

Pelo que,

$$\int f(z)\phi(z)dz = \int e^{Q(\hat{z})}dz$$

$$\approx w_1^+ f(z_1^+) = (2\pi)^{q/2} |Q''(\hat{z})|^{-1/2} e^{Q(\hat{z})}.$$

Assim, o método adaptativo da quadratura de Gauss-Hermite é equivalente à aproximação do integral que utiliza o método de aproximação de Laplace quando $Q = 1$.

Na maioria dos casos o método adaptativo da quadratura de Gauss-Hermite necessita de menos pontos de quadratura que o método da quadratura de Gauss-Hermite. Por outro lado o método adaptativo da quadratura de Gauss-Hermite necessita do cálculo de \hat{z} para cada indivíduo. Neste sentido o método adaptativo da quadratura de Gauss-Hermite é muito mais lento a convergir.

O método adaptativo da quadratura de Gauss-Hermite é o mais adequado para os GLMM e encontra-se implementado em vários programas estatísticos tais como o R (versão 2.12.2) na biblioteca `lme4` e na biblioteca `bind`.

Algumas alternativas têm sido propostas para evitar os problemas de integração numérica tais como os Métodos de Monte Carlo (MC). McCulloch (1997) propôs o método de Monte Carlo Newton-Raphson (MCNR), cujo procedimento incorpora um passo Metropolis-Hastings num algoritmo Monte Carlo EM (MCEM) que calcula e melhora a utilização dos métodos *simulated maximum likelihood* (SML). Booth & Hobert (1999) propuseram duas implementações do algoritmo EM baseadas na máxima verosimilhança para ajustar GLMMs. Uma outra metodologia (Gilks et al., 1996) consiste na utilização de Modelos de Cadeias de Markov via métodos MC (MCMC).

5.5 Inferência

Tendo em atenção que os estimadores do GLMM são obtidos com base no método da máxima verosimilhança, toda a inferência se baseia na função verosimilhança e nas distribuições assintóticas dos estimadores de máxima verosimilhança. Com efeito, assumindo que o modelo é apropriado, os estimadores obtidos têm assintoticamente uma distribuição de Gauss com valor médio igual ao parâmetro a estimar e variância dada pelo inverso da matriz de informação de Fisher.

Assim, testes tipo Wald podem ser usados comparando as estimativas padronizadas com a distribuição $N(0, 1)$. Hipóteses compostas podem ser igualmente testadas usando uma formulação mais geral da estatística de Wald que neste caso é comparada com a distribuição qui-quadrado.

Os testes de razão de verosimilhança são, em geral, os utilizados para comparar a estrutura fixa de dois modelos encaixados que tenham os mesmos efeitos aleatórios. Quando os modelos não estão encaixados usam-se os critérios de informação AIC ou BIC. Quer neste caso, quer quando os modelos estão encaixados, deve-se ter em atenção qual o método utilizado na maximização de (5.3). A função verosimilhança tem de se basear nos dados e não na aproximação dos dados.

Quando o objectivo é testar a presença de efeitos aleatórios, tem-se um problema de fronteira e, tal como foi referido na Secção 2.6.3, a distribuição assintótica sob a hipótese nula é uma mistura de qui-quadrados.

5.6 Análise de Resíduos

A análise de resíduos é feita recorrendo-se a métodos gráficos. Tendo em atenção que se está a trabalhar com dados longitudinais não-Gaussianos esta análise é particularmente complexa tal como se referiu no Capítulo 4.

No caso particular de dados longitudinais binários Gonçalves & Azzalini (2008) apresentam um método que consiste em agregar os resíduos ao longo dos indivíduos para cada instante de observação de forma a aliviar a discretização existente nos dados binários assim como a reduzir o número de gráficos a analisar. Esta metodologia encontra-se implementada na biblioteca `bild` do R e pode ser utilizada quer se tenha considerado o modelo marginal para dados binários com as estruturas de correlação referidas na Secção 4.8, quer se tenha considerado um modelo com um efeito aleatório na intersecção e qualquer uma daquelas estruturas de correlação.

5.7 Modelo Marginal *versus* Modelo com Efeitos Aleatórios

Os modelos marginais e os modelos lineares generalizados mistos são ambos uma extensão do modelo linear generalizado para dados longitudinais não Gaussianos mas possuem características diferentes, no sentido em que cada um possui diferentes especificações e pressupostos em relação à distribuição conjunta, assim como em relação à correlação ao longo das medições repetidas num mesmo indivíduo.

Quando se discutem as duas abordagens tem de se ter em conta que

$$E[g(Y_{it}|\mathbf{b}_i)] \neq g[E(Y_{it}|\mathbf{b}_i)].$$

Considere-se o caso simples de dados binários em que o GLMM tem efeito aleatório na intersecção. Neste caso o valor esperado condicional, $E(Y_{it}|\mathbf{b}_i)$, em função de t é dado por:

$$E(Y_{it}|\mathbf{b}_i) = \frac{\exp [(\beta_0 + b_{0i}) + \beta_1 t]}{1 + \exp [(\beta_0 + b_{0i}) + \beta_1 t]}$$

enquanto que valor esperado marginal $E(Y_{it})$ é dado por:

$$\begin{aligned} E(Y_{it}) &= E[E(Y_{it}|\mathbf{b}_i)] = E \left[\frac{\exp [(\beta_0 + b_{0i}) + \beta_1 t]}{1 + \exp [(\beta_0 + b_{0i}) + \beta_1 t]} \right] \\ &\neq \frac{\exp [\beta_0 + \beta_1 t]}{1 + \exp [\beta_0 + \beta_1 t]}. \end{aligned}$$

A representação gráfica das probabilidades de sucesso nas duas situações é dada na Figura 5.1 onde é evidente a diferença entre os modelos com e sem efeito aleatório.

Designa-se por β^A o vector dos parâmetros fixos no GLMM e por β^M vector dos parâmetros fixos no modelo marginal. O vector β^A modela a evolução de cada indivíduo separadamente enquanto que β^M exprime o modo como, em média, a probabilidade de sucesso evolui na população.

Não existe, em geral, uma relação directa entre os parâmetros do modelo marginal e o do GLMM a não ser em alguns casos especiais (Molenberghs & Verbeke, 2005). Estão nesses casos a regressão logística e a regressão de Poisson com efeito aleatório, b_{0i} , na intersecção e $b_{0i} \sim N(0, d_{11})$.

No caso da regressão logística Zeger et al. (1988) mostrou que

$$\text{logit}[E(Y_{it})] \approx \mathbf{x}_{it}^\top \beta (c^2 d_{11} + 1)^{-1/2},$$

onde $c = \frac{16\sqrt{3}}{15\pi}$.

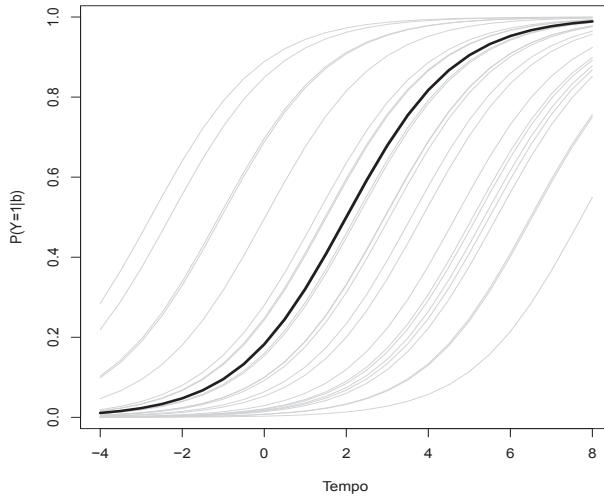


Figura 5.1 Representação gráfica da probabilidade de sucesso num modelo logístico com efeito aleatório na intersecção para diferentes realizações de b e da correspondente probabilidade no modelo marginal (curva a preto).

Para a regressão de Poisson, a relação entre os parâmetros dos dois modelos é dada por:

$$\begin{aligned}
 E(Y_{it}) &= \int \exp(\mathbf{x}_{it}^\top \boldsymbol{\beta}) \exp(b_{0i}) \frac{1}{\sqrt{2\pi d_{11}}} \exp\left(-\frac{b_{0i}^2}{2d_{11}}\right) db_{0i} \\
 &= \exp(\mathbf{x}_{it}^\top \boldsymbol{\beta}) \int \exp(b_{0i}) \frac{1}{\sqrt{2\pi d_{11}}} \exp\left(-\frac{b_{0i}^2}{2d_{11}}\right) db_{0i} \\
 &= \exp\left(\mathbf{x}_{it}^\top \boldsymbol{\beta} + \frac{d_{11}}{2}\right).
 \end{aligned}$$

A interpretação dos parâmetros no GLMM é diferente da interpretação desses parâmetros quando considerado o modelo marginal tal como a Figura 5.1 ilustra.

Os parâmetros de regressão do modelo marginal têm uma interpre-

tação que não depende da associação entre variáveis resposta de um mesmo indivíduo. Tem-se assim, que nos modelos marginais, os parâmetros de regressão descrevem o efeito das covariáveis na variável resposta média da população. Os GLMM têm como principal pressuposto a heterogeneidade de indivíduo para indivíduo. A função de ligação não-linear nos GLMM leva a que os parâmetros de regressão descrevam as alterações da variável resposta de cada indivíduo e a relação destas alterações com as covariáveis.

A escolha entre o modelo marginal e o GLMM para dados longitudinais (não-Gaussianos) só pode ser feita consoante o estudo e as questões a que se pretende dar resposta. Neste sentido, qualquer que seja o estudo longitudinal, a diferentes questões científicas correspondem diferentes modelos.

Por último, refira-se que no GLMM a utilização da máxima verosimilhança leva a que as inferências sejam válidas se o mecanismo de omissão de dados é MCAR ou MAR enquanto que no modelo marginal as inferências com base em estimadores obtidos com o método GEE só são válidas se o mecanismo de omissão de dados é MCAR.

Capítulo 6

Exemplos de Aplicação

6.1 Introdução

Os exemplos do Capítulo 1 vão ser analisados a fim de ilustrar as diferentes metodologias descritas. A selecção dos exemplos teve como objectivo dar a conhecer diferentes domínios de aplicação de estudos longitudinais e permitir ao leitor o fácil acesso aos mesmos assim como a reprodução das análises aqui apresentadas. O programa R é o usado e as bibliotecas são seleccionadas de acordo com o tipo de dados. No caso em que se têm dados longitudinais Gaussianos a biblioteca `nlme` (Pinheiro et al., 2011) é a escolhida quer se opte pelo modelo com efeitos aleatórios ou não. Quando os dados longitudinais são não-Gaussianos e o modelo marginal é considerado são usadas duas bibliotecas. A biblioteca `geepack` (Højsgaard et al., 2011) quando os dados são contagens sendo aplicado o método GEE. A biblioteca `bird` (Gonçalves et al., 2010) quando se têm dados binários sendo neste caso usado o método da máxima verosimilhança. Quando se opta por um modelo com efeitos aleatórios para dados longitudinais não-Gaussianos a biblioteca `bird` volta a ser a seleccionada quando se têm dados longitudinais binários. Se os dados são contagens ao longo do tempo é escolhida a biblioteca `lme4` (Bates et al., 2011). Seguidamente vai-se passar à análise dos dados correspondendo cada

secção a um dos exemplos que passarão a ser referidos pelo nome pelo qual são designados nas bibliotecas do R utilizadas para o efeito.

6.2 BodyWeight

Nesta secção o Exemplo 1.1 (**BodyWeight**) é analisado sendo a variável resposta o peso, e portanto, uma resposta contínua. Por outro lado, o objectivo da experiência é o de se saber se as taxas de crescimento das 3 dietas diferem ao longo do tempo. Isto é, pretende-se inferir para a população mas, como se está a trabalhar com dados longitudinais Gaussianos (pelo menos aproximadamente) a questão entre modelo marginal *versus* modelo com efeitos aleatórios não se coloca. Assim, neste exemplo, dois modelos podem ser ajustados aos dados: (i) o modelo linear misto e (ii) o modelo linear geral para dados longitudinais. Em ambos os casos é usada a biblioteca `nlme` do R.

6.2.1 Modelo Linear Misto

Começa-se pela análise exploratória dos dados que compreende a identificação dos diferentes elementos que constituem os dados, a sua representação gráfica, e a identificação da estrutura fixa do modelo através de um ajustamento não paramétrico. A função `loess` é usada para este ajustamento, sendo o método não paramétrico utilizado o método da regressão polinomial local, robusto face a *outliers* (Cleveland, 1979). A identificação de quais os efeitos aleatórios a considerar no modelo é muitas vezes sugerida nesta parte exploratória. Caso isso não aconteça é feito um ajustamento individual e identificadas quais as estimativas dos parâmetros fixos do modelo que mais variam de indivíduo para indivíduo. A selecção dos parâmetros da parte fixa do modelo é feita com base no teste de razão

de verosimilhanças, se os modelos estão encaixados, ou com base nos critérios AIC ou BIC caso contrário. O método de estimação usado é o método ML. Finda a modelação da parte fixa do modelo, este é re-ajustado usando-se o método REML a fim de se obterem estimadores não enviesados das componentes da variância e passa-se à selecção dos efeitos aleatórios assim como da modelação da matriz \mathbf{D} . A vali-

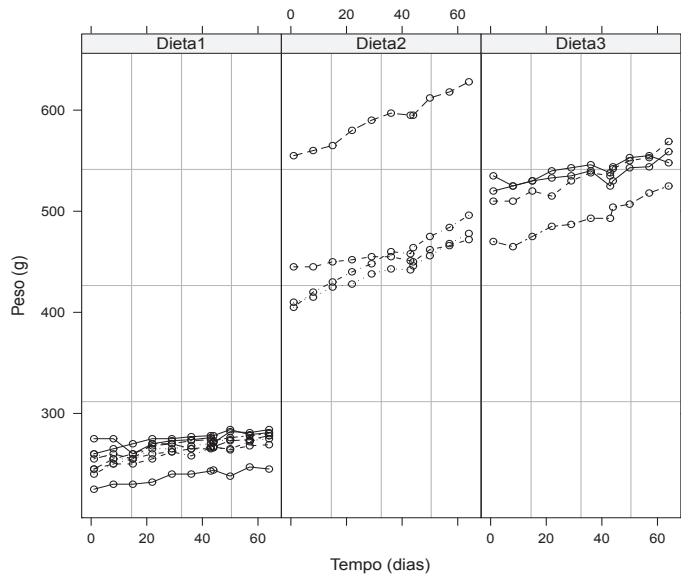


Figura 6.1 Gráfico dos perfis individuais, por dieta.

dação do modelo é feita tal como é descrita na Secção 2.8. No caso de existir heterocedasticidade e/ou correlação entre os erros aleatórios procede-se à modelação matriz de variâncias-covariâncias dos erros aleatórios, de acordo com os métodos descritos no Capítulo 3, e posterior validação do novo modelo. No caso do **BodyWeight** têm-se 11 observações por rato e 16 por cada instante de tempo num total de 176 observações. Os 16 ratos estão divididos por 3 dietas. A Dieta 1

116 Exemplos de Aplicação

com 8 ratos e as restantes dietas (Dieta 2 e Dieta 3) com 4 ratos cada. O peso dos ratos (em gramas) foi medido de sete em sete dias a partir do dia 1 até ao dia 64, com uma medição extra no dia 44 (os instantes de observação não estão igualmente espaçados). Cada grupo (rato) é caracterizado pela respectiva dieta. Na Figura 6.1 está representada a evolução do peso dos ratos ao longo do tempo por dieta.

Para cada dieta é feito o ajustamento não-paramétrico e a análise desse ajustamento sugere uma relação linear entre o peso e o tempo (Figura 6.2).

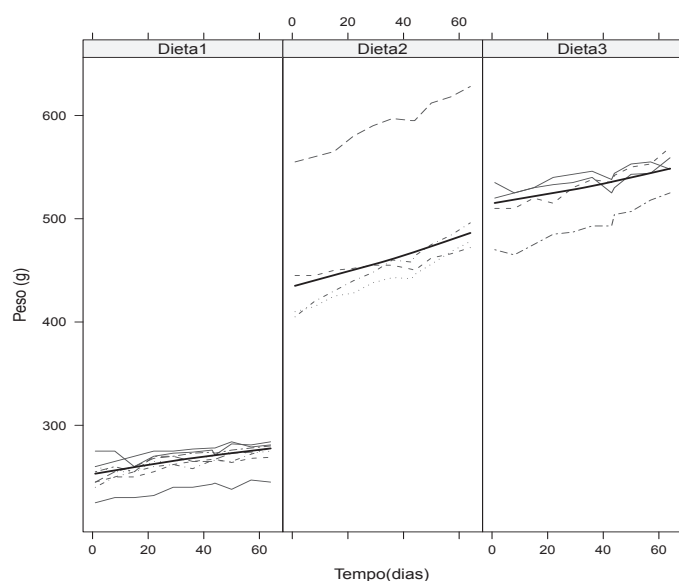


Figura 6.2 *Ajustamento não paramétrico (a negrito), por dieta.*

Quanto aos efeitos aleatórios a considerar faz-se um ajustamento linear individual a analisa-se quais os parâmetros que mais variam de rato para rato. Essa análise é feita com base no gráfico da estimativa dos intervalos de confiança para os parâmetros do modelo ajustado a

cada rato (Figura 6.3).

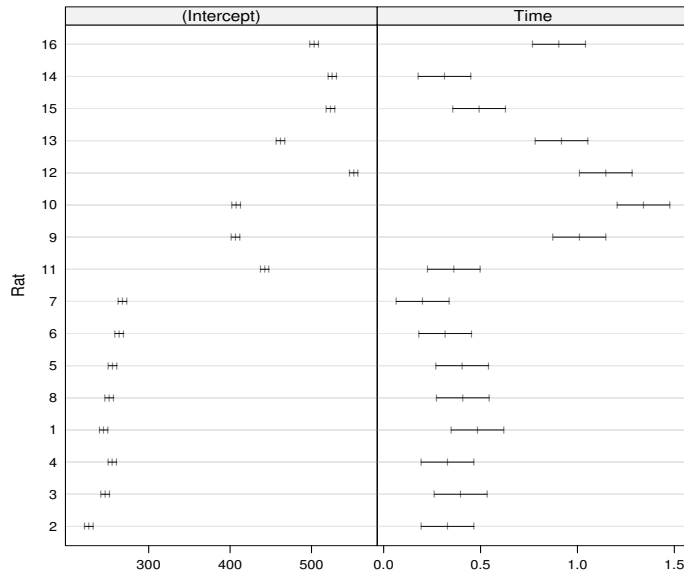


Figura 6.3 Estimativa dos intervalos de confiança para os parâmetros do ajustamento individual.

Como se pode verificar há grande variabilidade nas estimativas quer da intersecção quer do declive. Vão-se então considerar dois modelos ambos com dois efeitos aleatórios, um na intersecção e outro no declive:

(i) *Modelo I*

$$Y_{it} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})Tempo_{it} + \beta_2Dieta_{2i} + \beta_3Dieta_{3i} + \epsilon_{it}$$

(ii) *Modelo II*

$$Y_{it} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})Tempo_{it} + \beta_2Dieta_{2i} + \beta_3Dieta_{3i} + \beta_4(Tempo_{it} \times Dieta_{2i}) + \beta_5(Tempo_{it} \times Dieta_{3i}) + \epsilon_{it}$$

com $i = 1, \dots, 16$, $t = 1, \dots, 11$ e $Tempo_{it} = 1, 8, 15, 22, 29, 36, 43, 44, 50, 57, 64$ e onde Y_{it} é a variável variável aleatória associada ao peso, em $Tempo_{it}$, do i -ésimo rato, $Dieta_{2i}$ é uma variável binária que assume o valor 1 se o i -ésimo rato recebe a Dieta 2; $Dieta_{3i}$ é uma variável binária que assume o valor 1 se o i -ésimo rato recebe a Dieta 3. A Dieta 1 é a dieta de referência, β_0 e β_1 são, respectivamente, o *intercept* e o declive médios para os ratos que recebem a Dieta 1; β_2 é a diferença média entre os *intercepts* dos ratos que receberam a Dieta 2 e os que receberam a Dieta 1, β_3 é a diferença média entre os *intercepts* dos ratos que receberam a Dieta 3 e os que receberam a Dieta 1, β_4 é a diferença média entre os declives dos ratos que receberam a Dieta 2 e os que receberam a Dieta 1, β_5 é a diferença média entre os declives dos ratos que receberam a Dieta 3 e os que receberam a Dieta 1. $\mathbf{b}_i = (b_{0i}, b_{1i})^\top$ é o vector dos efeitos aleatórios e ϵ_{it} é o erro aleatório cujas hipóteses subjacentes são as dadas no Capítulo 2. Os dois modelos foram ajustados aos dados usando o método ML. Da comparação dos dois modelos, usando o teste de razão de verosimilhanças, o *Modelo I* foi rejeitado (valor-p= 0.0021). As estimativas ML dos parâmetros para o modelo seleccionado encontram-se na Tabela 6.1 assim como os erros padrão (SE), graus de liberdade (GL), valores da estatística de teste (valor-T) e respectivos valores-p.

Na Tabela 6.2 é dada a tabela da ANOVA para o *Modelo II* onde GLnum e GLden se referem aos graus de liberdade do numerador e do denominador, respectivamente.

Passa-se então ao reajustamento do modelo seleccionado, usando o método REML, e testa-se se é possível só haver um efeito aleatório no *intercept* (*Modelo III*) ou só um no declive (*Modelo IV*). Para comparar estes modelos com o *Modelo II* é usado o teste de razão de verosimilhanças e, em ambos os casos, os valores-p são inferiores a 0.0001 pelo que se continua com o *Modelo II*. Para estes dois tes-

Tabela 6.1 Estimativas ML, erros padrão, graus de liberdade, valores- T e valores- p para o Modelo II.

Parâmetro	Estimativa	SE	GL	valor - T	valor - p
β_0	251.652	12.009	157	20.955	0.0000
β_1	0.360	0.084	157	4.302	0.0000
β_2	200.666	20.801	13	9.647	0.0000
β_3	252.072	20.801	13	12.118	0.0000
β_4	0.606	0.145	157	4.185	0.0000
β_5	0.298	0.145	157	2.061	0.0410

Tabela 6.2 Tabela da ANOVA para o Modelo II.

Parâmetro	GLnum	GLden	valor - T	valor - p
<i>Intercept</i>	1	157	2036.631	< 0.0001
<i>Tempo</i>	1	157	98.187	< 0.0001
<i>Dieta</i>	2	13	101.466	< 0.0001
<i>Tempo × Dieta</i>	2	157	9.004	$2e - 04$

tes o valor- p obtido no *output* do R é o valor *naive* (Secção 2.6.3) em virtude de se estar perante um problema de fronteira, sendo a distribuição aproximada da estatística de teste, em qualquer um dos casos, uma mistura $0.5\chi_1^2 + 0.5\chi_2^2$ e o verdadeiro valor- p inferior ao indicado. Tendo em atenção que estimativa do intervalo de confiança a 95% para a correlação entre os efeitos aleatórios é $(-0.6121, 0.3901)$ vai-se testar se podem ser considerados independentes, isto é, se a matriz \mathbf{D} tem estrutura diagonal (*Modelo V*). O teste de razão de verosimilhanças é usado e o valor- p é de 0.6027, pelo que não se rejeita esta estrutura para a matriz \mathbf{D} . Na Tabela 6.3 são dados os valores do AIC e do BIC assim como o valor do teste de razão de verosimilhanças e o respectivo valor- p para as comparações acabadas

Tabela 6.3 Valores de AIC, BIC, logLIK, TRV e valores-p.

Modelos	AIC	BIC	logLIK	Teste	TRV	valor - p
<i>Modelo II</i>	1171.720	1203.078	-575.860			
<i>Modelo III</i>	1248.245	1273.332	-616.123	<i>III vs II</i>	80.525	< 0.0001
<i>Modelo IV</i>	1573.269	1598.355	-778.635	<i>IV vs II</i>	405.549	< 0.0001
<i>Modelo V</i>	1169.991	1198.213	-575.995	<i>V vs II</i>	0.271	0.6027

de referir.

O gráfico dos resíduos padronizados *versus* os valores ajustados para o *Modelo V* é dado na Figura 6.4 e evidência um aumento da variabilidade com os valores ajustados. A mesma representação gráfica, mas por dieta, é dada na Figura 6.5 e sugere uma eventual heterocedasticidade entre as Dietas.

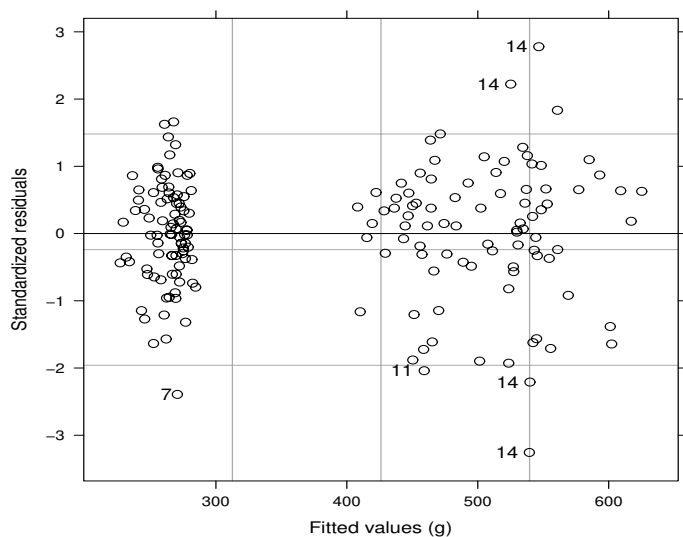


Figura 6.4 Resíduos padronizados *versus* valores ajustados para o Modelo V.

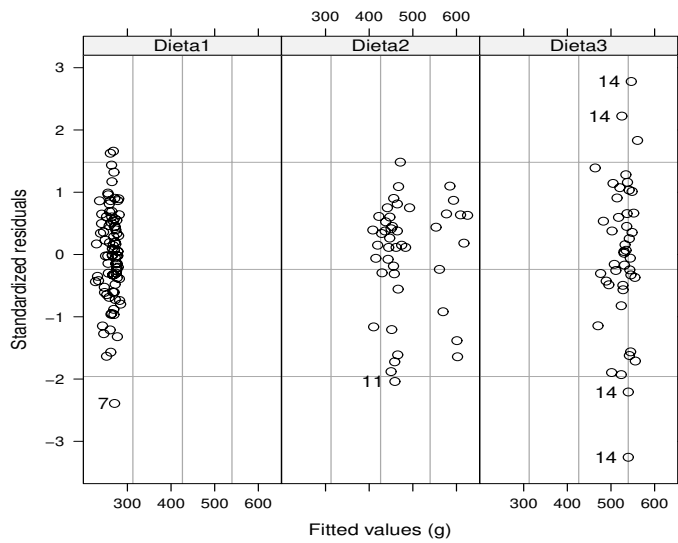


Figura 6.5 Resíduos padronizados versus valores ajustados por dieta para o Modelo V.

Vai-se modelar esta heterocedasticidade através da *função de variância potência de uma covariável*, $Var(\epsilon_{it}) = \sigma^2 |v_{it}|^{2\delta}$ (*Modelo VI*), sendo igualmente plausível a *função variâncias diferentes por estrato* $Var(\epsilon_{it}) = \sigma^2 \delta_{S_{it}}^2$ (*Modelo VII*). Com base no teste de razão de verossimilhanças compara-se cada um destes modelos com o *Modelo V*. Os resultados estão apresentados na Tabela 6.4, onde se verifica que os respectivos valores-p indicam forte evidência de heterocedasticidade pelo que o *Modelo V* é rejeitado em cada um dos testes.

Segue-se a comparação entre o *Modelo VI* e o *Modelo VII* feita com base no critério AIC em virtude destes dois modelos não estarem encaixados. Dessa comparação é seleccionado o *Modelo VI* em virtude de ser o que apresenta o menor valor do AIC (Tabela 6.4).

A estimativa de δ é de 0.5435, isto é, a variância cresce aproximadamente com a raiz quadrada dos valores ajustados. Para o novo

Tabela 6.4 Valores de AIC, BIC, logLik, TRV e valores-p.

Modelos	AIC	BIC	logLik	Teste	TRV	valor - p
Modelo V	1169.991	1198.213	-575.995			
Modelo VI	1162.172	1193.530	-571.086	V vs VI	9.812	< 0.0017
Modelo VII	1162.194	1196.688	-570.097	V vs VII	11.797	0.0027

modelo (Modelo VI) o gráfico dos resíduos padronizados *versus* valores ajustados é dado na Figura 6.6 onde é patente um padrão de homocedasticidade com muito poucos *outliers*. O mesmo acontece quando se considera o gráfico dos resíduos padronizados *versus* valores ajustados por dieta (Figura 6.7).

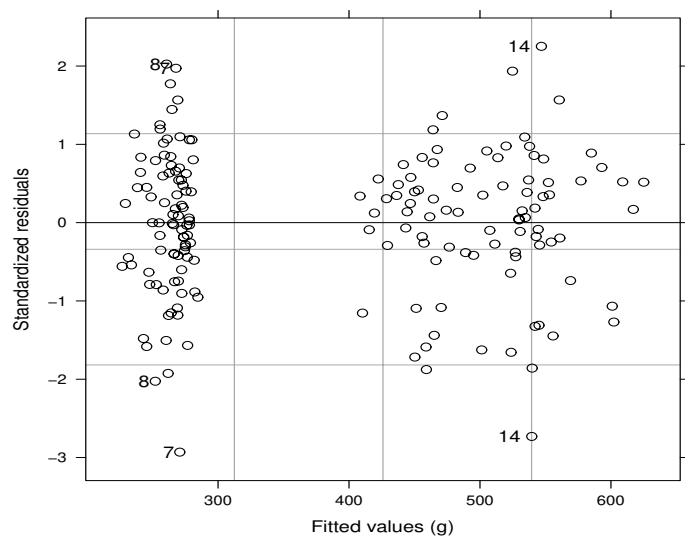


Figura 6.6 Resíduos padronizados *versus* valores ajustados para o Modelo VI.

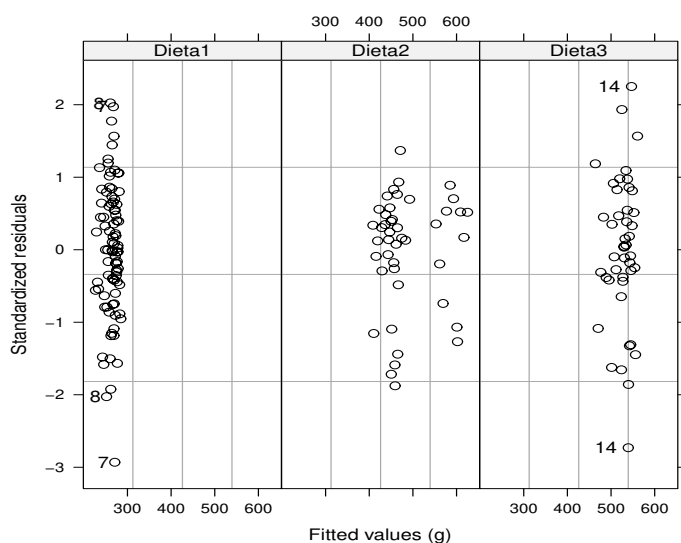


Figura 6.7 Resíduos padronizados versus valores ajustados por dieta para o Modelo VI.

Para analisar a correlação entre as observações de um mesmo rato usa-se o variograma em virtude dos dados terem sido obtidos os instantes de tempo não igualmente espaçados.

O gráfico do semivariograma amostral dos resíduos padronizados do *Modelo VI* está representado na Figura 6.8. A curva não paramétrica ajustada ao semivariograma permite visualizar o padrão existente concluindo-se que cresce até aos 20 dias após o que estabiliza. Este tipo de comportamento sugere uma correlação espacial exponencial. É então ajustado o *Modelo VIII* com esta estrutura de correlação e, com base no teste de razão de verosimilhança (não há correlação *versus* há correlação espacial exponencial) conclui-se (valor- $p < 0.001$) que a correlação existente é modelada através daquele modelo de correlação.

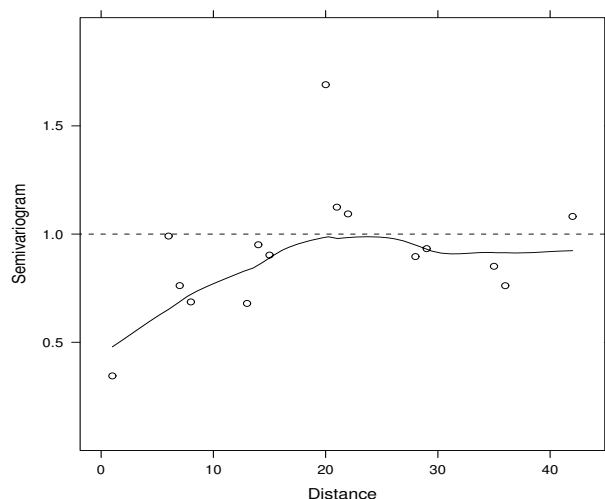


Figura 6.8 Semivariograma amostral dos resíduos padronizados do Modelo VI.

Na Figura 6.9 está representado o gráfico do semivariograma amostral dos resíduos normalizados para o novo modelo (*Modelo VIII*), verificando-se que o padrão agora apresentado pela curva não paramétrica corresponde a um ruído branco.

O passo seguinte é ver se este modelo verifica os pressupostos requeridos. O semivariograma é o indicado na Figura 6.9.

Nas Figuras 6.10 e 6.11 estão dados, respectivamente, o gráfico dos resíduos normalizados *versus* valores ajustados e o papel de probabilidades para os resíduos normalizados. Da análise das figuras conclui-se que as hipóteses subjacentes são verificadas.

As estimativas dos parâmetros para o modelo seleccionado (*Modelo VIII*) encontram-se na Tabela 6.5 assim como os erros padrão, valor da estatística de teste-t (valor-T) e respectivos valores-p.

A tabela da ANOVA para o *Modelo VIII* é dada na Tabela 6.6.

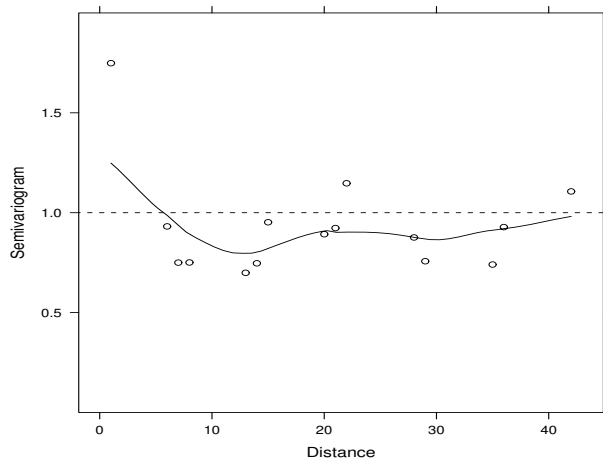


Figura 6.9 Semivariograma amostral dos resíduos normalizados do modelo de correlação espacial exponencial (Modelo VIII).

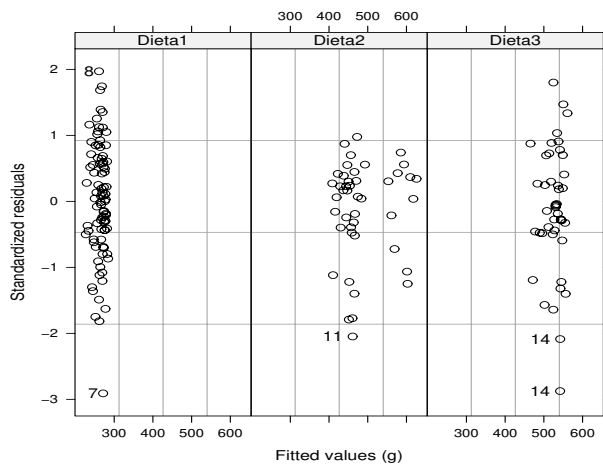


Figura 6.10 Resíduos padronizados versus valores ajustados por dieta (Modelo VIII).

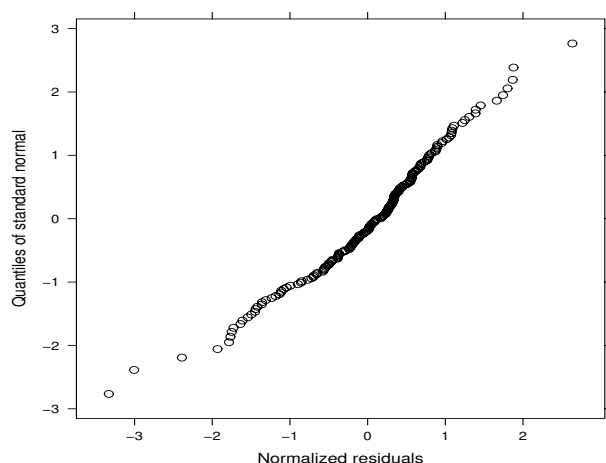


Figura 6.11 *Papel de probabilidades dos resíduos normalizados (Modelo VIII).*

Tabela 6.5 *Estimativas ML, erros padrão, graus de liberdade, valores-T e valores-p para o Modelo VIII.*

Parâmetro	Estimativa	SE	GL	valor - T	valor - p
β_0	251.485	13.023	157	19.311	0.0000
β_1	0.363	0.086	157	4.238	0.0000
β_2	200.749	22.597	13	8.884	0.0000
β_3	252.585	22.607	13	11.173	0.0000
β_4	0.625	0.153	157	4.101	0.0001
β_5	0.307	0.153	157	2.000	0.0472

Pode-se finalmente dizer que o crescimento dos ratos é uma função linear do tempo sendo o aumento verificado na taxa de crescimento da Dieta 2 em relação à da Dieta 1 estatisticamente significativo (valor-p=0.0001). Quanto ao aumento verificado na taxa de crescimento da Dieta 3 em relação à Dieta 1 (valor-p=0.0472) só se pode dizer que a

Tabela 6.6 Tabela da ANOVA para o Modelo VIII.

Parâmetro	GLnum	GLden	valor - T	valor - p
<i>Intercept</i>	1	157	1585.774	< 0.0001
<i>Tempo</i>	1	157	91.592	< 0.0001
<i>Dieta</i>	2	13	78.612	< 0.0001
<i>Tempo × Dieta</i>	2	157	8.691	3e - 04

significância estatística é *borderline*. A diferença entre a taxa de crescimento da Dieta 2 e da Dieta 3 não é estatisticamente significativa (valor-p= 0.0775). Tem-se ainda que a estimativa da matriz

$$\mathbf{D} = \begin{bmatrix} \sigma_0^2 & \sigma_{01}^2 \\ \sigma_{10}^2 & \sigma_1^2 \end{bmatrix}$$

é

$$\begin{bmatrix} (36.74)^2 & 0 \\ 0 & (0.232)^2 \end{bmatrix}.$$

Relativamente à matriz dos erros aleatórios, $\mathbf{\Lambda}_i \sigma^2$, a estimativa de σ^2 é de 0.0186 e $\mathbf{\Lambda}_i = \mathbf{W}_i \mathbf{C}_i \mathbf{W}_i$ tem \mathbf{W}_i modelada por uma *função de variância potência de uma covariável*, $Var(\epsilon_{it}) = \sigma^2 |v_{it}|^{2\delta}$ onde a estimativa de δ é 0.5953 e a estrutura de \mathbf{C}_i é explicada por um modelo exponencial com $\rho = 4.8931$. Nas Figuras 6.12 e 6.13 estão representados, respectivamente, os gráficos dos ajustamentos individual e por tratamento com base no *Modelo VIII*.

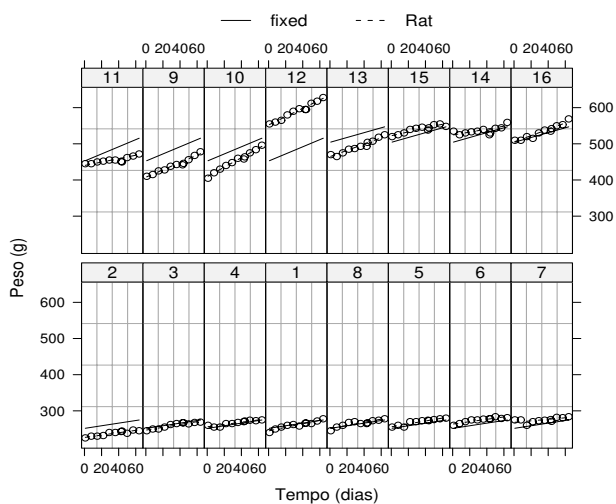


Figura 6.12 Previsões a partir do Modelo VIII.

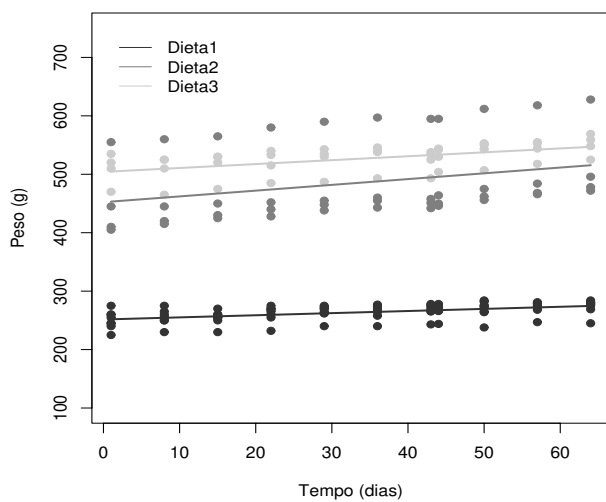


Figura 6.13 Estimativa do peso médio por dieta (Modelo VIII).

6.2.2 Modelo Linear Geral para Dados Longitudinais

Considere-se agora o ajustamento um modelo sem efeitos aleatórios aos dados **BodyWeight**. O método GLS é o escolhido e a metodologia seguida no ajustamento deste modelo é a mesma utilizada na secção anterior. Começa-se por ajustar o *Modelo IX*

$$Y_{it} = \beta_0 + \beta_1 \text{Tempo}_{it} + \beta_2 \text{Dieta}_{2i} + \beta_3 \text{Dieta}_{3i} \\ + \beta_4 (\text{Tempo}_{it} \times \text{Dieta}_{2i}) + \beta_5 (\text{Tempo}_{it} \times \text{Dieta}_{3i}) + \epsilon_{it}$$

seguinte-se os passos do ajustamento anterior para a modelação da estrutura de variâncias-covariâncias dos erros aleatórios. Neste caso o modelo de heterocedasticidade escolhido foi a *função variâncias diferentes por estrato* continuando a correlação a ser modelada pela estrutura de correlação espacial. As estimativas dos parâmetros para este modelo (*Modelo X*) encontram-se na Tabela 6.7 assim como os erros padrão.

Da análise da Tabela 6.7 verifica-se que as estimativas dos parâmetros fixos são semelhantes às do *Modelo VIII* ressaltando o facto de no modelo sem efeitos aleatórios o aumento verificado na taxa de crescimento da Dieta 3 em relação à Dieta 1 ($p=0.1828$) não ser estatisticamente significativo enquanto que no modelo com efeitos aleatórios essa significância estatística é *borderline*.

A comparação entre este modelo (*Modelo X*) e o melhor modelo obtido no parágrafo anterior (*Modelo VIII*) faz-se com base nos critérios de informação uma vez que os modelos não se encontram encaixados mas têm a mesma estrutura fixa. Para o *Modelo X* os valores do AIC e do BIC são, respectivamente, 1146.683 e 1178.041 enquanto que para o *Modelo VIII* são 1143.377 e 1177.871. O modelo com efeitos aleatórios é o que apresenta menores valores do AIC e do BIC e como tal parece dar uma melhor representação dos dados. Mas tal

Tabela 6.7 Estimativas ML, erros padrão, valores- T e valores- p para o Modelo X .

Parâmetro	Estimativa	SE	valor - T	valor - p
β_0	250.338	9.659	25.918	0.0000
β_1	0.3671	0.0761	4.831	0.0000
β_2	202.263	21.4174	9.442	0.0000
β_3	257.487	27.721	9.289	0.0001
β_4	0.661	0.168	3.921	0.0000
β_5	0.292	0.218	1.338	0.1828

como Pinheiro & Bates (2000) fazem notar a escolha entre um modelo com efeitos aleatórios e um modelo sem efeitos aleatórios deve ter em conta mais do que os critérios de informação e o teste de razão de verossimilhanças (Secção 3.7.1).

6.3 Seizure

Neste exemplo a variável resposta é o número de episódios de epilepsia e o objectivo do ensaio clínico é o de saber se o tratamento com *prograbide* reduz ou não a taxa de episódios de epilepsia quando comparado com o *placebo*. Para responder a esta questão vai-se comparar: (i) As alterações da *baseline* para o *follow-up*, na taxa de episódios de epilepsia nos dois grupos de tratamento. Neste caso o modelo marginal é considerado e é usado o método GEE. (ii) As alterações específicas dos doentes da *baseline* para o *follow-up*, relativamente à taxa de episódios de epilepsia nos dois grupos de tratamento, sendo neste caso usado o GLMM. A utilização destas duas abordagens tem como objectivo ilustrar as mesmas e, ao mesmo tempo, mostrar a diferença que existe entre elas no que diz respeito à interpretação dos

parâmetros. A representação gráfica do perfil de cada doente é dada na Figura 6.12 onde se vê que os dados apresentam um valor extremo ou *outlier* correspondente ao doente 49 (Thall & Vail, 1990). Este doente apresenta 151 episódios ao longo das 8 semanas que constituem o intervalo de tempo da *baseline* e 302 ($102 + 65 + 72 + 63$) episódios durante os quatro sucessivos períodos de 2 semanas cada. Este doente foi atribuído ao grupo com o tratamento *prograbide* e como pode produzir impacto nas análises considera-se, em ambas as abordagens, a análise dos dados com e sem o indivíduo 49.

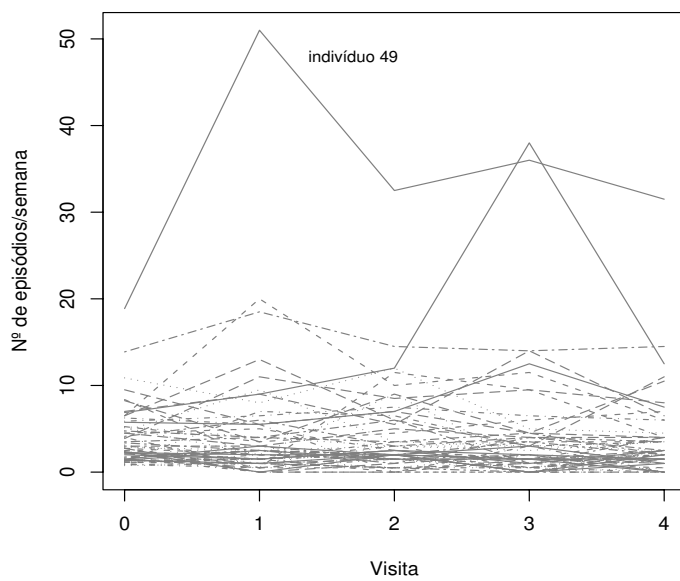


Figura 6.14 Gráfico dos perfis individuais da taxa semanal de episódios de epilepsia para os 59 doentes.

6.3.1 Modelo Marginal. Método GEE

A aplicação do método GEE para a estimação dos parâmetros pressupõe a utilização de procedimentos que são comuns, quer se esteja a modelar contagens ou dados binários. Começa-se por uma análise exploratória dos dados que inclui, quando necessário, um ajustamento não paramétrico tal como foi descrito na Secção 6.2.1 que permite estabelecer o tipo de relação existente entre a variável resposta e o tempo. A estrutura de correlação a utilizar é igualmente sugerida nesta fase de análise com base na matriz de correlação dos resíduos de Pearson resultante do ajustamento do modelo linear generalizado apropriado ignorando a dependência entre as observações de um mesmo indivíduo. O teste de Wald é utilizado para a comparação de modelos encaixados usados na modelação do valor esperado. É utilizada a biblioteca `geepack` do R.

Os dados do exemplo **Seizure** apresentam variabilidade superior à predita pelo valor esperado do modelo de Poisson pelo que o modelo marginal considerado é um modelo de regressão log-linear, em que se tem como pressuposto uma função de variância de Poisson, $Var(Y_{it}) = \phi\mu_{it}$, onde se assume $\phi > 1$.

A equação do modelo marginal para o número esperado de episódios de epilepsia por semana (*baseline* e *follow-up*) é dada por

$$\begin{aligned} \log[E(Y_{it})] = \log(\mu_{it}) = & \log(t_{it}) + \beta_0 + \beta_1 Tempo_{it} + \beta_2 Trat_i \\ & + \beta_3(Tempo_{it} \times Trat_i) \end{aligned}$$

com $t = 0, 1, 2, 3, 4$ e $i = 1, \dots, 59$ e onde Y_{it} é a variável aleatória de Poisson que dá o número de episódios de epilepsia para o i -ésimo doente no t -ésimo período de observação e t_{it} é o comprimento do período t (onde $t_{it} = 8$ se $t = 0$ e $t_{it} = 2$ se $t = 1, 2, 3, 4$).

O termo $\log(t_{it})$ é incluído no modelo para englobar diferentes pe-

Tabela 6.8 *Matriz de correlação para as observações.*

	<i>baseline</i>	<i>visita1</i>	<i>visita2</i>	<i>visita3</i>	<i>visita4</i>
<i>baseline</i>	1.000	0.792	0.828	0.676	0.840
<i>visita1</i>	0.792	1.000	0.872	0.745	0.897
<i>visita2</i>	0.828	0.872	1.000	0.745	0.897
<i>visita3</i>	0.676	0.745	0.807	1.000	0.829
<i>visita4</i>	0.84	0.897	0.897	0.829	1.000

ríodos de observações e entra no modelo como *offset* (Diggle et al., 2002; Fitzmaurice et al., 2004). A variável binária $Tempo_{it}$ corresponde aos períodos *baseline* e *follow-up*, assumindo o valor 0 se $t = 0$ e o valor 1 se $t = 1, 2, 3, 4$. $Trat$ é uma variável indicatriz que assume o valor 0 se ao indivíduo i foi atribuído o *placebo* e o valor 1 se foi atribuído ao *prograbide*.

Como $Tempo_{it}$ só assume dois valores não é necessário fazer-se um ajustamento não paramétrico para saber qual a relação entre a variável resposta e o tempo. Para obter a informação sobre a estrutura de correlação a usar no método GEE procede-se tal como o descrito anteriormente, isto é, ignora-se a dependência entre observações de um mesmo doente e ajusta-se um modelo de regressão de Poisson. Com base neste ajustamento calcula-se a matriz de correlação dos resíduos de Pearson para cada doente, que se apresenta na Tabela 6.8, assim como a estimativa do parâmetro de dispersão ϕ (19.4129).

Da análise da tabela é escolhida a estrutura $corr(Y_{it}, Y_{it'}) = \alpha$. Passa-se ao ajustamento daquele modelo usando o método GEE com aquela estrutura de correlação. O grupo do *placebo* é a classe de referência considerada e as estimativas dos erros padronizados, para os estimadores dos parâmetros de regressão, são obtidas através do estimador

Tabela 6.9 Estimativas dos parâmetros, erros padrão, valores- T e valores- p .

Parâmetro	Estimativa	SE	valor - T	valor - p
β_0	1.348	0.157	73.342	0.0000
β_1	0.112	0.116	0.931	0.3347
β_2	0.028	0.225	0.015	0.9024
β_3	-0.105	0.219	0.230	0.6317
ϕ	19.413	8.792	4.876	0.0272
α	0.777	0.080	94.654	0.0000

sandwich. O método *jackknife* foi igualmente utilizado não se tendo verificado qualquer alteração dos resultados. As estimativas dos parâmetros para o modelo encontram-se na Tabela 6.9 assim como os erros padrão (*sandwich*), valor da estatística do teste de Wald (valor- T) e respectivos valores - p . Na Tabela 6.10 encontra-se o sumário da interpretação dos parâmetros do modelo em termos do logaritmo do número esperado de episódios de epilepsia por semana, isto é, do logaritmo da taxa esperada de episódios (vai-se omitir a palavra “esperada” sem perigo de confusão). Por exemplo, a taxa de episódios na *baseline* ($t = 0$) no grupo do *placebo* é de e^{β_0} , já no *follow-up* ($t = 1$) é de $e^{\beta_0+\beta_1}$. Assim, e^{β_1} é o quociente da taxa de episódios entre o *follow-up* e a *baseline* no grupo do *placebo*. Do mesmo modo, no grupo do *prograbide*, $e^{\beta_0+\beta_2}$ é a taxa de episódios na *baseline* e $e^{\beta_0+\beta_1+\beta_2+\beta_3}$ é a taxa de episódios no *follow-up* e, como tal, $e^{\beta_1+\beta_3}$ é o quociente da taxa de episódios entre o *follow-up* e a *baseline*. Uma comparação directa dos dois tratamentos em termos da alteração da taxa de episódios é expressa em termos de β_3 , pelo que β_3 é o parâmetro de interesse. O parâmetro β_3 representa a diferença dos logaritmos das alterações das taxas comparando os doentes do grupo

Tabela 6.10 *Parâmetros do modelo de Poisson marginal.*

Grupo	Tempo	$\log(\mu_{it}) = \log \left[E \left(\frac{Y_{it}}{t_{it}} \right) \right]$
Placebo	<i>Baseline</i> ($t = 0$)	β_0
	<i>Follow-up</i> ($t = 1$)	$\beta_0 + \beta_1$
Prograbide	<i>Baseline</i> ($t = 0$)	$\beta_0 + \beta_2$
	<i>Follow-up</i> ($t = 1$)	$\beta_0 + \beta_1 + \beta_2 + \beta_3$

prograbide com os do grupo *placebo*. Um valor negativo deste parâmetro significa uma maior redução da taxa de episódios no grupo do *prograbide*. No presente caso (Tabela 6.9) a estimativa desse parâmetro é de -0.1047 não sendo a diferença entre os grupos do *placebo* e do *prograbide* estatisticamente significativa (valor-p=0.6317) se o doente 49 é considerado na análise. Por último faz-se notar que o valor de 0.78 para a estimativa do parâmetro α (Tabela 6.9) indica a existência de correlação significativa (valor-p=0.000) entre observações de um mesmo indivíduo e que o valor aproximado de 19.4129 para a estimativa do parâmetro de dispersão revela sobredispersão, relativamente à variabilidade de Poisson predita nestes dados.

Retirando o doente 49 e ajustando o mesmo modelo obtém-se as estimativas indicados na Tabela 6.11. A estimativa do parâmetro β_3 passa a ser de -0.3024 indicando, embora de forma modesta (valor-p = 0.0827), que o tratamento com *prograbide*, comparado com o *placebo*, reduz a taxa de episódios.

No que se refere à redução esperada da taxa de episódios esta é de aproximadamente 17% no grupo do *prograbide* enquanto que no grupo do *placebo* não se espera verificar qualquer alteração ($e^{0.112} \approx 1$).

Tabela 6.11 Estimativas dos parâmetros, erros padrão, valores- T e valores- p (sem o doente 49).

Parâmetro	Estimativa	SE	valor - T	valor - p
β_0	1.348	0.157	73.342	0.0000
β_1	0.112	0.116	0.931	0.3347
β_2	0.107	0.195	3.000	0.5839
β_3	-0.302	0.174	3.011	0.0827
ϕ	10.385	2.318	20.074	$7.45e - 06$
α	0.598	0.084	51.065	0.0000

6.3.2 Modelo Linear Generalizado Misto

A utilização do modelo linear generalizado misto significa que se responde à questão colocada no que se refere às alterações específicas dos doentes relativamente à taxa de episódios de epilepsia, da *baseline* para o *follow-up* nos dois grupos de tratamento. É utilizada a biblioteca `lme4` do R onde é usado o método adaptativo da quadratura de *Gauss-Hermite*. O modelo com efeitos aleatórios na intersecção e no declive é considerado sendo a parte fixa do modelo a dada na secção anterior.

A equação do modelo é então

$$\begin{aligned} \log[E(Y_{it}|\mathbf{b}_i)] &= \log(\mu_{it}^b) = \log(t_{it}) + \beta_0 + \beta_1 \text{Tempo}_{it} + \beta_2 \text{Trat}_i \\ &\quad + \beta_3 (\text{Tempo}_{it} \times \text{Trat}_i) + b_{0i} + b_{1i} \text{Tempo}_{it} \end{aligned}$$

com $t = 0, 1, 2, 3, 4$; $i = 1, \dots, 59$ e $\mathbf{b}_i = (b_{0i}, b_{1i})^T$ o vector de efeitos aleatórios com distribuição Gaussiana multivariada com valor esperado nulo e matriz de variâncias-covariâncias \mathbf{D} . Dado \mathbf{b}_i , assume-se que os Y_{it} são independentes com distribuição de Poisson e $\text{var}(Y_{it}|\mathbf{b}_i) = E(Y_{it}|\mathbf{b}_i)$, (isto é, $\phi = 1$). Neste modelo assume-se

Tabela 6.12 *Parâmetros do modelo de Poisson marginal.*

Grupo	Tempo	$\log(\mu_{it}) = \log \left[\frac{E(Y_{it} \mathbf{b}_i)}{t_{it}} \right]$
Placebo	<i>Baseline</i> ($t = 0$)	$(\beta_0 + b_{0i})$
	<i>Follow-up</i> ($t = 1$)	$(\beta_0 + b_{0i}) + \beta_1 + b_{1i}$
Prograbide	<i>Baseline</i> ($t = 0$)	$(\beta_0 + b_{0i}) + \beta_2$
	<i>Follow-up</i> ($t = 1$)	$(\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) + \beta_2 + \beta_3$

que existe heterogeneidade entre os doentes no que diz respeito ao número de episódios de epilepsia não só na *baseline* mas também na alteração desse número ao longo do tempo. Na Tabela 6.12 encontra-se o sumário da interpretação dos parâmetros do modelo em termos do logaritmo do número de episódios de epilepsia por semana. Neste caso, e^{β_1} é o quociente da taxa de episódios entre o *follow-up* e a *baseline* para um doente no grupo *placebo* com um declive aleatório não observado, $b_{1i} = 0$, (o valor médio e a mediana de b_{1i}). Quanto a $e^{\beta_1 + \beta_3}$, é o quociente da taxa de episódios entre o *follow-up* e a *baseline* para um doente no grupo do *prograbide* com um declive aleatório não observado $b_{1i} = 0$. Uma comparação directa entre os dois tratamentos em termos da alteração do número de episódios de epilepsia é expressa em termos de β_3 quando dois doentes são escolhidos, um de cada grupo, de modo a terem o mesmo valor não observado de b_{1i} . As estimativas dos parâmetros para o modelo seleccionado, considerando o doente 49, encontram-se na Tabela 6.13 assim como os erros padrão, valor da estatística de teste e respectivos valores-p.

A hipótese $H_0 : \beta_3 = 0$ é rejeitada ao nível de significância 5% (valor-p=0.0405) e como tal pode-se afirmar que há diferença entre os dois tratamentos em termos das alterações específicas dos doentes na taxa de episódios para aquele nível de significância. Para um doente no

Tabela 6.13 *Estimativas dos parâmetros, erros padrão, valores-T e valores-p.*

Parâmetro	Estimativa	SE	valor - T	valor - p
β_0	1.068	0.139	7.663	$1.82e - 14$
β_1	0.004	0.107	0.035	0.9721
β_2	0.053	0.192	0.273	0.7848
β_3	-0.306	0.149	-2.048	0.0405
d_{11}	0.496			
d_{22}	0.228			
d_{12}	0.025			

grupo *placebo* não há alteração da taxa de episódios entre o *follow-up* e a *baseline* ($e^{0.004} \approx 1$) mas, para um doente no grupo *progabide* já se espera uma redução de 26% ($e^{0.004-0.306} \approx 0.74$) no número de episódios de epilepsia.

A estimativa da variância de b_{0i} , d_{11} , é de 0.4962, o que implica grande variabilidade entre os doentes no que se refere à taxa de episódios na *baseline*. Por exemplo, aproximadamente, 95% dos doentes no grupo do *placebo* têm uma taxa de episódios na *baseline* que varia de $\exp(1.0684 - 1.96 \times \sqrt{0.4962})$ a $\exp(1.0684 + 1.96 \times \sqrt{0.4962})$, ou seja, entre 0.73 e 12 episódios de epilepsia por semana. Do mesmo modo a estimativa da variância de b_{1i} , d_{22} , é de 0.2279 o que implica grande variabilidade na alteração da taxa de episódios entre os doentes. Por exemplo, para os doentes tratados com o *progabide* a variação na taxa de episódios vai de uma redução de 71% a um aumento de 88% com aproximadamente 95% de confiança ($\exp[(0.0038 - 0.30573) - 1.96 \times \sqrt{0.2279}]$ a $\exp[(0.0038 - 0.30573) + 1.96 \times \sqrt{0.2279}]$).

Tabela 6.14 Estimativas dos parâmetros, erros padrão, valores-p e valores-p (sem o doente 49).

Parâmetro	Estimativa	SE	valor - T	valor - p
β_0	1.067	0.133	8.009	$1.16e - 15$
β_1	0.012	0.105	0.114	0.9091
β_2	-0.007	0.185	-0.036	0.9709
β_3	-0.346	0.148	-2.341	0.0192
d_{11}	0.449			
d_{22}	0.215			
d_{12}	-0.009			

Repetindo a análise anterior mas agora sem o doente 49, a hipótese $H_0 : \beta_3 = 0$ continua a ser rejeitada (valor-p=0.0192). Na Tabela 6.14 são dadas as estimativas dos parâmetros assim como os erros padrão, valor da estatística de teste e respectivos valores-p para o GLMM sem o doente 49. A redução da taxa esperada de episódios entre o *follow-up* e a *baseline* para um doente no grupo *prograbide* é agora de 28% ($e^{0.012-0.346} \approx 0.72$).

6.4 Locust

Neste exemplo a resposta é binária indicando se o gafanhoto se mexe (1) ou não (0). Os gafanhotos foram divididos em dois grupos de tratamento (“alimentado” e “não alimentado”) e, dentro de cada grupo, os indivíduos são alternadamente “macho” e “fêmea”. O objectivo desta experiência, tal como foi referido na Capítulo 1, é o de saber se a probabilidade de se mover é a mesma nos dois grupos. A probabilidade marginal de sucesso é o parâmetro de interesse. À se-

melhança dos exemplos anteriores o modelo marginal e o modelo com efeitos aleatórios são apresentados. A biblioteca `bind` é a utilizada no ajustamento dos dois modelos. Recorde-se que nesta biblioteca a dependência entre observações de um memo indivíduo é feita em termos de *odds ratio* e é usado o método ML para obter as estimativas dos parâmetros do modelo independentemente de ser marginal ou com efeitos aleatórios. Na análise que se segue o sexo não foi considerado pois em estudos realizados anteriormente o efeito correspondente mostrou não ser estatisticamente significativo (Gonçalves, 2002).

6.4.1 Modelo Marginal

A probabilidade marginal de locomoção é modelada de acordo com o modelo de regressão logística

$$\begin{aligned} \text{logit}[E(Y_{it})] = \text{logit}(\mu_{it}) = & \beta_0 + \beta_1 \text{Tempo}_{it} + \beta_2 \text{Tempo}_{it}^2 + \beta_3 \text{feed}_i \\ & + \beta_4 (\text{Tempo}_{it} \times \text{feed}_i) + \beta_5 (\text{Tempo}_{it}^2 \times \text{feed}_i) \end{aligned}$$

com $i = 1, \dots, 24$, $t = 1, 2, \dots, 161$, $\text{Tempo}_{it} = \frac{t}{120}$ (uma unidade de tempo corresponde a 1 hora) e onde Y_{it} é a variável aleatória binária, que assume o valor 1 quando o i -ésimo gafanho se mexe na t -ésima ocasião e 0 caso contrário. Quanto a feed_i é uma variável indicatriz que assume o valor 1 se o i -ésimo gafanho está no grupo “alimentado” e 0 se está no grupo “não alimentado”.

No ajustamento do modelo é usado o método ML para estimar os parâmetros e são consideradas três estruturas de correlação a que correspondem outros tantos modelos. O *Modelo I* com uma estrutura de correlação independente, o *Modelo II* com uma estrutura de correlação Markoviana de primeira ordem (MC1) e o *Modelo III* com uma estrutura de correlação Markoviana de segunda ordem (MC2). A classe de referência considerada foi $\text{feed}=0$. O teste de razão de

Tabela 6.15 Valores de AIC, logLik, ΔD e valor-p.

Modelos	AIC	logLik	Teste	ΔD	valor - p
<i>Modelo I</i>	3479.986	-1733.993			
<i>Modelo II</i>	3227.979	-1606.989	<i>IvsII</i>	254.007	< 0.0001
<i>Modelo III</i>	3141.527	-1562.763	<i>IIvsIII</i>	88.452	< 0.0001

Tabela 6.16 Estimativas dos parâmetros, erros padrão, valores-T e valores-p para o Modelo III-MC2.

Parâmetro	Estimativa	SE	valor - T	valor - p
β_0	-1.530	0.255	-6.008	< 0.0000
β_1	4.432	0.848	5.228	< 0.0000
β_2	-2.415	0.597	-4.047	0.0001
β_3	-2.153	0.533	-4.039	0.0001
β_4	4.826	1.637	-2.947	0.0032
β_5	4.217	1.098	3.842	0.0001
$\log \psi_1$	1.459	0.112	13.035	< 0.0000
$\log \psi_2$	0.914	0.097	9.454	< 0.0000

verossimilhanças é usado para comparar os modelos e na Tabela 6.15 encontram-se os valores do AIC, do logaritmo da verossimilhança, a diferença dos desvios (ΔD) e o respectivo valor-p. Da análise dos resultados é seleccionado o modelo com uma estrutura de correlação MC2 (*Modelo III*). Na Tabela 6.16 são dadas as estimativas dos parâmetros assim como os erros padrão, valores da estatística de teste e respectivos valores-p para o *Modelo III*. Todos os parâmetros são significativos ao nível de significância 5% mostrando que o ajustamento quadrático é adequado e que a diferença entre os dois grupos é esta-

tisticamente significativa. Em particular, os parâmetros $\lambda_1 = \log \psi_1$ e $\lambda_2 = \log \psi_2$ são marcadamente significativos indicando claramente uma correlação serial de segunda ordem.

A análise gráfica dos resíduos para o *Modelo I* e para o *Modelo III* encontra-se nas Figuras 6.15 e 6.16, respectivamente. Da comparação dos gráficos das funções de autocorrelação (ACF) e de autocorrelação parcial (PACF) empíricas dos resíduos para os dois modelos é possível ver o efeito da correlação MC2.

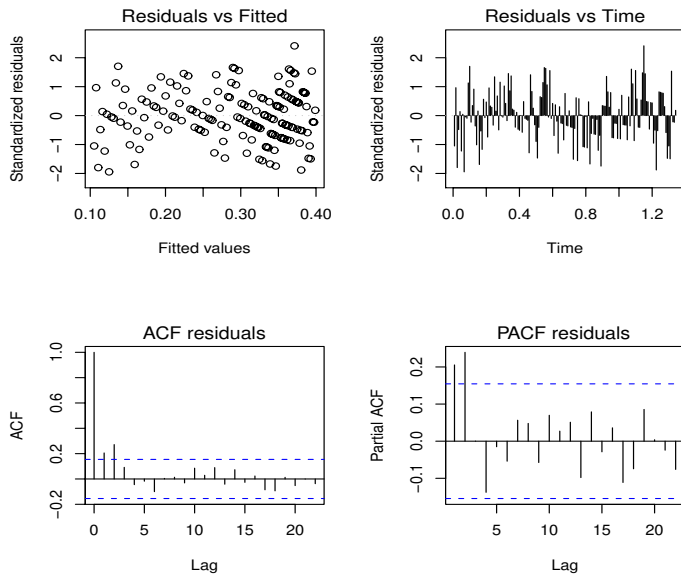


Figura 6.15 Gráficos da análise de resíduos para o Modelo I (independente).

A estimativa da probabilidade de locomoção é dada por:

(i) $feed=0$

$$\frac{\exp(-1.530 + 4.432 \times Tempo - 2.415 \times Tempo^2)}{1 + \exp(-1.530 + 4.432 \times Tempo - 2.415 \times Tempo^2)}$$

(ii) $feed=1$

$$\frac{\exp(-3.683 - 0.394 \times Tempo + 1.802 \times Tempo^2)}{1 + \exp(-3.683 - 0.394 \times Tempo + 1.802 \times Tempo^2)}$$

cujo gráfico está representado na Figura 6.17.

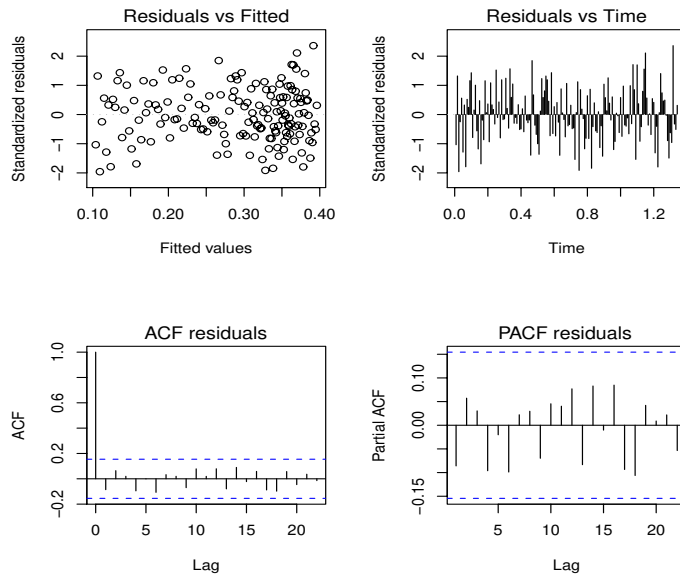


Figura 6.16 Gráficos da análise resíduos para o Modelo III (MC2).

Por exemplo, a estimativa da probabilidade de locomoção nos tempos 0.5, 0.8 e 1 é de, respectivamente, 0.5205, 0.6153 e 0.6193 para o grupo dos “não alimentados” ($feed=0$) e de 0.0314, 0.0549 e 0.093 para o grupo dos “alimentados” ($feed=1$).

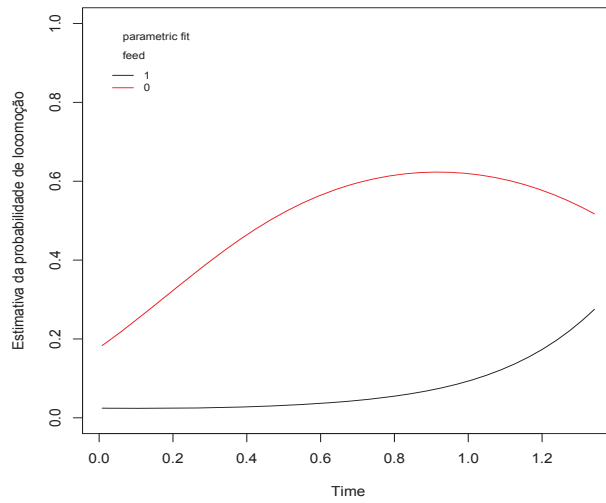


Figura 6.17 Gráfico da estimativa da probabilidade de locomoção (Modelo III - MC2).

6.4.2 Modelo Linear Generalizado Misto

Considere-se o modelo de regressão logística dado na secção anterior mas agora com efeito aleatório na intersecção

$$\begin{aligned} \text{logit}[E(Y_{it}|\mathbf{b}_i)] &= \text{logit}(\mu_{it}^b) \\ &= \beta_0 + b_{0i} + \beta_1 \text{Tempo}_{it} + \beta_2 \text{Tempo}_{it}^2 + \beta_3 \text{feed}_i \\ &\quad + \beta_4 (\text{Tempo}_{it} \times \text{feed}_i) + \beta_5 (\text{Tempo}_{it}^2 \times \text{feed}_i) \end{aligned}$$

em que b_{0i} é uma variável aleatória Gaussiana com valor médio 0 e variância d_{11} , $b_{0i} \sim N(0, d_{11})$. Para o modelo dado vão-se considerar duas estruturas de correlação a que correspondem dois modelos. O modelo com um efeito aleatório na intersecção e estrutura de correlação MC1 (*Modelo IV- MC1R*) e o modelo com um efeito aleatório na intersecção e estrutura de correlação MC2 (*Modelo V-MC2R*).

Após o ajustamento dos modelos aos dados é feita a comparação entre eles através do teste de razão de verosimilhanças obtendo-se o valor de $\Delta D = 28.8869$ a que corresponde um valor-p de $7.6729e - 08$ pelo que é seleccionado o *Modelo V*. Os gráficos referentes à análise dos resíduos para os *Modelos IV* (MC1R) e *V* (MC2R) encontram-se nas Figuras 6.18 e 6.19, respectivamente.

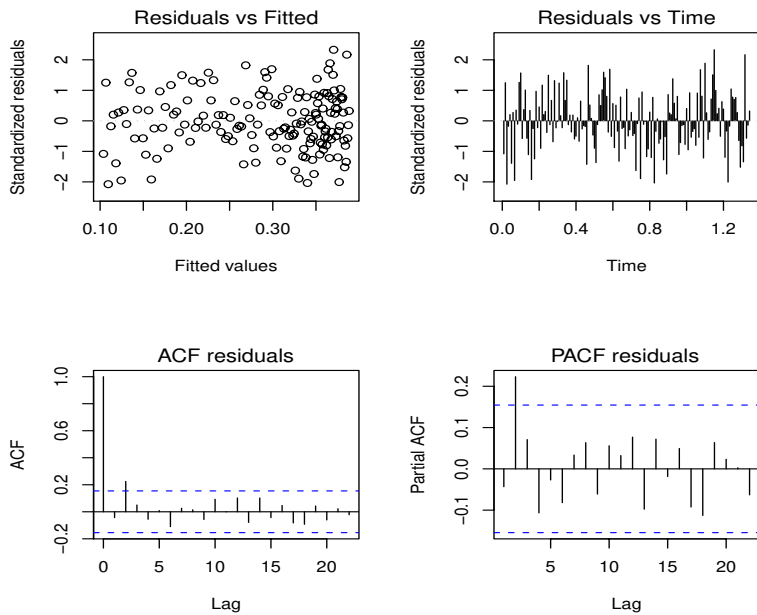


Figura 6.18 Gráficos da análise de resíduos para o *Modelo IV* (MC1R).

Na Tabela 6.17 são dadas as estimativas dos parâmetros assim como os erros padrão, valores da estatística de teste e respectivos valores-p para o *Modelo V*.

Para se obter a estimativa da probabilidade marginal utiliza-se o factor de aproximação indicado na Secção 5.7, sendo neste caso $d_{11} = 0.9972$. Na Figura 6.20 está representado o gráfico para aproximação

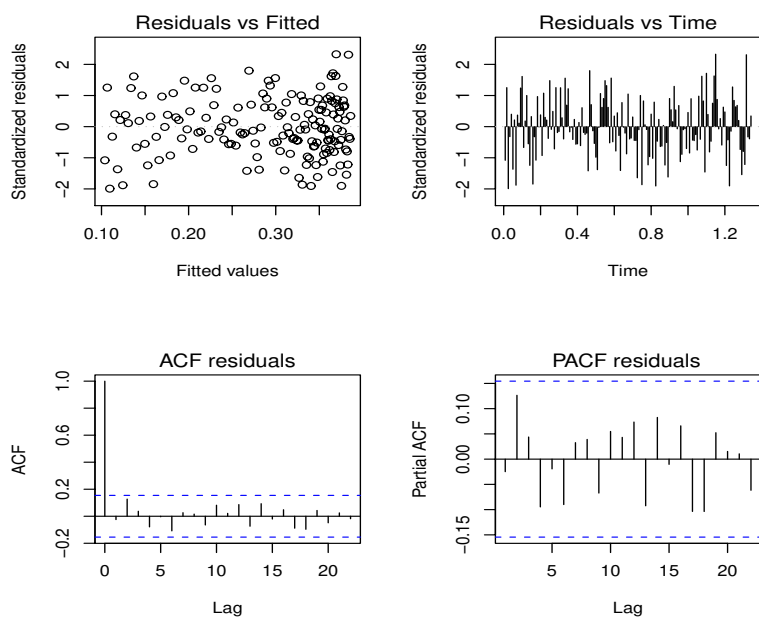


Figura 6.19 Gráficos da análise de resíduos para o Modelo *V* (MC2R).

da estimativa da probabilidade de locomoção marginal (*adjusted parametric fit*) e a estimativa da probabilidade de locomoção para um gafanhoto com o efeito aleatório não observado na intersecção $b_{0i} = 0$, em ambos os grupos, designada por *unadjusted fit*.

Com base no *Modelo V* podem ser calculadas as estimativas individuais da probabilidade de locomoção. Na Figura 6.21 apresenta-se o gráfico dessas estimativas para os gafanhotos 8, 10, 13, 22 e 24.

Por último faz-se notar que as estimativas dos efeitos fixos no *Modelo V* são maiores, em valor absolutos, do que as obtidas para o *Modelo III*. Esta diferença tem a ver com a diferente interpretação dos β 's nas duas classes de modelos.

Tabela 6.17 Estimativas dos parâmetros, erros padrão, valores- T e valores- p para o Modelo III.

Parâmetro	Estimativa	SE	valor - T	valor-p
β_0	-1.815	0.368	-4.93	< 0.0000
β_1	5.324	0.759	7.010	< 0.0000
β_2	-2.939	0.532	-5.529	< 0.0001
β_3	-2.305	0.662	-3.484	0.0005
β_4	-6.074	1.565	-3.882	0.0001
β_5	5.066	1.044	4.855	< 0.0000
$\log \psi_1$	0.981	0.113	8.717	< 0.0000
$\log \psi_2$	0.558	0.103	5.420	< 0.0000

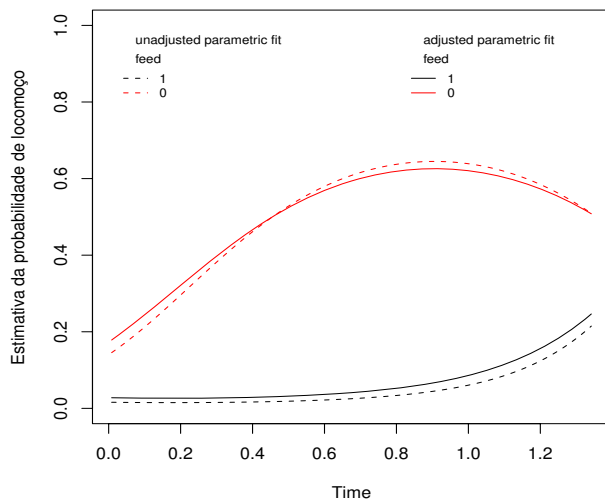


Figura 6.20 Gráfico da estimativa da probabilidade de locomoção para o Modelo V (MC2R).

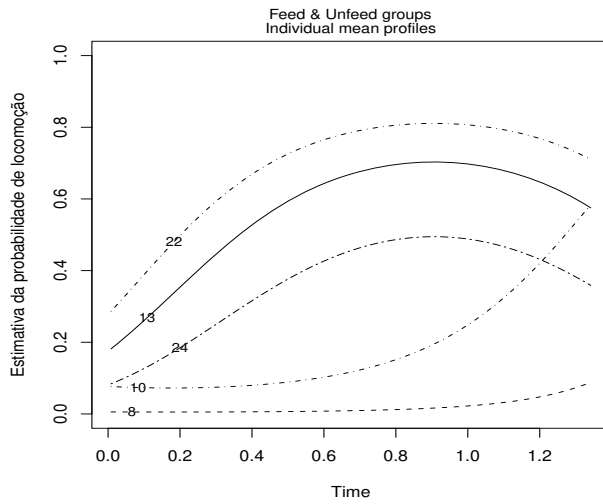


Figura 6.21 Gráfico da estimativa da probabilidade de locomoção para os indivíduos 8, 10, 13, 22 e 24 (Modelo V - MC2R).

Bibliografia

- AIRY, G. B. (1981). *On the Algebraical and Numerical Theory of Errors of Observation and the Combination of Observations*. Macmillan, London.
- AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–723.
- AZZALINI, A. (1994a). Logistic regression and other discrete data models for serially correlated observations. *J. Ital. Statist. Soc.* **3**, 169–179.
- AZZALINI, A. (1994b). Logistic regression for autocorrelated data with application to repeated measures. *Biometrika* **81**, 767–775. Amendment: vol. 84 (1997), 989.
- AZZALINI, A. (1996). *Statistical Inference*. Chapman & Hall, London, 2 edition.
- AZZALINI, A. & CHIOGNA, M. (1997). *S-plus* tools for the analysis of repeated measurements data. *Computational Statistics* **12**, 53–66.
- BATES, D., MAECHLER, M., & BOLKER, B. (2011). *lme4: Linear mixed-effects models using Eigen and Eigen++*. R package version 0.999375-39.
- BOLKER, B., BROOKS, M., CLARK, C., GEANGE, S., POULSEN, J., STEVENS, M., & WHITE, J. (2008). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution* **23**, 16–22.

Ecology and Evolution **24**, 127–135.

BOOTH, J. & HOBERT, J. (1999). Maximizing generalized linear mixed model likelihoods with an automated monte carlo em algorithm. *Journal of the Royal Statistical Society, Series B* **61**, 265–285.

BOX, G., JENKINS, G., & REINSEL, G. (1994). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 3 edition.

BRESLOW, N. & CLAYTON, D. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.

BRESLOW, N. & LIN, X. (1995). Bias correction in generalized linear mixed models with a single component of dispersion. *Biometrika* **82**, 81–91.

CLEVELAND, W. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74**, 829–836.

CRESSIE, N. (1993). *Statistics for Spatial Data*. wiley, New York.

CRESSIE, N. & HAWKINS, D. (1980). Robust estimation of the variogram. *Journal of the International Association of Mathematical Geology* **12**, 115–125.

CROWDER, M. & HAND, D. (1990). *Analysis of Repeated Measures*. Chapman & Hall, London.

DANIELS, M. & HOGAN, J. (2007). *Missing Data in Longitudinal Studies. Strategies for Bayesian Modelling and sensitivity Analysis*. Chapman & Hall /CRC, New York.

DAVIDIAN, M. & GILTINIAN, D. M. (1993). Some sample methods for estimating intraindividual variability in nonlinear mixed effects models. *Biometrics* **49**, 59–73.

DAVIDIAN, M. & GILTINIAN, D. M. (1995). *Nonlinear models for repeated measurement data*. Chapman & Hall, London.

- DEGRUTTOLA, V., LANGE, N., & DAFNI, U. (1991). Modeling the progression of hiv infection. *Journal of the American Statistical Association* **86**, 569–577.
- DEMPSTER, A., LAIRD, N., & RUBIN, D. (1977). Maximum likelihood for incomplete data via em algorithm (with discussion). *Journal of the Royal Statistical Society, Series B* **39**, 1–38.
- DIGGLE, P., HEAGERTY, P., LIANG, K.-Y., & ZEGER, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press, Oxford, 2nd edition.
- FISHER, R. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Transaction of the Royal Society of Edinburgh* **52**, 399–433.
- FISHER, R. (1925). *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh.
- FITZMAURICE, G., DAVIDIAN, M., VERBEKE, G., & MOLENBERGHS, G. (2008). *Longitudinal data analysis: a handbook of modern statistical methods*. Chapman & Hall/CRC, New York.
- FITZMAURICE, G. & LAIRD, N. (1993). A likelihood-based method for analyzing longitudinal binary responses. *Biometrika* **80**, 141–151.
- FITZMAURICE, G., LAIRD, N., & LIPSITZ, S. (1994). Analyzing incomplete longitudinal binary responses: A likelihood-based approach. *Biometrics* **50**, 601–612.
- FITZMAURICE, G., LAIRD, N., & ROTNITZKY, A. (1993). Regression models for discrete longitudinal responses. *Statistical Science* **8**, 284–309.
- FITZMAURICE, G., LAIRD, N., & WARE, J. (2004). *Applied Longitudinal Analysis*. Wiley, New York.
- FITZMAURICE, G., MOLENBERGHS, G., & LIPSITZ, S. (1995). Regression models for longitudinal binary responses with informative

- dropouts. *Journal of the Royal Statistical Society , Series B* **57**, 691–704.
- GILKS, W., RICHARDSON, S., & SPEIGELHALTER, D. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.
- GOLDENSTEIN, H. (1991). Nonlinear multilevel models, with an application to discrete response data. *Biometrika* **78**, 45–51.
- GOLDENSTEIN, H. & RASBASH, J. (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A* **159**, 505–513.
- GONÇALVES, M. (2002). *Likelihood methods for discrete longitudinal data*. PhD thesis, University of Lisbon.
- GONÇALVES, M. & AZZALINI, A. (2008). Using markov chains for marginal modelling of binary longitudinal data in an exact likelihood approach. *Metron* **LXVI**, 157–181.
- GONÇALVES, M., CABRAL, M., & AZZALINI, A. (2010). *bild: a package for BInary Longitudinal Data*. version 1.0-0.
- GONÇALVES, M., CABRAL, M., RUIZ DE VILLA, M., ESCRICH, E., & SOLANAS, M. (2007). Likelihood approach for count data in longitudinal experiments. *Computational Statistics and Data Analysis* **51**, 6511–6520.
- HAND, D. & CROWDER, M. (1996). *Practical Longitudinal Data Analysis*. Chapman & Hall, London.
- HARVILLE, D. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* **61**, 383–385.
- HARVILLE, D. (1976). Extensions of gauss-markov theorem to include the estimation of random effects. *The Annals of Statistics Associations* **72**, 320–340.
- HARVILLE, D. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of American Statistical Association* **72**, 320–338.

- HEAGERTY, P. (1999). Marginally specified logistic-normal models for longitudinal data. *Biometrics* **55**, 668–698.
- HEAGERTY, P. (2002). Marginalized transition models and likelihood inference for longitudinal categorical data. *Biometrics* **58**, 342–351.
- HENDERSON, C., KEMPTHORNE, O., SEARLE, S., & VON KROSIG, C. (1959). Estimation of environmental and genetic trends from records subject to culling. *Biometrics* **15**, 192–218.
- HØJSGAARD, S., HALEKOH, U., & YAN, J. (2011). *geepack: Generalized Estimating Equation Package*. R package version 1.0-18.
- JARDIM, J. (2001). Análise do recrutamento de pescada (*merluccius merluccius*) na costa continental portuguesa. Uma abordagem espaço-temporal. Master's thesis, Faculdade de Ciências da Universidade de Lisboa.
- JOWAHEER, V. & SUTRADHAR, B. (2002). Analysing longitudinal count data with overdispersion. *Biometrika* **89**, 389–399.
- KORN, E. & WHITTEMORE, A. (1979). Methods for analyzing panel studies of acute health effects of air pollution. *Biometrics* **35**, 795–802.
- LAIRD, N., LANGE, N., & STRAM, D. (1987). Maximum likelihood computations with repeated measures: Application of the em algorithm. *Journal of the American Statistical Association* **82**, 97–105.
- LAIRD, N. & WARE, J. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.
- LIANG, K.-Y. & ZEGER, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.
- LIANG, K.-Y., ZEGER, S., & QAQISH, B. (1992). Multivariate regression analysis for categorical data. *Journal of the Royal Statistical Society, Series B* **54**, 3–40.
- LIN, D. Y., WEI, L., & YING, Z. (2002). Model-checking techniques based on cumulative residuals. *Biometrics* **58**, 1–12.

- LINDSTROM, M. & BATES, D. (1988). Newton-raphson and em algorithms for linear mixed-effects models for repeated measures data. *Journal of American Statistical Association* **83**, 1014–1022.
- LITTELL, R., MILLIKEN, G., STROUP, W., & WOLFINGER, R. (1996). Sas system for mixed models. *SAS Institute Inc., Cary, NC*.
- LITTLE, R. & RUBIN, D. (1987). *Statistical Analysis with Missing data*. Wiley & Sons, New York.
- LONGFORD, N. (1993). *Random coefficients models*. Oxford: Oxford University Press.
- LONGFORD, N. (2005). *Missing Data and Small Area Estimation*. Springer, New York.
- MACDONALD & RAUBENHEIMER, D. (1995). Hidden markov models and animal behaviour. *Biometrical Journal* **37**, 701–712.
- MATHERON, G. (1997). *Traite de Geostatistique Appliquee, Vol. I of Memoires du Bureau de Recherches Geologiques et Minières*. Editions Technip, Paris.
- MCCULLOCH, C. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* **92**, 162–170.
- MCCULLOCH, C. & SEARLE, S. (2001). *Generalized, Linear and Mixed Models*. Wiley, New York.
- MOLENBERGHS, G. & KENWARD, M. (2007). *Missing Data in Clinical Studies*. Wiley.
- MOLENBERGHS, G. & VERBEKE, G. (2005). *Models for Discrete Longitudinal Data*. Springer, New York.
- MONTE-PEGADO, C. (2003). Modelos lineares mistos em S. Uma aplicação à modelação de populações biológicas. Master's thesis, Faculdade de Ciências da Universidade de Lisboa.

- PAIK, M. (1988). Repeated measurement analysis for nonnormal data in small samples. *Communications in Statistics- Simulation and Computation* **17**, 1155–1171.
- PATTERSON, H. & THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554.
- PINHEIRO, J. AND BATES, D., DEBROY, S., SARKAR, D., & R DEVELOPMENT CORE TEAM (2011). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-98.
- PINHEIRO, J. (1994). *Topics in Mixed-Effects Models*. PhD thesis, University of Wisconsin, Madison, WI.
- PINHEIRO, J. & BATES, D. (1995). Approximations to the log-likelihood function in nonlinear mixed-effects models. *Journal of Computational and Graphical Statistics* **4**, 12–35.
- PINHEIRO, J. & BATES, D. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag, New York.
- R DEVELOPMENT CORE TEAM (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- ROBINS, J., ROTNIZKY, A., & ZHAO, L. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90**, 106–121.
- ROTNIZKY, A. & JEWELL, N. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* **77**, 485–497.
- RUBIN, D. (1976). Inference and missing data. *Biometrika* **63**, 81–92.
- SCHWARZ (1978). Estimating the dimension of a model. *The annals of Statistics* **6**, 461–464.
- SEARLE, S., CASELLA, G., & MCCULLOCH, C. (1992). *Variance Componentes*. Wiley, New York.

- STRAM, D. & LEE, J. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171–1177.
- THALL, P. & VAIL, S. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics* **46**, 657–671.
- THISTED, R. (1988). *Elements of Statistical Computing*. Chapman & Hall, London.
- TIERNY & KADANE (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* **81**, 82–86.
- VERBEKE, G. & MOLENBERGHS, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer, New York.
- WARE, J., LIPSITZ, S., & SPEIZER, F. (1988). Issues in the analysis of repeated categorical outcomes. *Statistics in Medicine* **7**, 95–108.
- WEDDERBURN, R. (1974). Quasi-likelihood, generalized linear models and the gauss-newton method. *Biometrika* **61**, 439–447.
- WOLFINGER, R. & O'CONNELL, M. (1993). Generalized linear mixed models: A pseudo-likelihood approach. *Journal of statistical Computation and Simulation* **48**, 233–243.
- WOLFINGER, R., TOBIAS, R., & SALL, J. (1994). Computing gaussian likelihoods and their derivatives for general linear mixed models. *SIAM Journal on Scientific Computing* **15**, 1294–1310.
- ZEGER, S., LIANG, K.-Y., & ALBERT, P. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44**, 1049–1060.
- ZEGER, S. & QAQISH, B. (1988). Markov regression models for time series. *Biometrics* **44**, 1019–1031.
- ZUUR, A., IENO, E., WALKER, N., SAVELIEV, A., & SMITH, G. (2009). *Mixed Effects Models and Extensions in Ecology with R*. Springer, New York.

ISBN: 978-972-8890-24-7
Depósito Legal: 332365/11



VA160610

